Selected Works in Probability and Statistics

For futher volumes:
http://www.springer.com/series/8556

Peter Guttorp • David Brillinger
Editors

Selected Works of David Brillinger

Springer

Editors
Peter Guttorp
University of Washington
Seattle, USA
peter@stat.washington.edu
and
Norwegian Computing Center
114 Blindern, Oslo
Norway
peter.guttorp@nr.no

David Brillinger
Department of Statistics
University of California
Berkeley, CA, USA
brill@stat.berkeley.edu

ISBN 978-1-4614-1343-1 e-ISBN 978-1-4614-1344-8
DOI 10.1007/978-1-4614-1344-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011942481

Printed on acid-free paper

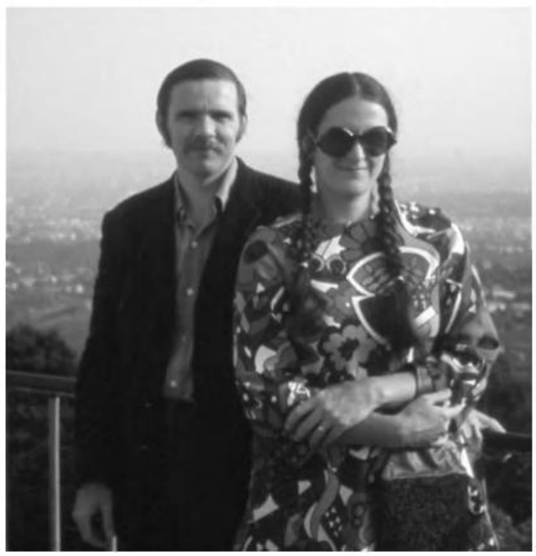
Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Series

Springer's Selected Works in Probability and Statistics series offers scientists and scholars the opportunity of assembling and commenting upon major classical works in statistics, and honors the work of distinguished scholars in probability and statistics. Each volume contains the original papers, original commentary by experts on the subject's papers, and relevant biographies and bibliographies.

Springer is committed to maintaining the volumes in the series with free access of SpringerLink, as well as to the distribution of print volumes. The full text of the volumes is available on SpringerLink with the exception of a small number of articles for which links to their original publisher are included instead. These publishers have graciously agreed to make the articles freely available on their websites. The goal is maximum dissemination of this material.

The subjects of the volumes have been selected by an editorial board consisting of Anirban DasGupta, Peter Hall, Jim Pitman, Michael Sörensen, and Jon Wellner.



Lorie and David at the ISI meeting in Vienna 1973 (where he delivered the paper [32] in section 27).

Preface

When I was asked to put together a Selected Works volume for David Brillinger, I never even considered saying no. But I realized that the hardest part of the job would be to convince David to let me do it. "The question that I keep asking myself is 'Is that really me?'" he wrote to me. But eventually he relented, when I argued that this would make some of his interesting papers, that are now hard to find, available to everyone. He started sending me thick envelopes full of papers. Going through all of them was pure joy. The breadth of David's contributions is incredible. Selecting which 600 pages out of the 224 entries in his bibliography (when I write this–when you read it there are undoubtedly more) to include was not so much joy. Some of his richest papers are just too long ([104] with Tukey, *Spectrum analysis in the presence of noise: some issues and examples*, is 141 pages, but can be found in Tukey's Collected Works that David edited; [53] *Comparative aspects of the study of ordinary time series and of point processes* is 101. Both these papers are full of new ideas, many of which have not yet been fully developed). However, after a sequence of emails and a long session at a Berkeley coffee shop, we agreed on the current selection. It was fortunate that Victor Panaretos' interview with David for *Statistical Science* was finished during this process, and could be included in this volume.

The selection contains all of David's named lectures (Wald, Fisher, Herzberg, Hunter and Neyman), in which he carefully presents material from his research, always containing important answers to scientific questions and illustrated with LOTS of pictures. In addition, there are papers from his main methodological areas: time series and point processes; and from his three main scientific interests: neurophysiology, seismology and population biology. We tried to make sure that there were papers with the main people he has worked with: Bruce Bolt, Jose Segundo, Alan Ager, Brett Stewart and Haiganoush Preisler. Some important work, for example his papers in demography, or on using wavelets, there simply was not room for. As I said, his scientific work is very broad.

I was fortunate to have both Jerzy Neyman and David Brillinger as teachers and advisers. The principal lesson I learned from both of them is the importance of working hard at understanding the science behind the questions you are trying to address. Collaboration is key to modern statistical science.

In finding appropriate researchers to comment on David's papers I needed not go beyond his list of former PhD students (39 at last count). David's more theoretical work is discussed by Victor Panaretos (PhD 2007), a Greek working in Switzerland. Time series papers are discussed by Pedro Morettin (PhD 1973) from Brazil, a country David loves and visits as often as he can. Some biological papers are addressed by Tore Schweder (PhD 1975) from Norway, another country that David frequently visits, assisted by Haiganoush Preisler (PhD 1977), a Palestinian working in the US, while I (PhD 1980), a Swede working in the US, deal with point processes, neurophysiology and earth sciences.

David is a very close friend of mine. Apart from statistics, we share interests in politics, hockey, and soccer, which we discuss at great length in person or briefly in emails. When he recently was selected as honorary member of the Canadian Statistical Society, he remarked to me "Somehow that's all about another person. I am just me, a kid from Toronto who a lot of people have helped." Well, the kid from Toronto is a member of the Canadian, Brazilian and Norwegian Academies, has three honorary doctorates, and a share of the Nobel Peace Prize for work done for the IPCC [112]. Not only is he a most accomplished scientist, but he is the epitome of the modern statistical scientist.

So David, this is for you. We all hope you will enjoy it. Thanks for teaching us what a statistician can and should be, being there for us to talk about science, soccer, and survival, writing poetry in a dissertation or signing in Hollerith. We owe you so much.

Peter Guttorp

Acknowledgements

This series of selected works is possible only because of the efforts and cooperation of many people, societies, and publishers. The series editors originated the series and directed its development. The volume editors spent a great deal of time organizing the volumes and compiling the previously published material. The contributors provided comments on the significance of the papers. The societies and publishers who own the copyright to the original material made the volumes possible and affordable by their generous cooperation:

American Mathematical Society
American Statistical Association
Berkeley Electronic Press
Canadian Statistical Society
Elsevier
Institute of Electrical and Electronics Engineers
Institute of Mathematical Statistics
International Statistical Institute
John Wiley & Sons
Mathematical Association of America
National Academy of Sciences
Nova Science
Oxford University Press
Royal Astronomical Society
Springer Science+Business Media
Statistical Society of Canada
Taylor and Francis
University of California Press

Contents

Contributors

Pedro A. Morettin
Department of Statistics, Universidade de Sao Paulo, C.P. 66281,
05311-970- Sao Paulo, Brazil

Victor Panaretos
Institut de mathématiques d'analyse et applications, École Polytechnique
Federale de Lausanne, Station 8 - Bâtiment MA, CH-1015 Lausanne, Switzerland

Haiganoush Preisler
Southwest Research Station, United States Forest Service, 800 Buchanan Street,
West Annex Building, Albany, CA 94710-0011, USA

Tore Schweder
Department of Economics, Universitetet i Oslo, Postboks 1095 Blindern,
NO-0317 OSLO, Norway

Bibliography

[1] Brillinger, D., 1961. *Asymptotic means and variances in the k-dimensional case*. Ph.D. thesis, Princeton University. 70 pp

[2] Brillinger, D., 1961. A note on the definition of asymptotic normality for multi- dimensional random variables. *Estadística* **20** 371

[3] Brillinger, D. R., 1961. A justification of some common laws of mortality. *Soc. Actuar. Trans.* **13** 116–126

[4] Brillinger, D. R., 1962. Examples bearing on the definition of fiducial probability with a bibliography. *Annals of Mathematical Statistics* **33** 1349–1355

[5] Brillinger, D. R., 1962. A note on the rate of convergence of a mean. *Biometrika* **49** 574–576

[6] Brillinger, D. R., 1963. Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group. *Ann. Math. Statist.* **34** 492–500

[7] Brillinger, D. R., 1963. A note on the re-use of samples. *Ann. Math. Statist.* **34** 341–343

[8] Brillinger, D., 1964. A technique for measuring the spectral density matrix of two signals. *Proc. IEEE* **52** 103–104

[9] Brillinger, D. R., 1964. The asymptotic behaviour to Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. *Rev. Internat. Statist. Inst., Int. Stat. Rev.* **32** 202–206

[10] Brillinger, D., 1965. A property of low-pass filters. *SIAM Review* **7** 65–67

[11] Brillinger, D. R., 1965. An introduction to polyspectra. *Ann. Math. Statist.* **36** 1351–1374

[12] Brillinger, D. R., 1965. A moving average representation for random variables covariance stationary on a finite time interval. *Biometrika* **52** 295–297

[13] Brillinger, D., 1966. The application of the jackknife to the analysis of sample surveys. *Commentary. The Journal of the Market Research Society* **8** 74–80

[14] Brillinger, D. R., 1966. The analyticity of the roots of a polynomial as functions of the coefficients. *Mathematics Magazine* **39** 145–147

[15] Brillinger, D. R., 1966. An extremal property of the conditional expectation. *Biometrika* **53** 594–595

[16] Brillinger, D. R., 1967. Bounded polymeasures and associated translation commutative polynomial operators. *Proc. Amer. Math. Soc.* **18** 487–491

[17] Brillinger, D. R. and Rosenblatt, M., 1967. Asymptotic theory of estimates of k-th order spectra. In Harris, B. (ed.), *Advanced Seminar on Spectral Analysis of Time Series (Madison, WI, 1966)*, 153–188. John Wiley, New York, NY

[18] Brillinger, D. R. and Rosenblatt, M., 1967. Asymptotic theory of estimates of kth-order spectra. *Proc. Natl. Acad. Sci. U.S.A.* **57** 206–210

[19] Brillinger, D. R. and Rosenblatt, M., 1967. Computation and interpretation of k-th order spectra. In Harris, B. (ed.), *Advanced Seminar on Spectral Analysis of Time Series (Proc. Advanced Sem., Madison, Wis., 1966)*, 189–232. John Wiley, New York

[20] Brillinger, D. R., 1968. Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros. *J. Roy. Statist. Soc. Ser. B* **30** 145–159

[21] Brillinger, D., 1969. The calculation of cumulants via conditioning. *Ann. Inst. Stat. Math.* **21** 215–218

[22] Brillinger, D., 1969. A search for a relationship between monthly sunspot numbers and certain climatic series. In *Bulletin of the International Statistical Institute. Volume 43, Book 1*, 293–307. International Statistical Institute

[23] Brillinger, D. R., 1969. Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375–390

[24] Brillinger, D. R., 1969. An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* **75** 545–547

[25] Brillinger, D. R., 1969. The canonical analysis of stationary time series. In Krishnaiah, P. (ed.), *Multivariate Analysis, II (Dayton, OH, 1968)*, 331–350. Academic Press, New York, NY

[26] Brillinger, D. R. and Hatanaka, M., 1969. An harmonic analysis of nonstationary multivariate economic processes. *Econometrica* **37** 131–141

[27] Brillinger, D., 1970. The identification of polynomial systems by means of higher order spectra. *Journal of Sound and Vibration* **12** 301–313

[28] Brillinger, D. and Hatanaka, M., 1970. A permanent income hypothesis relating to the aggregate demand for money (an application of spectral and moving spectral analysis). *Economic Studies Quarterly* **21** 44–71

[29] Brillinger, D. R., 1970. The frequency analysis of relations between stationary spatial series. In Pyke, R. (ed.), *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress on Time Series and Stochastic Processes. Convexity and Combinatorics (Vancouver, BC, 1969)*, 39–81. Canadian Mathematical Society, Ottawa

[30] Brillinger, D. R., 1972. On the number of solutions of systems of random equations. *Ann. Math. Statist.* **43** 534–540

[31] Brillinger, D. R., 1972. The spectral analysis of stationary interval functions. In Le Cam, L., Neyman, J. and Scott, E. (eds.), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, CA, 1970/1971)*, vol. 1, 483–513. University of California Press, Berkeley, CA

[32] Brillinger, D., 1973. An empirical investigation of the Chandler wobble and two proposed excitation processes. In *Bulletin of the International Statistical Institute. 39th Session: Vienna 1973. Book 3*, 413–434. International Statistical Institute

[33] Brillinger, D., 1973. A power spectral estimate insensitive to transients. *Technometrics* **15** 559–562

[34] Brillinger, D. R., 1973. The analysis of time series collected in an experimental design. In Krishnaiah, P. (ed.), *Multivariate analysis, III (Dayton, OH, 1972)*, 241–256. Academic Press, New York, NY

[35] Brillinger, D. R., 1973. Estimation of the mean of a stationary time series by sampling. *J. Appl. Probab.* **10** 419–431

[36] Brillinger, D. R., 1974. The asymptotic distribution of the Whittaker periodogram and a related chi-squared statistic for stationary processes. *Biometrika* **61** 419–422

[37] Brillinger, D. R., 1974. Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *Ann. Probab.* **2** 815–827

[38] Brillinger, D. R., 1974. Fourier analysis of stationary processes. *Proc. IEEE* **62** 1628–1643

[39] Brillinger, D. R., 1974. *Time Series: Data Analysis and Theory.* International Series in Decision Processes. Holt, Rinehart and Winston, New York, NY

[40] Brillinger, D., 1975. Estimation of product densities. In Franc, J. (ed.), *Computer Science and Statistics: 8th Annual Symposium*, 431–438. University of California

[41] Brillinger, D., Bryant, H. and Segundo, J., 1975. Identification of synaptic operators. In McCann, G. and Marmarelis, P. (eds.), *Proceedings of the First Symposium on Testing and Identification of Nonlinear Systems*, 221–226. California Institute of Technology, Pasadena, CA

[42] Brillinger, D. R., 1975. The identification of point process systems. *Ann. Probability* **3** 909–929

[43] Brillinger, D. R., 1975. Statistical inference for stationary point processes. In Puri, M. (ed.), *Stochastic Processes and Related Topics (Bloomington, IN, 1974)*, vol. 1, 55–99. Academic Press, New York, NY

[44] Brillinger, D., 1976. Approximate estimation of the standard errors of complex statistics based on sample surveys. *New Zealand Statistician* **11** 35–41

[45] Brillinger, D., 1976. Some history of the data analysis of time series in the united states. In Owen, D. (ed.), *On the History of Statistics and Probability*, 267–280. Marcel Dekker, New York, NY

[46] Brillinger, D., Bryant, H. and Segundo, J., 1976. Identification of synaptic interactions. *Biological Cybernetics* **22** 213–228

[47] Brillinger, D. R., 1976. Estimation of the second-order intensities of a bivariate stationary point process. *J. Roy. Statist. Soc. Ser. B* **38** 60–66

[48] Brillinger, D. R., 1976. Measuring the association of point processes: a case history. *Amer. Math. Monthly* **83** 16–22

[49] Brillinger, D. R., 1977. The identification of a particular nonlinear time series system. *Biometrika* **64** 509–515

[50] Brillinger, D., 1978. A note on the estimation of evoked response. *Biological Cybernetics* **31** 141–144

[51] Brillinger, D., Jones, L. and Tukey, J., 1978. The role of statistics in weather resources management. Tech. Rep. II, Weather Modification Advisory Board, U.S. Department of Commerce, Washington, DC. U.S. Government Printing Office No. 003-0180-0091-1

[52] Brillinger, D., Udias, A. and Bolt, B., 1978. Un modelo probabílistico para la determinación del mechanismo region de los terremotos. In *II Asamblea de la Unión Geodésica y Geofísica, Barcelona*, 637–640. IGN, Madrid

[53] Brillinger, D. R., 1978. Comparative aspects of the study of ordinary time series and of point processes. In Krishnaiah, P. (ed.), *Developments in statistics, Vol. 1*, 33–133. Academic Press, New York, NY

[54] Brillinger, D. R., 1978. Henry Scheffé, 1907–1977. *J. Roy. Statist. Soc. Ser. A* **141** 406–407

[55] Brillinger, D. R., 1978. A note on a representation for the Gauss-Poisson process. *Stochastic Process. Appl.* **6** 135–137

[56] Lee, W. and Brillinger, D., 1978. A point process analysis of Chinese earthquake history. In *Proc. Conf. Seismic Gaps.* MIT Press, Cambrige, MA

[57] Silveria, W. and Brillinger, D., 1978. On the maximum wave heights in severe seas. In *Tenth Annual Offshore Technology Conference (Houston, TX, 1978)*, 1563–1568. Conference Paper no. 3232

[58] Bolt, B. and Brillinger, D., 1979. Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series. *Geophys. J. Roy. Astron. Soc.* **59** 593–603

[59] Brillinger, D., 1979. Confidence intervals for the cross covariance function. In Behara, M., Blyth, C., Campbell, L., Csörgő, M., Dawson, D., Giri, N., Joffe, A., Stanton, R. and Theodorescu, R. (eds.), *Mathematical Statistics*, vol. 5 of *Selecta Statistica Canadiana*, 1–16. Department of Mathematics, McMaster University, Hamilton, ON

[60] Brillinger, D. and Segundo, J., 1979. Empirical examination of the threshold model of neuron firing. *Biological Cybernetics* **35** 213–220

[61] Brillinger, D. R., 1979. Analyzing point processes subjected to random deletions. *Canad. J. Statist.* **7** 21–27

[62] Brillinger, D. R., 1979. A continuous form of post-stratification. *Ann. Inst. Statist. Math.* **31** 271–277

[63] Hung, G., Brillinger, D. and Stark, L., 1979. Interpretation of kernels. ii. same signed lst- and 2nd- degree (main-diagonal) kernels of the human pupillary system. *Math. Biosci.* **46** 159–187

[64] Lee, W. and Brillinger, D., 1979. On Chinese earthquake history - an attempt to model an incomplete data set by point process analysis. *Pure and Applied Geophysics* **117** 1229–1257

[65] Brillinger, D., 1980. *Vremennye ryady. Obrabotka dannykh i teoriya.* Mir, Moscow. Russian translation of first edition of Time Series: Data Analysis and Theory (Expanded Edition) Holden-Day Inc., San Francisco (1981)

[66] Brillinger, D., Guckenheimer, J., Guttorp, P. and Oster, G., 1980. Empirical modelling of population time series data: The case of age and density dependent vital rates. In Oster, G. F. (ed.), *Some Mathematical Questions in Biology, 14th Symposium (San Francisco, 1980)*, vol. 13 of *Lectures on Mathematics in the Life Sciences*, 65–90. American Mathematical Society, Providence, RI

[67] Brillinger, D. and Tiao, G. (eds.), 1980. *Directions in time series. Proceedings of the IMS Special Topics Meeting on Time Series Analysis (Ames, IA, 1978.).* Institute of Mathematical Statistics, Hayward, CA

[68] Brillinger, D., Udias, A. and Bolt, B., 1980. A probability model for regional focal mechanism solutions. *Bul. Seismol. Soc. Amer.* **70** 149–170

[69] Brillinger, D. R., 1980. Analysis of variance problems under time series models. In Krishnaiah, P. R. (ed.), *Analysis of Variance*, vol. 1 of *Handbook of Statistics*, 237–277. North-Holland, Amsterdam

[70] Brillinger, D. R., 1980. The comparison of least squares and third-order periodogram procedures in the estimation of bifrequency. *J. Time Ser. Anal.* **1** 95–102

[71] Brillinger, D. R. and Tiao, G. C. (eds.), 1980. *Directions in time series.* Institute of Mathematical Statistics, Hayward, CA

[72] Brillinger, D. R. and Tiao, G. C. (eds.), 1980. *Directions in time series.* Institute of Mathematical Statistics, Hayward, Calif.

[73] Udias, A., Buforn, E., Brillinger, D. and Bolt, B., 1980. A numerical method for individual and regional fault plane determinations. *Publ. Inst. Geophys. Pol. Acad. Sc.* **A-10** 87–94

[74] Brillinger, D., 1981. Problem 6372. *Amer. Math. Monthly* **88** 769

[75] Brillinger, D. R., 1981. The general linear model in the design and analysis of evoked response experiments. *J. Theoret. Neurobiol.* **1** 105–119

[76] Brillinger, D. R., 1981. The key role of tapering in spectrum estimation. *IEEE Trans. Acoust., Speech, Signal Processing* **29** 1075–1076. Reprinted in Carter, G.C. (ed.), Coherence and Time Delay Estimation (IEEE, 1993) 37-38

[77] Brillinger, D. R., 1981. Some aspects of modern population mathematics. *Canad. J. Statist.* **9** 173–194

[78] Brillinger, D. R., 1981. Some aspects of the analysis of evoked response experiments. In Csörgö, M., Dawson, D., Rao, J. and Saleh, A. (eds.), *Statistics and related topics (Ottawa, ON, 1980)*, 155–168. North-Holland, Amsterdam

[79] Brillinger, D. R., 1981. *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco, CA, expanded edition edn.

[80] Brillinger, D., 1982. Seismic risk assessment: some statistical aspects. *Journal of Earthquake Prediction Research* **1** 183–195

[81] Brillinger, D., 1982. Some contrasting examples of the time and frequency domain approaches to time series analysis. In El-Shaarawi, A. and Esterby, S. (eds.), *Time Series Methods in Hydrosciences (Burlington, ON, 1981)*, 1–15. Elsevier, Amsterdam

[82] Brillinger, D. R., 1982. Asymptotic normality of finite Fourier transforms of stationary generalized processes. *J. Multivariate Anal.* **12** 64–71

[83] Brillinger, D. R., 1982. A generalized linear model with "Gaussian" regressor variables. In Bickel, P. J., Doksum, K. and Hodges, J., Jr. (eds.), *A Festschrift for Erich L. Lehmann*, Wadsworth Statistics/Probability Series, 97–114. Chapman & Hall/CRC

[84] Brillinger, D. R., 1982. Some bounds for seismic risk. *Bull. Seismol. Soc. Amer.* **72** 1403–1410

[85] Comis, J., Marg, E., Brillinger, D., Rudiak, D. and Keller, E., 1982. Comparison of means and medians in the reduction of visual evoked potential data. In Riddle, W. (ed.), *Proceedings of the Fifteenth Hawaii International Conference on System Sciences*, vol. 2, 242–253. Western Periodicals

[86] Udias, A., Buforn, E., Brillinger, D. and Bolt, B., 1982. Joint statistical determination of fault plane parameters. *Physics of Earth and Planetary Interiors* **30** 178–184

[87] Brillinger, D., 1983. Jump processes. In *Encyclopedia of Statistical Sciences*, vol. 4, 336–338. John Wiley and Sons, New York, NY

[88] Brillinger, D., 1983. Statistics in fertility research: value and limitations. Tech. Rep. 19, Committee on Population and Demography, National Research Council

[89] Brillinger, D. R., 1983. The finite Fourier transform of a stationary process. In Brillinger, D. and Krishnaiah, P. (eds.), *Time Series in the Frequency Domain*, vol. 3 of *Handbook of Statistics*, 21–37. Elsevier, Amsterdam. This constitutes Chapter 2 of the book.

[90] Brillinger, D. R., 1983. The finite Fourier transform of a stationary process. In *Time series in the frequency domain*, vol. 3 of *Handbook of Statist.*, 21–37. North-Holland, Amsterdam

[91] Brillinger, D. R. and Krishnaiah, P. R. (eds.), 1983. *Time series in the frequency domain*, vol. 3 of *Handbook of Statistics.* North-Holland Publishing Co., Amsterdam

[92] Brillinger, D. R. and Preisler, H. K., 1983. Maximum likelihood estimation in a latent variable problem. In Karlin, S., Amemiya, T. and Goodman, L. (eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics*, 31–65. Academic Press, New York, NY

[93] Brillinger, D., 1984. Tukey's work on time series analysis. In Brillinger, D. (ed.), *The Collected Works of J.W. Tukey*, vol. 1, xliii–xlvi. Chapman & Hall, London

[94] Brillinger, D. and Preisler, H., 1984. An exploratory analysis of the Joyner-Boore attenuation data. *Bull. Seismol. Soc. America* **74** 1441–1450

[95] Brillinger, D. R., 1984. Statistical inference for irregularly observed processes. In Parzen, E. (ed.), *Time Series Analysis of Irregularly Observed Data (College Station, TX, 1983)*, vol. 25 of *Lecture Notes in Statistics*, 38–57. Springer, New York, NY

[96] Brillinger, D. R., 1984. Statistical inference for random processes. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983)*, 1049–1061. PWN, Warsaw

[97] Tukey, J. W., 1984. *The collected works of John W. Tukey. Vol. I.* Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA. Time series: 1949–1964, Edited and with commentary by David R. Brillinger, With introductory material by William S. Cleveland and Frederick Mosteller

[98] Tukey, J. W., 1984. *The collected works of John W. Tukey. Vol. II.* Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA. Time series: 1965–1984, Edited and with commentary by David R. Brillinger, With introductory material by William S. Cleveland and Frederick Mosteller

[99] Brillinger, D., 1985. Fourier inference: some methods for the analysis of array and nongaussian series data. *Journal of the American Water Resources Association* **21** 743–756

[100] Brillinger, D., 1985. A maximum likelihood approach to frequency-wavenumber analysis. *IEEE Trans. Acoustics, Speech, Signal Processing* **33** 1076–1085

[101] Brillinger, D., 1985. Some statistical aspects of nmr spectroscopy. In *Actas del Segundo Congreso Latinoamericano de Probabilidad y Estatística Mathematica, Caracas*, 9–17

[102] Brillinger, D., 1985. Statistics. In *New Canadian Encyclopedia*, 1754. Hurtig

[103] Brillinger, D. and Preisler, H., 1985. Further analysis of the Joyner-Boore attenuation data. *Bull. Seismol. Soc. America* **75** 611–614

[104] Brillinger, D. and Tukey, J., 1985. Spectrum analysis in the presence of noise: some issues and examples. In Brillinger, D. (ed.), *The Collected Works of J.W. Tukey*, vol. 2, 1001–1141. Chapman & Hall, London

[105] Brillinger, D. and Preisler, H., 1986. Two examples of quantal data analysis: a) multivariate point process b) pure death process in an experimental design. In *Proceedings of the 13th International Biometric Conference, Seattle*, 94–113

[106] Brillinger, D. R., 1986. The natural variability of vital rates and associated statistics. *Biometrics* **42** 693–734

[107] Brillinger, D. R., 1986. Regression for randomly sampled spatial series: the trigonometric case. *J. Appl. Probab.* **23A** 275–289

[108] Brillinger, D. R., 1986. What do seismology and neurophysiology have in common? - Statistics! *C. R. Math. Rep. Acad. Sci. Canada* **8** 1–11

[109] Brillinger, D., 1987. Analysing interacting nerve cell spike trains to assess causal connections. In Marmarelis, V. (ed.), *Advanced Methods of Physiological System Modeling*, vol. 1, 29–40. University of Southern California, Biomedical

[110] Brillinger, D., 1987. Fitting cosines: some procedures and some physical examples. In MacNeill, I. and Umphrey, G. (eds.), *Applied Probability, Stochastic Processes, and Sampling Theory*, vol. 1 of *Advances in the Statistical Sciences: Festschrift in Honor of Professor V.M. Joshi*, 75–100. D. Reidel, Boston, MA

[111] Brillinger, D., 1987. Statistiques. In *L'encyclopedie du Canada*, 1871. Hurtig. Translation of Statistics, New Canadian Encyclopedia, Hurtig (1985), p. 1754.

[112] Bloomfield, P., Brillinger, D., Nychka, D. and Stolarski, R., 1988. Statistical approaches to ozone trend detection. Tech. Rep. 18, World Meteorological Organization. Appendix A of "Report of the International Ozone Trends Panel" (WMO technical report)

[113] Brillinger, D., 1988. An elementary trend analysis of Rio Negro levels at Manaus, 1903-1985. *Revista Brasileira de Probabilidade e Estatística* **2** 63–79

[114] Brillinger, D., 1988. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics* **59** 189–200

[115] Brillinger, D. R., 1988. The maximum likelihood approach to the identification of neuronal firing systems. *Annals of Biomedical Engineering* **16** 3–16

[116] Brillinger, D. R., 1988. Some statistical methods for random process data from seismology and neurophysiology (in the 1983 wald memorial lectures). *Ann. Statist.* **16** 1–54

[117] Brillinger, D., 1989. Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. In Tanur, J. M., Mosteller, F., Kruskal, W. H., Lehmann, E. L., Link, R. F., Pieters, R. S. and Rising, G. R. (eds.), *Statistics: A Guide to the Unknown*, 249–260. Wadsworth & Brooks/Cole, 3rd ed. edn.

[118] Brillinger, D., 1989. Parameter estimation for nongaussian processes via second and third order spectra with an application to some endocrine data. In Marmarelis, V. (ed.), *Advanced Methods of Physiological System Modeling*, 53–61. Plenum, New York, NY

[119] Brillinger, D., 1989. Some examples of the statistical analysis of seismological data. In Litehiser, J. (ed.), *Observatory Seismology: A Centennial Symposium for the Berkeley Seismographic Stations*, 266–278. University of California Press, Berkeley, CA

[120] Brillinger, D., Downing, K., Glaeser, R. and Perkins, G., 1989. Combining noisy images of small crystalline domains in high resolution electron microscopy. *J. Appl. Stat.* **16** 165–175

[121] Brillinger, D. R., 1989. Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika* **76** 23–30

[122] Rosenberg, J., Amjad, A., Breeze, P., Brillinger, D. and Halliday, D., 1989. The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. Mol. Biol.* **53** 1–31

[123] Sieh, K., Stuiver, M. and Brillinger, D., 1989. A more precise chronology of earthquakes produced by the San Andreas Fault in Southern California. *J. Geophys. Res.* **94** 603–623. Reprinted in Pipkin, B. and Proctor, R.J., eds., Engineering Geology Practice in Southern California (Star, 1992)

[124] Brillinger, D., 1990. Mapping aggregate birth data. In Singh, A. and Whitridge, P. (eds.), *Analysis of Data in Time. Proceedings of the 1989 International Symposium*, 77–83. Statistics Canada, Ottawa

[125] Brillinger, D., 1990. Spatial-temporal modelling of spatially aggregate birth data. *Survey Methodology* **16** 255–269

[126] Brillinger, D., 1990. A study of second-and-third-order spectral procedures and maximum likelihood in the identification of a bilinear system. *IEEE Trans. Acoustics, Speech, and Signal Processing* **38** 1238–1245

[127] Brillinger, D. R., Downing, K. H. and Glaeser, R. M., 1990. Some statistical aspects of low-dose electron imaging of crystals. *J. Statist. Plann. Inference* **25** 235–259

[128] Brillinger, D., 1991. Basic aspects of higher-order spectra and some of their uses. In Lacoume, J. (ed.), *Higher Order Statistics: Proceedings of the International Signal Processing Workshop on Higher Order Statistics (Chamrousse, France, 1991)*, 7–16. CEPHAG, Saint Martin d'Heres, France. Reprinted in Lacoume, J.L. (ed.), Higher Order Statistics (Elsevier, 1992)

[129] Brillinger, D. R., 1991. Some asymptotics of finite Fourier transforms of a stationary p-adic process. *J. Combin. Inform. System Sci.* **16** 155–169

[130] Brillinger, D. R., 1991. Some history of the study of higher-order moments and spectra. *Statist. Sinica* **1** 465–476

[131] Stoffer, D. S., 1991. Walsh-Fourier analysis and its statistical applications. *J. Amer. Statist. Assoc.* **86** 461–485. With comments by K. G. Beauchamp, David R. Brillinger and Pedro A. Morettin and a rejoinder by the author

[132] Brillinger, D., 1992. Locally weighted analysis of spatially aggregate birth data: uncertainty estimation and display. In *Spatial Issues in Statistics (Proceedings of Statistics Canada Symposium 91)*, 71–79. Statistics Canada, Ottawa

[133] Brillinger, D., 1992. Moments, cumulants and some applications to stationary random process. In Purdue, P. and Solomon, H. (eds.), *Proceedings of the Conference on Moments and Signal Processing, March, 1992 (Monterey, CA)*, 108–126. Stanford University Department of Statistics. This volume of proceedings is Technical Report No. 459, dated 1992-09-21, prepared under contract N00014-92-J-1264 (NR-042-267) for the Office of Naval Research

[134] Brillinger, D., 1992. Nerve cell spike train data analysis: a progression of technique. *J. Amer. Statist. Assoc.* **87** 260–271. 1991 R.A. Fisher Memorial Lecture

[135] Brillinger, D., Caines, P., Geweke, J., Parzen, E., Rosenblatt, M. and Taqqu, M. S. (eds.), 1992. *New directions in time series analysis. Part II (Minneapolis, MN, 1990)*, vol. 46 of *IMA Volumes in Mathematics and its Applications*. Springer-Verlag, New York, NY

[136] Brillinger, D. and Kaiser, R., 1992. Fourier and likelihood analysis in nmr spectroscopy. In Brillinger, D., Caines, P., Geweke, J., Parzen, E., Rosenblatt, M. and Taqqu, M. (eds.), *New Directions in Time Series Analysis, Part I*, vol. 45 of *The IMA Volumes in Mathematics and its Applications*, 41–64. Springer-Verlag, New York, NY

[137] Brillinger, D. R. and Kaiser, R., 1992. Fourier and likelihood analysis in NMR spectroscopy. In *New directions in time series analysis, Part I*, vol. 45 of *IMA Vol. Math. Appl.*, 41–64. Springer, New York

[138] Brillinger, D., 1993. An application of statistics to seismology: dispersion and modes. In Subba Rao, T. (ed.), *Developments in Time Series Analysis: In Honour of Maurice B. Priestley*, 331–340. Chapman & Hall/CRC Press, London

[139] Brillinger, D., 1993. Distribution of particle displacements via higher-order moment functions. *IEE Proc. F* **140** 390–394

[140] Brillinger, D., 1993. Earthquake risk and insurance. *Environmetrics* **4** 1–21

[141] Brillinger, D., 1993. Some examples of statistical analysis and computing in science. In *Computing Science and Statistics: Proceedings of the 25th Symposium of the Interface (San Diego, CA, 1993)*, 1–13. Interface Foundation, Fairfax Station, VA. Keynote talk

[142] Brillinger, D. R., 1993. Comparative aspects of the analysis of stationary time series, point processes and hybrids. *Resenhas* **1** 263–282

[143] Brillinger, D. R., 1993. The digital rainbow: some history and applications of numerical spectrum analysis. *Canad. J. Statist.* **21** 1–19

[144] Brillinger, D., 1994. Examples of scientific problems and data analyses in demography, neurophysiology and seismology. *J. Computational and Graphical Statistics* **3** 239–249

[145] Brillinger, D., 1994. Some basic aspects and uses of higher-order spectra. *Signal Processing* **36** 239–249

[146] Brillinger, D., 1994. Some river wavelets. *Environmetrics* **5** 211–220

[147] Brillinger, D., 1994. Trend analysis: time series and point process problems. *Environmetrics* **5** 1–19. Also in Proc. Stochastic and Statistical Methods in Hydrology and Environmental Engineering (Kluwer, 1994)

[148] Brillinger, D., 1994. Uses of cumulants in wavelet analysis. In Luk, F. T. (ed.), *Advanced Signal Processing: Algorithms, Architectures, and Implementations V*, vol. 2296 of *Proceedings of SPIE*, 2–18

[149] Brillinger, D. and Villa, A., 1994. Examples of the investigation of neural information processing by point process analysis. In Marmarelis, V. (ed.), *Advanced Methods of Physiological System Modeling*, vol. 3, 111–127. Plenum, New York, NY

[150] Brillinger, D. R., 1994. Time series, point processes, and hybrids. *Canad. J. Statist.* **22** 177–206

[151] Lange, N., Billard, L., Conquest, L., Ryan, L., Brillinger, D. and Greenhouse, J. (eds.), 1994. *Case studies in biometry.* Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley-Interscience, New York, NY

[152] Malik, F., Brillinger, D. and Vale, R., 1994. High-resolution tracking of microtubule motility driven by a single kinesin motor. *Proc. Natl. Acad. Sci. USA* **91** 4584–4588

[153] Brillinger, D., 1995. On a weather modification problem of Professor Neyman. *Prob. Math. Statistics* **15** 115–125

[154] Brillinger, D., 1995. Trend analysis: binary-valued and point cases. *Stochastic Hydrology and Hydraulics* **9** 207–213

[155] Brillinger, D., 1996. Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16** 1–23

[156] Brillinger, D. and Stewart, B., 1996. Elephant seal movements: some frequency based studies. *Brazilian J. Probability and Statistics* **10** 119–134

[157] Brillinger, D. R., 1996. An analysis of an ordinal-valued time series. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. II (1995)*, vol. 115 of *Lecture Notes in Statistics*, 73–87. Springer, New York, NY

[158] Brillinger, D. R., 1996. Some uses of cumulants in wavelet analysis. *J. Nonparametr. Statist.* **6** 93–114

[159] Bolt, B. and Brillinger, D., 1997. Maximum likelihood solution for layer parameters based on dynamic wave spectra. *Physics of the Earth and Planetary Interiors* **103** 337–342

[160] Brillinger, D., 1997. Earthquakes, statistics of. In Kotz, S., Banks, D. L. and Read, C. B. (eds.), *Encyclopedia of Statistical Science, Update*, vol. 1, 208–212. John Wiley and Sons, New York, NY

[161] Brillinger, D., 1997. Random process methods and environmental data: the 1996 hunter lecture. *Environmetrics* **8** 269–281

[162] Brillinger, D., Fernholz, L. and Morgenthaler, S. (eds.), 1997. *The practice of data analysis. Essays in honor of John W. Tukey. Papers from the symposium held in honor of John W. Tukey on his 80th birthday (Princeton, NJ, 1995)*. Princeton University Press, Princeton, NJ

[163] Brillinger, D. and Villa, A., 1997. Assessing connections in networks of biological neurons. In Brillinger, D., Fernholz, L. and Morgenthaler, S. (eds.), *The Practice of Data Analysis: Essays in Honor of John W. Tukey. Papers from the symposium held in honor of John W. Tukey on his 80th birthday (Princeton, NJ, 1995)*, 77–92. Princeton University Press, Princeton, NJ

[164] Brillinger, D. R., 1997. An application of statistics to meteorology: Estimation of motion. In Pollard, D., Torgersen, E. and Yang, G. L. (eds.), *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, 93–106. Springer, New York

[165] Brillinger, D. R., 1997. A particle migrating randomly on a sphere. *J. Theoret. Probab.* **10** 429–443

[166] Brillinger, D. R. and Stewart, B. S., 1997. Elephant seal movements: Dive types and their sequences. In Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G. and Wolfinger, R. D. (eds.), *Modelling Longitudinal and Spatially Correlated Data: Methods, applications, and future directions (Nantucket, MA, 1996)*, vol. 122 of *Lecture Notes in Statistics*, 275–288. Springer-Verlag, New York

[167] Gregoire, T. G., Brillinger, D. R., Diggle, P. J., Russek-Cohen, E., Warren, W. G. and Wolfinger, R. D. (eds.), 1997. *Modelling longitudinal and spatially correlated data: methods, applications, and future directions (Nantucket, MA, 1996.)*, vol. 122 of *Lecture Notes in Statistics*. Springer, New York, NY

[168] Barton, D. K., Brillinger, D., El-Shaarawi, A., McDaniel, P., Pollock, K. H. and Tuley, M. T., 1998. Final report of the haystack orbital debris data review panel. Tech. Rep. 4809, NASA

[169] Brillinger, D., 1998. Some wavelet analyses of point process data. In *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Pacific Grove, CA, 1997)*, vol. 2, 1087–1091. IEEE

[170] Brillinger, D. and Irizarry, R., 1998. An investigation of the second- and higher-order spectra of music. *Signal Processing* **65** 161–179

[171] Brillinger, D. and Stewart, B., 1998. Elephant seal movements: modelling migration. *Can. J. Stat.* **26** 431–443

[172] Brillinger, D., 1999. Some examples of empirical Fourier analysis in scientific problems. In Ghosh, S. (ed.), *Asymptotics, Nonparametrics, and Time Series: A Tribute to Madan Lal Puri*, vol. 158 of *Statistics, Textbooks and Monographs*, 1–36. Dekker, New York, NY

[173] Preisler, H., Brillinger, D., Ager, A. and Kie, J., 1999. Analysis of animal movement using telemetry and gis data. In *1999 Proceedings of the Section on Statistics and the Environment*, 100–105. American Statistical Association

[174] Brillinger, D., 2000. Some examples of random process environmental data analysis. In Rao, C. and Sen, P. (eds.), *Handbook of Statistics*, vol. 18, 33–56. Elsevier, Amsterdam

[175] Brillinger, D., Morettin, P., Irizarry, R. and Chiann, C., 2000. Some wavelet-based analyses of markov chain data. *Signal Processing* **80** 1607–1627

[176] Brillinger, D. R., 2000. Examining an irregularly sampled time series for whiteness. *Resenhas* **4** 423–431

[177] Brillinger, D., 2001. John Tukey and the correlation coefficient. In *Computing Science and Statistics: Proceedings of the 33rd Symposium of the Interface (Costa Mesa, CA, 2001)*, 204–218. Interface Foundation, Fairfax Station, VA

[178] Brillinger, D., 2001. Space debris: flux in a two dimensional orbit. In Fernholz, L. T., Morgenthaler, S. and Stahel, W. (eds.), *Statistics in Genetics and in the Environmental Sciences*, 105–116. Birkhauser Verlag, Basel

[179] Brillinger, D., 2001. Time series: General. In Smelser, N. and Baltes, P. (eds.), *International Encylopedia of the Social and Behavioral Sciences*, vol. 23, 15 724–15 731. Elsevier, Oxford

[180] Brillinger, D., Preisler, H. K., Ager, A. A. and Kie, J. G., 2001. The use of potential functions in modelling animal movement. In Saleh, A. (ed.), *Data Analysis from Statistical Foundations: A Festschrift in Honour of the 75th Birthday of D.A.S. Fraser*, 369–386. Nova Science Publishers, Huntington, NY

[181] Brillinger, D. and Yang, G., 2001. Lucien le cam, statisticien extraordinaire. *STATS* **31** 3–6

[182] Brillinger, D. R., 2001. Does anyone know when the correlation coefficient is useful? A study of the times of extreme river flows. *Technometrics* **43** 266–273

[183] Brillinger, D. R., 2001. Preface. In *Data analysis from statistical foundations*, ix–xi. Nova Sci. Publ., Huntington, NY

[184] Brillinger, D. R., 2001. *Time Series: Data Analysis and Theory*, vol. 36 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelpha, PA. This SIAM edition is an unabridged republication of the work first published by Holden Day, Inc., San Francisco, 1981

[185] Brillinger, D. R., Chiann, C., Irizarry, R. A. and Morettin, P. A., 2001. Automatic methods for generating seismic intensity maps. *J. Appl. Probab.* **38A** 188–201

[186] Preisler, H., Brillinger, D., Ager, A., Kie, J. and Akers, P., 2001. Stochastic differential equations: a tool for studying animal movement. In Rennolls, K. (ed.), *Forest Biometry, Modelling and Information Science (Greenwich, 2001).* These proceedings exist only in online form, rather than as a printed book

[187] Brillinger, D., 2002. Risk assessment, seismological. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 3, 1843–1848. John Wiley and Sons

[188] Brillinger, D., 2002. Some examples of risk analysis in environmental problems. In *Computing Science and Statistics: Proceedings of the 34th Symposium of the Interface (Montreal, QC, 2002)*, 3–24. Interface Foundation, Fairfax Station, VA. Keynote address

[189] Brillinger, D., 2002. Space debris risk. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 4, 2036–2040. John Wiley and Sons

[190] Brillinger, D., 2002. Waiting time. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 4, 2323–2324. John Wiley and Sons

[191] Brillinger, D., Guttorp, P. and Schoenberg, F., 2002. Point process: temporal. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 3, 1577–1582. John Wiley and Sons

[192] Brillinger, D. R., 2002. John W. Tukey: his life and professional contributions. *Ann. Statist.* **30** 1535–1575

[193] Brillinger, D. R., 2002. John W. Tukey's work on time series and spectrum analysis. *Ann. Statist.* **30** 1595–1618

[194] Brillinger, D. R., 2002. John Wilder Tukey (1915–2000). *Notices Amer. Math. Soc.* **49** 193–201

[195] Brillinger, D. R., 2002. Second-order moments and mutual information in the analysis of time series. In Chaubey, Y. (ed.), *Recent Advances in Statistical Methods (Montreal, QC, 2001)*, 64–76. Imperial College Press, London

[196] Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G. and Stewart, B. S., 2002. Employing stochastic differential equations to model wildlife motion. *Bull. Braz. Math. Soc. (N.S.)* **33** 385–408

[197] Guttorp, P., Brillinger, D. and Schoenberg, F., 2002. Point process: spatial. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 3, 1571–1573. John Wiley and Sons

[198] Schoenberg, F., Brillinger, D. and Guttorp, P., 2002. Point process: spatial-temporal. In El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), *Encyclopedia of Environmetrics*, vol. 3, 1573–1577. John Wiley and Sons

[199] Brillinger, D., 2003. Modelling and analysis of some random process data from neurophysiology. *Rev. Invest. Oper.* **24** 206–218. Includes Spanish summary

[200] Brillinger, D., 2003. Risk analysis: Examples and discussion (plenary talk). In der Kiureghian, A., Madanat, S. and Pestura, J. (eds.), *Applications of Statistics and Probability in Civil Engineering: Proceedings, 9th Conference, San Francisco, CA*, 3–11. Millpress, Rotterdam

[201] Brillinger, D. R., 2003. An analysis of a bivariate time series in which the components are sampled at different instants. In Moore, M., van Eeden, C. and Froda, S. (eds.), *Mathematical statistics and applications: Festschrift for Constance van Eeden*, vol. 42 of *IMS Lecture Notes Monograph Series*, 375–396. Institute of Mathematical Statistics, Beachwood, OH

[202] Brillinger, D. R., 2003. Simulating constrained animal motion using stochastic differential equations. In Athreya, K., Majumdar, M., Puri, M. and Waymire, E. (eds.), *Probability, Statistics and their Applications: Papers in Honor of Rabi Bhattacharya*, vol. 41 of *IMS Lecture Notes Monograph Series*, 35–48. Institute of Mathematics Statistics, Beachwood, OH

[203] Brillinger, D. R., 2003. Three environmental probabilistic risk problems. *Statist. Sci.* **18** 412–421

[204] Brillinger, D. R., Preisler, H. K. and Benoit, J. W., 2003. Risk assessment: A forest fire example. In Goldstein, D. and Speed, T. (eds.), *Statistics and Science: a Festschrift for Terry Speed*, vol. 40 of *IMS Lecture Notes Monograph Series*, 177–196. Institute of Mathematical Statistics, Beachwood, OH

[205] Brillinger, D., 2004. Everyone was an environmentalist. *Newsletter, ASA Section on Statistics and the Environment* **6** 2–3

[206] Brillinger, D., Preisler, H. and Naderi, H., 2004. Wild fire chances and probabilistic risk assessment. In *Proceedings of Accuracy 2004 and Ties (Portland, ME, 2004)*

[207] Brillinger, D. R., 2004. Some data analyses using mutual information. *Braz. J. Probab. Stat.* **18** 163–182

[208] Brillinger, D. R., Preisler, H. K., Ager, A. A. and Kie, J. G., 2004. An exploratory data analysis (EDA) of the paths of moving animals. *J. Statist. Plann. Inference* **122** 43–63

[209] Brillinger, D. R., Robinson, E. A. and Schoenberg, F. P. (eds.), 2004. *Time series analysis and applications to geophysical systems*, vol. 139 of *The IMA Volumes in Mathematics and its Applications*. Springer-Verlag, New York. Papers from the workshop held at the University of Minnesota, Minneapolis, MN, November 12–15, 2001

[210] Preisler, H., Brillinger, D., Burgan, R. and Benoit, J., 2004. Probability based models for estimation of wildfire risk. *Int. J. Wildland Fire* **13** 133–142

[211] Brillinger, D., 2005. Tukey, John Wilder. In *Encyclopedia of Statistical Sciences*. John Wiley and Sons

[212] Brillinger, D. R., 2005. The beginnings of some exploratory environmental risk analyses. Proceedings of the International Statistical Institute, CD

[213] Brillinger, D. R., 2005. A journey through random process data analysis: one type to the next. Video, ASA Committee to Film Distinguished Statisticians. Transcript at www.stat.berkeley.edu/brill/papers.html

[214] Brillinger, D., 2006. Space debris: yet another environmental problem. In Peck, R., Casella, G., Cobb, G. W., Hoerl, R., Nolan, D., Starbuch, R. and Stern, H. (eds.), *Statistics: A Guide to the Unknown*, 183–194. Thomson Brooks/Cole, Belmont, CA

[215] Brillinger, D. R., 2006. Anscombe, Francis John. In *Encyclopedia of Statistical Sciences*, vol. 1, 174–176. John Wiley and Sons

[216] Brillinger, D. R., 2006. Discussion: "Tukey's paper after 40 years" [Technometrics **48** (2006), no. 3, 319–325; mr2248365] by C. Mallows. *Technometrics* **48** 325–327

[217] Brillinger, D. R., 2006. The magic years: 1960-1964 at the Labs. discussion of C. Mallows *tukey's paper after 40 years. Technometrics* **48** 325

[218] Brillinger, D. R., Preisler, H. K. and Benoit, J. W., 2006. Probabilistic risk assessment for wildfires. *Environmetrics* **17** 623–633

[219] Brillinger, D. R., Stewart, B. S. and Littnan, C. S., 2006. A meandering hylje. In Liski, E. P., Isotalo, J., Puntanen, S. and Styan, G. (eds.), *Festschrift for Tarmo Pukkila on his 60th birthday*, 79–92. Dep. Math. Stat. Philos. Univ. Tampere, Tampere

[220] Brillinger, D., Penzien, J. and Romanowicz, B., 2007. Bruce alan bolt: 1930-2005. Memorial Tributes, National Academy of Engineering 31-35. See also Pp.103-107 in The 1755 Lisbon Earthquake: Revisited. Springer (2009)

[221] Brillinger, D. R., 2007. Learning a potential function from a trajectory. *Signal Processing Letters* **14** 867–870

[222] Brillinger, D. R., 2007. Modelling some Norwegian soccer data. In *Advances in statistical modeling and inference*, vol. 3 of *Ser. Biostat.*, 3–20. World Sci. Publ., Hackensack, NJ

[223] Brillinger, D. R., 2007. A potential function approach to the flow of play in soccer. *J. Quant. Anal. Sports* **3** Art. 3, 21 pp. (electronic)

[224] Brillinger, D. R., 2007. Volatility for point and marked point processes. Bulletin of the International Statistical Institute 56th Session Proceedings ISI 2007- Lisbon. CD

[225] Brillinger, D. R. and Guha, A., 2007. Mutual information in the frequency domain. *J. Statist. Plann. Inference* **137** 1076–1084

[226] Brillinger, D. R., 2008. The 2005 Neyman lecture: dynamic indeterminism in science. *Statist. Sci.* **23** 48–69

[227] Brillinger, D. R., 2008. Extending the volatility concept to point processes. *J. Statist. Plann. Inference* **138** 2607–2614

[228] Brillinger, D. R., 2008. Extreme event risk. In Melnick, E. and Everitt, B. (eds.), *Encyclopedia of Quantitative Risk Assessment and Analysis*, 676–677. John Wiley and Sons

[229] Brillinger, D. R., 2008. Modelling game outcomes of the Brazilian 2006 Series A championship as ordinal-valued. *Braz. J. Probab. Stat.* **22** 89–104

[230] Brillinger, D. R., 2008. Rejoinder: The 2005 Neyman lecture: dynamic indeterminism in science [mr2523939]. *Statist. Sci.* **23** 76–77

[231] Brillinger, D. R., Stewart, B. S. and Littnan, C. L., 2008. Three months journeying of a Hawaiian monk seal. In Nolan, D. and Speed, T. (eds.), *Probability and statistics: essays in honor of David A. Freedman*, vol. 2 of *Inst. Math. Stat. Collect.*, 246–264. Inst. Math. Statist., Beachwood, OH

[232] Brillinger, D. and Bolt, B., 2009. A statistical study of the seismic intensities of the 1755 lisbon earthquake. In *The 1755 Lisbon Earthquake: Revisited*, 455–469. Springer

[233] Brillinger, D. R., 2009. An analysis of Chinese Super League partial results. *Sci. China Ser. A* **52** 1139–1156. With a comment by Alan Lee and a reply by the author

[234] Brillinger, D. R., Autrey, B. S. and Cattaneo, M. D., 2009. Probabilistic risk modelling at the wildland urban interface: the 2003 cedar fire. *Environmetrics* **20** 607–620

[235] Brillinger, D. R. and Davis, R. A., 2009. A conversation with Murray Rosenblatt. *Statist. Sci.* **24** 116–140

[236] Brillinger, D. R., Lindsay, K. A. and Rosenberg, J. R., 2009. Combining frequency and time domain approaches to systems with multiple spike train input and output. *Biol. Cybernet.* **100** 459–474

[237] Brillinger, D. R., 2010. Modelling spatial trajectories. In Gelfand, A., Diggle, P., Fuentes, M. and Guttorp, P. (eds.), *Handbook of Spatial Statistics*, 463–475. CRC Press

[238] Brillinger, D. R., 2010. Some examples of the communication of risk and uncertainty. *Environmetrics* **21** 719727

[239] Brillinger, D. R. and Stewart, B. S., 2010. Stochastic modeling of particle movement with application to marine biology and oceanography. *Journal of Statistical Planning and Inference* **140** 3597–3607

Statistical Science Interview

A Conversation with David R. Brillinger

Victor M. Panaretos

Ecole Polytechnique Federale de Lausanne

Abstract. David Ross Brillinger was born on the 27th of October 1937, in Toronto, Canada. In 1955, he entered the University of Toronto, graduating with a B.A. with Honours in Pure Mathematics in 1959, while also serving as a Lieutenant in the Royal Canadian Naval Reserve. He was one of the five winners of the Putnam mathematical competition in 1958. He then went on to obtain his M.A. and Ph.D. in Mathematics at Princeton University, in 1960 and 1961, the latter under the guidance of John W. Tukey. During the period 1962-1964 he held halftime appointments as a Lecturer in Mathematics at Princeton, and a Member of Technical Staff at Bell Telephone Laboratories, Murray Hill, New Jersey. In 1964, he was appointed Lecturer and, two years later, Reader in Statistics at the London School of Economics. After spending a sabbatical year at Berkeley in 1967-68, he returned to become Professor of Statistics in 1970, and has been there ever since. During his 40 years (and counting) as a faculty member at Berkeley, he has supervised 40 doctoral theses. He has a record of academic and professional service and has received a number of honors and awards.

This conversation took place on September 9th 2009, in the Swiss Alps of Valais, during David's visit to give a doctoral course on "Modeling Random Trajectories" in the Swiss Doctoral School in Statistics and Applied Probability.

1. GROWING UP IN TORONTO

Victor: I suppose this is an interesting setting to be doing this, as one story would suggest you originally come not from very far from here...

David: Indeed! Now I don't know the specifics, but there were Brillingers in Basel at the end of 1400's. Once we were in Zurich, at Peter Buhlmann's invitation, and we saw a statue that was close: B-U-L-L-I-N-G-E-R. Now, the Brillingers in Basel became protestant at the time of Martin Luther. The next time I find them is in the 1700's when Brillingers went to Pennsylvania as Mennonites. They finally got up to Canada after the American Revolution. They were the original draft dodgers. You see then, in America, men had to be in the militia, but the Brillingers were pacifists. So they went to Ontario where they could practice their religion as they wished. So I'd like to think that there is some Swiss background and presumably it would have been through some great-great uncle who was "Rektor" of the University of Basel.

Victor: I see, I see, so it would then be Brillinger [German pronunciation] rather than Brillinger [French pronunciation]?

David: That's right. And you Victor told me that you've seen a truck on the Swiss highway with Brillinger on it. Also Alessandro [Villa] told me he saw a mailbox with Brillinger on it, or something like that.

Victor: Jumping much further into the future: you grew up in Canada.

David: Yes!

Victor: Could you tell us a bit about your family?

Victor M. Panaretos is Assistant Professor of Mathematical Statistics, Institut de Mathématiques, Ecole Polytechnique Federale de Lausanne. EPFL-IMA-SMAT Station 8, 1015 Switzerland (e-mail: victor.panaretos@epfl.ch).

imsart-sts ver. 2009/08/13 file: BrillingerPanaretos_Conversation.tex date: March 15, 2010

2

David: My father died –let's just work it out– when I was 7 months old, so this was very harsh on my mother. She woke up in the middle of the night and he seemed to be in some trouble, but then she fell back asleep and I think she felt guilty about that ever after. I doubt there was anything that could have been done back then because he died of a cerebral hemorrhage. I wish I could have gotten to know them together better. You know, they had their house, a cottage, a dog and so on. They had a Harley motorcycle and went off on that on their honeymoon, they had a sailing canoe... Lakes, and Canadian things were very much part of their lives. My mother was actually a very beautiful woman, when you see the pictures, with smiles [Figure 1]. But the smiles mostly disappeared after my father's death. Then, it was World War II times and most of the men were gone. It's hard for me to imagine she wouldn't have remarried. But it just never happened.

She really cared a great deal about my education and structured things so that I got a fine education. At the start, there was a bit of money –because my father was going to be an actuary, so she had some insurance money. I went to a private boys' school in Toronto until the money ran out. Then, there was this school for bright kids in Toronto, the University of Toronto Schools (UTS). I took the exam and got into it. UTS was very important for me. I should mention that my maternal grandmother was also very important, and perhaps she raised me. She had had her husband die in the great flu epidemic and found herself with five children to raise. So I had, I think, a beginning that made me appreciate being alive and not really expecting too much to come from it. I really have been pretty content and non-aggressive about things in my life and feel very lucky. You know, all four of my uncles –and I've decided they were my role-models– were taxi cab drivers at some point in their lives. The way they could just talk to anybody and the way they engaged people, to some extent formulated the way I have become. I had a lot of paying jobs as I was growing up including caddying, delivering prescriptions, salesperson in a small shop.

I had a lot of cousins that were important to me because I didn't have siblings. And there were a lot of wonderful mother's side family gatherings. So, I don't think I really thought about not having a father when young, but I do wish I could have asked my father certain questions since we did not have much contact with the Brillinger side of the family. That was a shame.

Victor: Did you have any influential teachers at school?

David: Oh, yes! There is one very influential teacher who taught me when I was at Upper Canada College –that was the private boy's school. I had not started the year there and when I transferred, he found out that I was not very good at fractions. So, he spent some time tutoring me. Now he was also an important person in Ontario hockey. And after tutoring me he came in the class one day and said he had 5 hockey rulebooks and he was going to give one of them to whoever answered a mathematical problem first. So first question, my hand went up, one rulebook; second question, second rulebook; third question, third rulebook! So he said "David that's it, you can't get anymore of those"! I really learned I was good at sports. Or no, actually, I wasn't good at sports, I was good at math, but I was very motivated when it came to sports! [laughs]. The teacher's name was H. Earl Elliott.

Victor: And those were the same rulebook?

David: [laughs] Oh yes! I don't know what I was going to do with all of them! He had not specified any rules, so I had three and gave my cousins two! I had realised I was good at math, and I loved working on math problems. A lot of books had problems without the solutions in the back. I had a lot of fun doing them. Perhaps I had more time to do that because the weather was bad in the winter and I did not have siblings. Afterwards, I went to UTS. I said that was for bright kids, but part of the definition of "bright kids" then was being male [both laugh]... Luckily things changed, although UTS no longer wins the Toronto high school hockey championship like it used to! I had a very influential mathematics teacher there, Bruce McLean. He was also the hockey coach and is still alive. He would just let me work at the back of the room on my own. Everybody else was up towards the front, but he would just leave me alone at this table and bring these books full of problems, [e.g. Loney (1930)]. Statistics was one of the topics. And there were these British problems that you've probably seen in the Tripos,

FIG. 1. *Young David in his mother's arms at the King and Queens visit to Toronto, as a Cub Scout, and with his ski gear.*

Victor, things like that. I don't know about what level I would have been at had I been in England, because students there started working with these concepts very early on. I read a book where I think Dyson said he had solved all the problems in Piaggio's differential equation book [Piaggio (1920)], but when he was at public school—I did that when I got to University, so I guess I was lagging behind. But I think I was very independently driven to work on these things. I thought I solved them, but, you know, I didn't quite know; but anyway, I solved them to my satisfaction. Then, Ontario used to have some pretty tough High School exams, for the last year—grade 13—and four of them were on algebra, geometry, trigonometry and problems respectively. I got 100, 99, and 100 on the first three and 96 on the last. I still think about that 96. You see you were to do 10 problems, but there were 12. So I "solved" all 12. Later "Mr." McLean told me that the person who was grading kept getting a total of 118 on my exam, and he could not figure out what was going on for a while. Eventually, he realized that I had attempted all 12. My error was that one of them was finding the maximum or minimum of something, so to show off I used calculus, but I forgot about checking the second derivative! I've never forgotten that since! [both laugh] But anyway, that brought me a scholarship that helped me make my way at University. Back then, prizes were important because there weren't many bursaries. Now, in America, they've switched to means tests. But I won a lot of prizes as an undergraduate which kept my mother and me with food and so on.

Victor: Evidently, mathematics was one thing you enjoyed, but what about sports?

David: I love sports, I always have and I have always been a Toronto Maple Leafs fan. I don't know if I still have it, but there was a wonderful picture of me about 3 years old with hockey stick in hand and skates on feet. I was often the last guy to make the team or the first guy not to make the team—but I was always there! When I was growing up, they would flood the whole neighborhood park so there would be 5 or more hockey games going on. You didn't need all this fancy equipment. I guess I could make the formal teams until I was 13 or so, but then that stopped. It returned for a while when I went to Princeton as a graduate student. There I got to be like an intramural star, because I could raise the puck, knew the rules and played left-handed. Now, I mentioned my high school teacher, Bruce McLean. There's a story I love concerning him: there was my 50th High School reunion a couple of years back and I was in Edmonton the week before the reunion and was going to need to be in Toronto the week after, so it was just too much time to be away from Berkeley. One of my dear friends from High School and University, John Gardner [now Chair of the Board of Directors of the Fields Mathematical Institute], asked if I'd like him to arrange a lunch with "Nails" McLean—his nickname for UTS students was "Nails". I said of course! So, when I went to Toronto the week after, we had lunch. McLean was 96, and had driven in through all the traffic to central Toronto for the lunch. We had a wonderful time. It turned out he had also been in the Navy, so we discussed that.

4

But at the end of the meal he got this incredibly serious look on his face. So I'm thinking "What's this all about?" And he says "David, when you were at school, there was something I really worried about, I worried about it for a long time". So I'm sitting there with my eyes rolled back and wondering. He continued "I really wanted you on the hockey team, but there were a lot of good players that year!" [both laugh]. I just grin when I remember that. And indeed the team was good. They won the Toronto championship. I just wanted to get the sweater, go to practice, and, if we're winning 7-2, get to skate around a bit. But I had to wait until Princeton to do that.

2. UNIVERSITY OF TORONTO AND THE CANADIAN NAVY

Victor: You mentioned before that you were in the Navy, can you tell us a bit more about that?

David: That was at University. I knew that by joining the Navy I was going to get to go outside of Toronto and perhaps Canada for a bit; because Toronto was really a bit boring back then. Canada did not have a draft –still doesn't– so the way the government thought they could get officers for the regular military was by having army, navy and air force programs at the universities. That was a bit like Boy Scouts, and I'd been a Cub [Figure 1], and a Boy Scout. For me, it was obvious to join the Navy because I loved to canoe and sail, and you got to go to Europe and Mexico. Whereas if you were in the Army you got to march around in the dust of Ontario; and if you were in the Air force, you were in Saskatchewan, which is flat, and with not so much to do then. So, I was on my way to seeing the world and at the same time got paid very well, the food and the clothing were obviously provided. Plus it was a lot of fun, I just loved it. I mean guns were only 5% or less of the life. So it was a no-brainer to be in the Navy. Second year I was based on the West Coast [Figure 2]. In the program there was a prize for the person who was best in navigation and I think I won probably easily as I had taken an astronomy course and had learned all this spherical trigonometry previously. The way things worked I ended up being a communications officer learning about radio and coding. This was great since I had been learning physics as well as mathematics. You know, in my career I've gotten to study mostly the things I was good at and enjoyed. I was principally good at math, and it was obvious what my career was to be.

Victor: You once told me a story about doing some very applied statistics in the Navy.

David: That was my first independent statistical research activity, I would say! So let's think. My fourth summer, I had already gone through a lot of basic training, becoming a communications specialist and a sub lieutenant. I was going to be in the aircraft carrier, the Bonnaventure, and we were supposed to sail into the middle of Atlantic because the Queen was going to fly over there on her way to visit Canada. And so we were to be stationed out there. I don't know why, maybe in case she leapt out with a parachute or something like that! I mean it was awfully ill-defined! [both laugh]

Victor: ...after all it is the *Royal* Canadian Navy!

David: Exactly! So we had to toast to the Queen at banquets and such and such. Anyway, they had to find something for me to do during the open period before the mission. So, they decided that, since I was studying statistics, they would like to know how many messages were sent out by the fleet weekly for several years. They took me to this room, and here were these huge stacks of signals by week. I would still be counting them if I had done it directly! But instead I thought why don't I just get 100 and weigh them and estimate a weight per signal. And then I asked for a scale, which they found. And I just measured how heavy the piles were, and so I gave them nice graphs. When the fleet was at sea, there were a lot more signals, and things like that. I guess it sounds nutty to be saying the following, I mean I'm totally a pacifist and I think I've been that all my life – but I did enjoy the Navy! I suppose back then Canada was doing peace keeping. Like Brazil's these days that was the Canadian role then. Our Prime Minister Lester Pearson won the Nobel peace prize for the idea of creating a UN Peace Force. My thought was that the world needs policemen, and since Canada was not in an aggressive posture at that point, I signed up. By the way, in the remaining time before the cruise, I did a lot of dinghy sailing in Halifax harbor.

Victor: Shall we talk a bit about the University of Toronto (U of T)? You did your bachelors honours in pure mathematics. I recall you telling me in Berkeley

FIG 2. *David in the Navy off Santa Barbara in 1957, and upon graduation from the University of Toronto in 1959.*

that you were already reading Bourbaki as a first year undergraduate— in French.

David: Yes, that's true! I was lucky because Canada was trying to be bilingual to support its francophones and I studied french for seven years. So there was a professor at U of T, John Coleman—who is still alive, aged a hundred or so I think; these Canadian mathematicians live a long time. He found out I could learn and read in French. I think he identified me especially because I had won this prize for algebra/geometry/trigonometry and problems. He found what I looked like by watching where my homework handed back ended up in the classroom. He invited me for a coffee or whatever. Actually, he was remembering when I talked to him a couple of years ago that we had butter tarts and tea when we met. He got me reading Bourbaki. And then he said why don't you do some of these problems? So we met the each week; I couldn't do the problems, and perhaps he had trouble too. I don't know if I could do them now, it would be fun to try. The first book was on algebra and I believe that Coleman bought it for me. I still have it [Bourbaki (1951)]. The later ones on analysis have probably been the most important to me. Coleman got me reading Bourbaki and I remain very appreciative. Going through them really stood me in good stead when I got to Princeton. I found myself a couple of years ahead of the American students. You see I'd gotten to do mainly maths and physics at Toronto, and I also had this secret weapon: French! I mean the French probabilists were then doing all this wonderful stuff, E. Borel, P. Lévy and M. Fréchet for example. And most of their things were not being translated. Nowadays the French mathematicians write in English most of the time so that's not an issue. That was first year. That year I also had a course from Ralph Wormleighton, he had been at Princeton—there was a real Toronto-Princeton railroad including Don Fraser, Art Dempster, Ralph Wormleighton; and when I applied to grad school I only applied to Princeton. It never occurred to me to apply anywhere else. I don't think that was a statement of confidence, but I didn't have anyone who had been at university at home, so I just was not getting that kind of advice. The second year was Dempster. Dempster has often taken the geometric approach. When I took a course from Coxter, I later saw where that approach was coming from. And then in the third year was Don Fraser—he was certainly using a lot of algebra. The fourth year was Dan DeLury. He was this skeptical older guy. He'd been out doing biometrical studies. His attitude was that one might have thought that they had designed an experiment well, but there were many ways that

6

an experiment might have gone wrong. His course was very maturing for me. It's important to have some training in criticism when you're an applied statistician.

Victor: So, that means that you would have had quite a rigorous maths background but also would have been exposed to quite a bit of statistics, which is rather atypical for that time period.

David: Although I was in pure mathematics –that's what my degree was in– I went to all the statistics courses. As a matter of fact, I probably went to *all* the courses, including the actuarial ones. Back then, I could just sit there and absorb things. It's not as though I'm boasting; I used to feel embarrassed about saying things like that, but I think I was just lucky: it was not really anything I did, it's just the way it was. I wish I could have played hockey better, but I didn't get that skill nor the ability to run 100 meters in less than 10 seconds. I guess I'm saying there may be a gene that I was lucky enough to get.

Victor: Do you recall any lectures that you particularly enjoyed? Coxeter had a fine reputation as a lecturer I suppose.

David: Oh yes, Coxeter was wonderful. He had left England after World War II. Also Tutte, who is another geometer, was great. In fact, Tutte had broken one of the important Nazi codes in World War II – and none of us knew that. But some people in the class were mean to him because he was a little shy, and they teased him. I'm sure if they had known about his breaking the code, they would have been more like "wow" instead. Regarding Coxeter, I remember one funny story, where he was talking about a particular geometry for many classes. His course became his book [Coxeter (1961)] or the book was part of his course. So, there was this particular finite geometry he was talking about a lot, with very bare assumptions and he was talking about it during a number of classes. So, finally, I asked "Why are you spending so much time on this, is it *that* important?" And he said something like: "Well you seemed so interested, Mr. Brillinger!" I mean, I was just asking questions to keep up with where he was going! I was intending to become an actuary for many years, in part because my father worked for Imperial Life. And they were very good to my mother and me. I had realized that if you are poor but good at mathematics then an actuarial career was a route to the middle class. I'm not sure I was after being middle class, but I needed to help my mother, so I was going to be an actuary. But Don Fraser, who had great influence on me, said something like: "Well David, sure that's nice, that you're going to be an actuary, but why don't you go to Princeton first?" So, I did! I went to Princeton, the plan being to become an actuary after I was done with all this childish fun, namely mathematics.

Victor: Apparently it was *too* much fun...!

David: I guess that's right. And I realized at some point that anything I could do as an actuary, I could probably do as a statistician – with the added benefit that I would get to travel and be an academic. I did take enough of the exams to become an Associate of the Society of Actuaries.

Victor: Just before going off to Princeton, you were among the winning 5 of the Putnam competition of Spring '58.

David: It was again Coleman, who got me involved.

Victor: And I recognized a couple of other famous names on the same honours list, Richard Dudley and Larry Shepp.

David: Yes, I got to know them both. You see, both of them went to Princeton for graduate studies. I really had no idea of what was involved. I just went and took the exam! I remember that Erdős visited Toronto for a month and he gave a course. One of the problems he taught us was on the Putnam exam! [laughs] Some number theory thing [continues laughing]... So on the exam day that one was out of the way pretty quickly! He was just a real gem, a real role model. I mean he had these simple direct ways to approach problems, and would advocate that you should take a breath before you start writing down a lot of equations and things like that. U of T was absolutely super. I got a super education in mathematics there and at high school. I mean some people might think of Canada as being a backwater, or as having been one, but there were some very fine researchers and teachers. You know, Coleman had also gone to Princeton just before the War started. I was lucky.

I can't resist adding that, while I was at U of T, I was actually at Victoria University. There, I earned a letter for playing on the soccer and squash teams, each for four years. I can show the letter to you!

FIG 3. *David with Bruce "Nails" McLean and with Don Fraser.*

3. PRINCETON

Victor: When did you move to New Jersey?

David: In the summer of '59. That was my last summer in the Navy, and I had become a Lieutenant. I turned up there in the beginning of August having left the Bonnaventure. I had asked if there was some work for me, and it turned out that Sam Wilks had just finished writing his book Mathematical Statistics [Wilks (1963)]. My job was to work on the problems. I remember I just lay out under the trees at Graduate School working on them, right by the golf course which I would golf on most days, illegally. I remember going over to Wilks' office just before term started. One of my Canadian friends, Irwin Guttman, was there. I said "Well here are the solutions, but I couldn't get one of them". And Wilks went "What???". In the end he took that problem out of the book. It was about proving that the median and the mean were jointly asymptotically normal. It took me a while to figure out a neat way to do that.

Victor: You got right into mathematical statistics upon arriving at Princeton.

David: Oh yes. Already at Toronto, I could see that statistics, perhaps as an actuary, was for me, because you interact with people a lot. Math was a lot of fun too, but you interact with a much narrower group of people. DeLury had impressed me, because he was really working at the frontier of the applications of statistics. I have found myself realizing that statisticians are the keepers of the scientific method. When a scientist comes up with something, what can they reasonably conclude? That appealed to me, to be able to get involved in many fields.

Victor: And when did you meet Tukey?

David: [laughs] Aaaaah, John Tukey... I watched him like a hawk! Because he was so interesting generally and so much fun to watch. I had been told about Tukey by Coleman. Coleman had been a graduate student when Tukey was at Princeton. And Coleman told me that I was going to meet someone who, at beer parties, was always drinking milk, he just had a big glass of milk. So I knew before meeting him that Tukey was different. Because at a beer party in Canada you drink beer, that's part of your manhood, or something like that. Princeton; at Princeton you didn't have to take any courses. You could sign up for one and would get an A, even if you never turned up. You had to write a thesis and pass an oral exam, so that was pretty good! So let's see, Tukey gave a time series course. And here was this person, unlike any other person I had ever met. He was from New England, very Canadian in a lot of ways. He had pride in his background. He was careful with money, and he had apple pie for breakfast. So I went to his time series course and this involved a lot of Fourier analysis – and I had a strong background in trigonometry and that made the course attractive.

Victor: Did you attend any of these courses along with David Freedman?

David: Oh yes! David F. was a year ahead of me, and he was influential on me [pauses and reflects for a moment]. I guess, oh my, most of these people are dead now, goodness. OK, whatever. I have these two stories about David, one involving Frank Anscombe and the other John Tukey. Now, David was a year ahead of me at Princeton. He was from Montreal, I was from Toronto so we were natural "rivals", right from the beginning! That's just the way it was. Of course I don't mean that in a bad way. Anyway, Frank had asked David F. to be his teaching assistant in a course. And David said: but I am on a scholarship, I don't have to do that! "OK, fine.", said Frank, and then Frank asked me [laughs]. And I knew what David had said, and got to give the same answer! David analyzed a lot of situations very clearly, and I observed David as I do a lot of people.

David F. never changed in terms of his intellectual calibre and wit, and the character of his questions. David was also in Tukey's time series course. Early in the term Tukey used the word spectrum several times. And David after, I don't know, 20 minutes or some such, asked what the definition of a spectrum was. So, Tukey said something like: "Well, suppose you've got a radar transmitting signals up and it bounces off an airplane and a signal returns ... so you see... well that's a spectrum". So, David's manner was "Well, ok." Then the next class the same thing happened. Tukey mentioned the spectrum, David wanted a definition, and Tukey said: "Well, suppose you have a sonar system and it bounces a signal off a submarine, or some such"... David never came back! [both laugh]

That was really pure David F., wanting clear explicit definitions. Tukey and David were the opposites of each other. You see, Tukey believed in vague concepts. He believed that if you tried to define something too precisely, then you would have lost important aspects going along with it. But David didn't think that you could talk about things properly unless you were completely clear. Of course, Tukey's and David's great confrontation was over census adjustment. I picture that David took a strict interpretation over what was required while JWT was after an effective estimate of the counts. It is no surprise that David was debating champion at McGill. He surely could have been a fine lawyer, and then a judge, and then ...

Victor: He did get involved with statistics and the law.

David: Yes, he was involved in statistics and economics, too. He worked at the Bank of Canada for a while. I think he might have expected that he would be going down that road. He probably thought that being a statistician you can do anything you want to – that was my own reason for choosing statistics.

David was a very sweet person. I am thinking just now of his taking Lorie and me out to dinner in a nice Princeton restaurant after we got back from our honeymoon.

Victor: Going back to Tukey, what did you learn from him as a researcher, what was his style?

David: I learned that there are novel ways to solve most problems. I think JWT could add two four-digit numbers in ten different ways that no one else in human history would ever have thought of! I mean he was like Richard Feynman. He was of the same ilk. There are people, and there are lots of historical examples, who just think differently than almost everyone else. Also what I have learned from Tukey is that there is a physical interpretation of so many of these concepts when you look at the history of mathematics. That's what I tried to bring up in my talk this morning about how some of these things came out of Kepler and Lagrange and so on [David was lecturing on SDE modeling of random trajectories using potential functions]. That you can understand a lot of this contemporary work if you think about how it had been generated in the first place. I think Tukey often found himself explaining things to people who didn't know much mathematics. I paid attention to how he did that. I would like to think that I'm not bad at doing that too. In a sense, you

FIG. 4. *David with John Tukey at the NBC Election Centre in 1962.*

probably lie a bit. I mean you probably use an analogy or a metaphor at some point, which is not quite right, but people get the idea.

Victor: That's the advantage of vagueness.

David: Yes, indeed! Tukey's vagueness meant for example we that could start out with standard errors and later find ourselves talking about the interquartile range, just letting the idea of "spread" be vague.

Victor: What was your relationship like when he became your advisor?

David: Re advising: there were lots of good problems around Fine Hall and the Labs that I worked on. Eventually, JWT suggested a particular one. The deal seemed to be that if I started to have trouble, I should go see him. Maybe his not being around town often was part of the breaks in our meetings. When I would meet him, if I seemed a bit too cocky he would knock me down; and if I looked discouraged he would build me up. My thesis concerned formalizing Gauss's delta method by working with truncated random variables asymptotically. Another thing was that during the school year I had the day-a-week job at Bell Labs, so often I drove back and forth to Bell Labs with him, sometimes in his convertible. During those drives, we talked about a lot of things. Sometimes, there were other passengers too. I learned while working with him that, when he used some new word, I shouldn't worry about it. I should just let him talk a while and then try to figure out what it was all about. I think a lot of people had a hard time understanding what he was trying to get at. I would eventually come up with something; now if it's really what he meant, I don't know. I'd say I had a wonderful relationship with him. I would kid him – I mean I didn't know you shouldn't tease professors until much later! Because I was working class Canadian and had my uncles as role models. That's how they'd approach people. Not mean teasing, just seeking a smile. I have also teased David Cox. David was patient with me.

Victor: There was good chemistry between you, then. Because, you know, he was relatively conservative and you've been pretty progressive and open about it all along.

David: There was, yes sure. We could talk about things just like that. No tension. He was on the conservative side, true. But it was more about different cultures. He was American and I am Canadian. Canadians are progressively conservative. In those days, there was a conservative spirit in Canadians when it comes to the way one dresses or the way you talk to other people. So, there was conservatism

10

in me. but it was social conservatism. not political conservatism.

Victor: Well, it would appear that Tukey had a very high opinion of you. It has been rumoured that he used a "milli-Brillingers" scale to measure people up?

David: [laughs] Yes, I have heard that from several people, including Mike Godfrey and Bill Williams, but what does one say? Bill told me that once Tukey asked about a prospective student, "How many milli-Brillingers?". Bill's reply was "four or five hundred mB's". John responded with something like, "Well that's very good." I don't know, I guess that I was quick on my feet, I don't mean at running. If I had to do something I would go and do it.

Victor: What about Sam Wilks whom you just mentioned earlier?

David: Sam was wonderful too. He was just a gem. It's a shame that he died way too soon. One story is that he was taking shingles medicine and drank some alcohol that night and there was a bad synergy. Another is that there was an unpleasant meeting over the admission of a student to the program. Sam was conservative politically, but that was never an issue. He had me work on these problems in the draft of his book as I mentioned. I also sat in on the course that was based on the book he was writing. He was a social animal. I can tell you one story. The Tukeys – God knows for what reason – had decided to have a come-as-your-spouse party. So Lorie was supposed to dress like me and I like Lorie, and so on and so forth, Mrs. Tukey like John Tukey, and John Tukey like Mrs. Tukey. That happened, but Gena and Sam Wilks came along as themselves! Near the end of my studying, I went off for an interview at the University of Michigan, before I knew whether I would receive a postdoc. Jimmy Savage was there then. I told him about the party. And I think he went like this [David holding his chin down] and said: "I know too much Freud to ever do something like that!" I didn't know a lot about Freud and I still don't know what Savage meant, but he did know a great deal about many of things.

Victor: So how did you meet Lorie?

David: Blind date! and we're both proud of that! One has to take risks sometimes. She went to Antioch College with its work-study program. She was studying sociology and had taken a statistics course using Mood and Graybill – not an easy book. She was in Princeton in the "work" component at the commercial side of the Gallup Poll. The Riehms introduced us. Carl was in mathematics, eventually becoming a professor at McMaster University, and Elaine was also working at Gallup. I think hers and Lorie's desks were next to each other. The Riehms were often trying to get Lorie and me together, but Elaine kept complaining because I was always out of town! I went back to Toronto a lot – no course responsibilities, remember? Lorie was attractive and we found lots of things to talk about. Anyway, it was a blind date. And, I don't know, we just hit it off quickly! One thing that I loved about Lorie was that she was very political –my politics weren't well formed at all yet– and she was also very analytical. Her parents even more so! Later, we realized that we each had a parent who had been born in China, the child of Methodist missionaries.

Victor: What a coincidence!

David: Oh yes! They were, in fact, in the same part of China: Sichuan province. And now with the web, you can find surprising things. So, I entered my Brillinger grandfather's name and her Yard grandfather's name, into Google, and then found them in the same book! [Bondfield (1912)]. Lorie's grandfather was in an American missionary and my grandfather was a Canadian medical missionary. Her parents were very political and they had a huge wealth of political literature. Probably like the literature you, Victor, grew up with. I was a bit shy with them, and since they had all these magazines and books on the coffee table, I could always check something out while I was listening. So, there was a very political side to it all, too. Anyway, we fell in love and it's been good. Almost 50 years now! People often say about us that we don't need to talk, that we just simply communicate. Lorie changed her career goals quite drastically after meeting me. If she had returned to Antioch College then I would have gone to Yellow Springs with her, probably to teach statistics. But in the meantime, I completed my PhD and had applied for a post-doctoral fellowship at London, which I was awarded. Lorie decided she preferred to go to London. She was actually studying British Trade Unions at Oxford when I asked her to marry me, so she got back to England quite quickly.

Victor: Indeed you really dashed through your PhD

FIG 5. *David in his Princeton PhD Regalia in 1961.*

in less than two years! How did that work? Did the lack of coursework requirements have anything to do with that?

David: I don't think so.

Victor: I guess that your "milli-Brillingers" had!

David: [laughs] Aaaah, I don't know, I guess Tukey gave me a problem, and said "see what you can do with it". So, I graduated that following May. Why didn't he give me something like Fermat's last theorem, I don't know! But I actually had a try at proving that in high school. I read a lot of the history of mathematics.

Victor: I suppose nowadays in Berkeley, as well as many other US universities, there is quite a bit of structure with a lot of coursework and exams. How do you compare those two different systems?

David: Well Freedman and I talked about that once. And we agreed that we would not have gone to Berkeley, which is pathetic. But that's the system. Plus Princeton was very selective when I went there, I think, two statisticians admitted each year.

Victor: David Cox once told me that the less structured approach is appropriate for the very brightest of students.

David: Yes, I think so, but I certainly don't claim to be a member of that group.

Victor: What do you think happened with the Princeton group?

David: From hearsay, I think I can make a reasoned guess. Tukey was a dominating figure. I know he had tremendous respect for Sam Wilks, but I'm not sure about some of the other people there. Also, he had the mathematicians to contend with. Yet, he needed people. He asked Don Fraser various times to go to Princeton, he asked Art Dempster various times, he asked me several times. Clearly, I can only speak for myself. I just wanted to do some things that were mine. It sounds selfish, but Tukey was so dominant and so quick. I don't think that he thought any less of me because I refused. A lot of people were afraid of him. For example, if they had a cockeyed idea, he didn't mince words. He told me once that he thought the best way to get a scientific discussion going on something was to start an argument. Now that's just the reverse of my personality. I did see him do a lot of that. It was possibly he wanted to get beyond the early pleasantries that go on. He did run over quite a number of people. He liked to argue and expected to win. I think that he wanted to win because he had a goal and wanted to get there quickly. I did love interacting with him during my thesis research. I found I could communicate very easily with him. But still, I felt a need to do my own thing. Princeton did get a viable group at one point, and it became a department. The members included Geoff Watson, Peter Bloomfield and Don McNeil. They each had a definite presence in the statistics world. However, I think that Peter Bloomfield just got fed up with being Department Chair. So he went off to a large department at North Carolina State. And McNeil went back to Australia. Also, I gather that Watson was treated quite terribly by the Mathematics Department. I was very sad when Geoff died for he had spoken truth many times. Eventually, Tukey was the only senior person left and when he retired the department went away. So, it is a sad story, but part of Princeton's strength in statistics was that the people it was producing for many years came through mathematics, so there was no messing with them in terms of mathematical stuff, but yet these people wanted to apply mathematics as opposed to doing research in some mathematical specialty. To deviate from the present topic slightly, I have long found classical applied mathematics a bit boring and old-fashioned, but I do know that Fisher, [Fisher (1925)] wrote that "statistics is essentially a branch of Applied Mathematics". Nowadays, one might say that statistics is a combination of applied mathematics and applied computing, the two driving the field. A Princeton review committee was set up, and recommended against continuing the Statistics Department, and that was that. But I did have a lot of fun at Princeton.

4. BELL LABS

Victor: Could you please tell us a bit about your summers at Bell Labs?

David: The first summer in grad school, there was a group of us from Princeton that had summer jobs at Bell Labs. I would drive up there with my friend Carl Riehm, an engineer and a logician. I don't know if the Labs had this program to find future employees or if it was just a good deed for science. I had learned some computing at Toronto on their IBM 650. Toronto had these computing services very early on. for example, they had a Feranti from the mid '50's. So, I had started out learning computing in a course in the physics department. This was before Fortran existed, so we were using machine language. Princeton had a 650 also, which I didn't really use that much – I guess I was a lot more interested in group theory then. But when I went to my summer job at Bell Labs, they had an IBM 701. Fortran got created and so they had me programming various things for Tukey. That was pretty much the story during my first summer; it was nice to make the money. Then, the second summer... Let's think... I guess the second summer Lorie had appeared on the scene! So, we had a lot of fun. I think that's when Tukey had me writing some programs involved in discriminating earthquakes from underground explosions. He was then involved in the Geneva negotiations for a nuclear test ban treaty with the Russians. Tukey had one of those out of the box ideas, the cepstrum. He thought this might also work for pitch detection. That's what I was doing. Specifically, taking speech signal, digitizing it, doing things to it on the computer, then reconstituting it and listening to it. Really, the spectrum and a lot of these time series things had a real meaning for me at that point. I also golfed a lot. The Labs had a short 3 hole course.

Victor: You got experience with getting your hands dirty with data.

David: Oh yes, right away. I really loved that. But, more importantly, I got exposed to a whole cast of characters creating exploratory data analysis! John Tukey, was the leader, obviously. But there were others right up there with him, Martin Wilk, in particular – he wrote some important papers with John. There were also Roger Pinkam, Bill Williams my buddy, Dick Hamming, Ram Gnanadesikan, Colin Mallows who had a strong influence on me. I was in an office with Colin so that was enjoyable and educational. And lunch was where I became a statistician, really. The whole group of us would go down to the cafeteria and sit around a big circular table. So, lunch was about this communal group trying to help each other with their scientific and statistical problems. Then, people would go back to their offices and do their own things. I mean the old Bell Labs worked wonderfully and it's just pathetic that it went away. There was an open door policy and everybody shared the problem they were working on. We had a lot of fun playing pranks up there, too. You know, it was all a gentler world back then in the early 60's. It had an incredible influence on my becoming a statistician because really they were creating a lot of applied statistics. I was very lucky. I mean I got onto a pretty good escalator going up. You don't realize at that time how special it all is scientifically and socially. When I've talked to some of the other Bell Labs people we've all said "Those were magic years.", and that we were so lucky to be right in the middle of them. Bell Labs was clearly years ahead of people in digital signal processing. Tukey coming up with the Fast Fourier Transform was just part of it. He was working on EDA methods too...

Victor: Did you "witness" the FFT being developed?

David: Tukey's form, yes. In his time series course, John had some way of doing it by complex demodulation. Filtering this and filtering that and then putting things together. But one day in '63, he turned up at a class with an iterative algebraic approach to computing the discrete Fourier transform for the case when one could factor the number of observations into a product of two integers [Tukey (1963)]. It turned out that F. Yates and I.J. Good had a related way for getting the effects in factorial experiments. The FFT idea switched a lot of Bell Labs effort from analogue to digital signal processing. It was wonderful to be there. It gave me things to do in statistics. The people involved got to be five years, maybe even more, ahead of the rest of the world.

5. LONDON SCHOOL OF ECONOMICS

Victor: How did England come about?

David: Well, part of the Canadian educational perspective – and maybe you felt this too even though you are from Greece– was that your education wasn't complete until you spent some time in England. It was that simple. So, I finished my doctorate, applied for a post-doc and got one! And then Lorie and I were off to England and to the London School of Economics. Actually, come to think of it, I've applied for only one job in my life that I wasn't offered. See I've been in the Navy, and then Lorie and I met up. She had strong political beliefs and I had strong social ones. Both of us were concerned with doing things about poverty and helping the developing world. So, I applied for a job at the United Nations – they were advertising for a statistician. Didn't even get interviewed! Didn't get it! Sometimes I think of how different our lives would have been. It is impossible to know, but things have certainly worked out.

Victor: ...for statistics definitely, but maybe not so for the United Nations!

David: [laughs] Sample surveys. I think that's what they were looking for.

Victor: But you've been involved in the International Statistical Institute, which has this attitude of solidarity too.

David: Oh, yes, definitely! That's been traditional and I'm glad I've had the chance to get involved in that. Anyway, England was about completing my education and I guess something led me to the London School of Economics. I am not sure just what it was, but that was wonderful. Because Kendall had just retired but was still around, Jim Durbin had just become a Professor. Alan Stuart was about to become one too. Maurice Quenouille was a Reader, Claus Moser was a Professor, as was R. G. D. Allen. I was surrounded by these senior people who were right in the middle of analyzing fundamental economic and political structures. It was pretty good, exciting even. They used to call these grants "post-doctoral drinking fellowships" [both laugh]. Lorie and I bought a Renault Dauphine and we went all over Europe. It was pretty cheap and safe then. Fred Mosteller wanted to offer me a job at Harvard when I came back, but he could never track me down. We were traveling to Austria for skiing!

Victor: Was there any difficulty in adjusting to the British view on statistics having been raised to the American attitude?

David: No, not really. I mean in Toronto then there was a very British background culture was there. Dan DeLury was a common sense person who said once that he re-read Fisher's Design of Experiments every year. I think I was different from the other British statisticians at the time, however, as I knew a fair amount of mathematics. Nowadays there are a lot of British statisticians who know a lot of mathematics. I'm afraid it sounds like I'm boasting too much just now. I saw Jim Durbin one time and he had some paper. He said he had tried to figure out something in it a few times but failed. He asked me: "David can you explain this?" I could tell at a glance that it was incorrect and said so. Jim said, "I wish I had your confidence." What he didn't have was my training, that's what the difference was.

Victor: Did you enjoy the RSS meetings?

David: Very much. I had never seen anything like them before in my life. There were people like Jack Good. He would stand up and be coming from a totally outside-the-box angle. I respected that because I had seen Tukey doing that all the time. At this point in my life, I believe that I have read most of Good's papers. I was honored to be asked to speak at his 65th birthday. I paid a lot of attention to what David Cox, Maurice Bartlett and George Barnard had to say, in particular. The way the meetings worked back then was that people could get the galleys of a meeting's paper before it was presented. So, you could compete with all these famous guys. You could read the papers and see if you had something to add to the discussion. That was a lot of fun. I'm not sure whether they do that now. I mean there certainly are discussions that go on. Back then, it seemed mostly in a spirit of friendliness, but now there seems to be real antagonism in the discussions as well as in referees' reports. They would make some strong remarks, but I wouldn't say they were mean then. Being a postdoc in England in the early sixties was great. We had a wonderful time. During the summer we went to the International Congress of Mathematicians in Stockholm. I found that I was reasonably well prepared for the level of the talks, having been to the various Princeton and Institute for Advanced Study seminars. It was exciting to see faces attached to many of the names that I had only read before. Hadamard is one I can mention. I went to one lecture in Stockholm – I think it was Linnik's. I got there early and talked with him. After I sat down, in comes Cramer, who sits right next to me! Then, in comes Kolmogorov and he sits on the other side of me! [both laugh] I was speechless! As you well know, I am usually quite talkative. I guess that I could have asked for autographs. That would have surprised them I am sure. Sadly I don't have a photograph to preserve the moment. It was pretty special and perhaps justified my having gotten a doctorate.

Then, we went back to Princeton. Lorie was pregnant so our life was going to change a lot. I went back to a job that was half time at Bell Labs, as Member of Technical Staff, and half time as a Lecturer in Mathematics at Princeton, teaching. The two positions were complementary in important ways. Tukey had created such a structure for himself; however, he was probably half-time in Princeton, half-time at Bell Labs and half-time in Washington. I guess that I then set out to have my own research career. I had done some writing of papers before, but now I settled into a more adult research program.

Victor: You seemed to be quite spread out at the time, I can see stuff in asymptotics [Brillinger (1962a)], Lie group invariance [Brillinger (1963a)], fiducial probability [Brillinger (1962b)], resampling [Brillinger (1963b)]... Really going off into many directions.

David: Well that was based on material I had learned. I would pick up a journal and see somebody had done something and if I thought there would be a way to contribute I would try. The Lie group material was motivated by Don Fraser. He was creating this area he called structural probability. I was trying to see if fiducial probability could be more formalized. R. A. Fisher kept pushing the idea of fiducial probability. It seemed as if in all his examples the fiducial probability was a Haar measure. So that was a natural thing to do. The Lie group paper arose also because people had wondered whether or not working with the correlation coefficient would lead to a fiducial distribution. I showed there was no prior –at least no Lie group measure that lead to one. But I was still solving problems, minor ones I suppose.

Victor: You mentioned reading papers and thinking about problems. I remember reading Tukey's Statistical Science interview [Fernholz & Morgenthaler (2000)] where he said that he would pick up journals and read papers, but not really study them. Which did you do?

David: I think I read them over. Because I had a reasonable memory and I could read quite quickly. So, a lot of my life has been working on something and then suddenly thinking: "Oh, yes, I've seen something like that before..." That's a problem with changing universities: because in Princeton library, I might have picked up some journal, but then having moved on to, say, LSE I had to search seriously. Anyway, I would pick up some journal, and read a paper that I sought in it, then. just as I was taught to read the dictionary, I'd look at the paper just before and the paper just after. That way you build up your knowledge. Also, when I have a journal issue in my hand, I don't think I read it to study it; rather. I read it to enjoy it.

Victor: And then came the baby and a decision to make: moving back to England.

David: Yes, that's right. Returning was an easy decision. Because Lorie and I both had loved living in London. Her being from New York city, and me from Toronto, we were used to "Which movie do we want to see? Then, where is it showing? OK, let's go!" Princeton was a small town and Lorie felt pretty restricted. Now we had the baby at home, but her parents lived up near New York City. I think it was pretty hard for her. Now women do keep working albeit part time or volunteering. But back then, they were right in the middle of the world, interacting with many people and ideas. Then, all of a sudden, they were at home for many hours with a baby. Well, Jim Durbin wrote me about there being a lectureship at the LSE, and was I interested. I think Lorie and I just had to look at each other for a moment to know we were interested. I stayed at Bell Labs through that summer to finish some projects and to build up some savings to go to England with. We had a VW van, so we were ahead of the hippies, and we shipped it over with us. We were driving around London for six years with this left hand drive big red VW van.

I have remarked many times that Bell Labs was the best job I had had in my life. Stimulating facilities, stimulating colleagues, stimulating problems and minimal restrictions on what one worked on. It is just that Murray Hill was in the middle of New Jersey. We were very fortunate to have the opportunity to decide how important was the choice of job as compared with the choice of where to live. My salary went down considerably of course.

Victor: What was life as a lecturer at the LSE like, and what was the contrast with Princeton?

David: Well, there were students of both sexes in the classroom at the LSE! They were left, not rightwing. In both cases, the students were very bright. Bill Cleveland was in a class that I took over when Sam Wilks died. Princeton and LSE were very different in many ways. I did prefer the English system in important ones. The thing I remember most about LSE is that there were five perhaps six of us, who were lecturers at the same time. We were of about the same age, having kids at the same time, watching the same TV programs. When Monty Python came along, we would all be talking about it the following Monday morning. They were teaching me about football/soccer and were learning about hockey and frisbee from Alastair Scott and me. We pretty much have all had successful careers. Fred Smith became the President of the Royal Statistical Society, Alastair Scott went back to New Zealand and was elected to the Royal Society of New Zealand, Graham Karlton moved to the Survey Research Center at the University of Michigan and became prominent in the US survey community, Wynn Lewis died young, Ken Wallace the econometrician amongst us was elected a Fellow of the British Academy [Most of the LSE statistics group in Fall 1969 are pictured and listed in Figure 6]. We were all together, all the time. We would go to the morning coffee, then have lunch and then afternoon tea again together. We drove across and around London to visit each other. At Princeton I was pretty much alone as a young person doing statistics.

Victor: But did your decidedly mathematical outlook tie in well with what was expected to be published in the British stats journals at the time?

David: I think that I know what you have in mind with that question. Just before we moved to England, I had submitted a paper to the Series B of the Journal of the Royal Statistical Society. It wasn't all that complicated, it was doing factor analysis with time series, getting latent values of spectral density matrices. I had in mind the problems Tukey had had me thinking about, concerning a signal from an earthquake or an explosion coming across an array of sensors. In an appendix, there was a derivation of approximate distributions of spectral estimates using prolate spheroidal functions, which Pollack and Slepian had come up with [Slepian & Pollack (1961)]. The referee said he didn't understand it and the paper was rejected! And I mean back then I didn't know about protesting an Editor's or Referee's decision. I probably should have re-written it and sent it back to JRSSB, but what does it matter? I did give a talk at an RSS meeting. Eventually. I put it on my website, and it's still there now. I developed the dimension reduction aspect further and have a paper on that in one of the multivariate analysis symposia and a chapter in my book. I don't think this occurrence affected me too much. but some of my students have been very disappointed by similar things in their career. Best I can tell them is that parts of life are arbitrary. resubmit.

Victor: By that time, you had been doing quite a lot of work on spectral analysis and then in '65 came the influential paper on polyspectra. That sounds like a Tukey term.

David: Yes, that is a Tukey term. One of the first things Alan Stuart said to me in London–you know how picky the English can be– was: "David, poly is a Greek prefix and spectrum is a Latin word. You are committing linguistic miscegenation!" He was just teasing me. But in Volume 1 of Kendall and Stuart [Kendall & Stuart (1963)] they say this against Tukey regarding "k-statistics".

FIG 6. *The Statistics Department at the London School of Economics in Fall 1969.*

Victor: Surely, there are many such examples - I can think of the word *bureaucracy* off the top of my head...

David: ... there's another thing that's wrong with bureaucracy! [both laugh] But anyway, I mean I was into all this non-linear stuff. Tukey, in an early memorandum had done something on the bispectrum. So that motivated me to do some research. You know, when you have a math background you seek to generalize things, to abstract them. It turned out I was unknowingly at first competing with the Russians -like Sinai and Kolmogorov - when I was doing that work. I heard that Kolmogorov had said some nice things about my work from Igor Zurbenko. That was really nice. Later on, the Russians translated my book into russian. I learned to read Russian mathematics in a fashion, in particular the works of Leonov and Shiryaev. That's what got me into the ergodicity results. For example, what I talked about today, was the Chandler wobble. Arato, Kolmogorov, and Sinai had a paper using stochastic differential equations to explain that motion [Arato et al. (1962)]. I was strongly influenced by French mathematics and a lot by Russian probability. I read the journals of both regularly. The work on cumulant functions and polyspectra let me get away from the restrictve assumption of Gaussianity in much of my later research.

Victor: Then, into the picture must have come Murray Rosenblatt, judging from your three joint papers on higher order spectra [Brillinger & Rosenblatt (1967a,b,c)]. I suppose he was in touch with the Russian school.

David: Oh yes, for sure. I had met Murray in New Jersey when he consulted at Bell Labs in 1963. I remember they had him working on the cepstrum, which is the inverse Fourier transform of the log of the spectrum. That work was part of estimating how deep earthquakes and explosions were, and so on. Then, Murray came to London. And again, I didn't know I shouldn't do something like this, being a young jerk, but I just went up to Murray and said something like "How about we write a paper and do some work together?" And he said "Fine." Murray has been my statistical role model, in many

FIG 7. *David with Murray Rosenblatt, and with Emanuel Parzen and his son, Michael.*

senses. Tukey was a creative role model. But at one point he said "Well David now that you are finishing what do you think you want to do?" He might have thought that I still wanted to become an actuary. What just came out of my mouth was: "I really don't want a life like what you have and I am concerned about whether I want to be an academic." And then Tukey, put his hands on his chin as he would often do and said: "What about Willy Feller? He has a pretty good life". So, he found a role model more to my liking. But then, I found Murray Rosenblatt. He just seemed to love his wife and his kids and had a lot going on in his life outside academia as well as a fine academic career. So he was a good role model. I don't think I really managed to express that to him until Richard Davis and I interviewed him for that article in Statistical Science [Brillinger & Davis (2009)]. He was a lot more of a mathematician than me, but in terms of his life, and interacting with people, I respected him.

Victor: Am I right that you also met Emanuel Parzen in England?

David: Oh yes, and we've been continually in touch since! We also met the Chernoffs then. This year, 2010, Manny and Carol are moving back to Palo Alto to a retirement home. So we expect to see a lot of them even though Palo Alto and Stanford have gotten steadily farther apart during our Berkeley years, in part because of the growth in traffic. But, with the Parzens moving there, I expect Palo Alto to come much closer. Manny and Carol are role models for us in a different ways. One is being a loving couple that were equal, with each member of the couple helping the other. And the other is Manny certainly helped me a lot by getting invitations to conferences, and by describing research that someone else was doing, so I was being kept up. And I think also by describing my research to other people. He was really the troubadour who was carrying the information of what was going on in other places around.

Victor: While maintaining a very strong concentration on cumulants and polymeasures, you also did some things on economics on the side.

David: Bell Labs had a lot of signal processing, so I was going into spectral analysis in detail. I think Kolmogorov and Sinai defined cumulant spectra in some sense, or cumulant functions. These functions turned out to provide a natural way to describe ergodicity and asymptotic independence. That's what I grabbed on to. That was the '65 paper. I think I might have been the first one to show that spectral estimates were asymptotically Gaussian without assuming that the time series itself was Gaussian. The economic work started in Princeton. Clive Granger –the Nobel prize winner– was at Princeton before I went to London. He and Michio Hatanaka were working on a book on spectrum analysis of economic series with John Tukey providing advice. When I moved to England, Clive was also there, at Nottingham, and would come down to the LSE every so often, so we had some contact over important periods. Hatanaka began working together and wrote a paper [Brillinger & Hatanaka (1970)]. I presented the work as an invited talk at the First World Econometric Meeting in Rome in 1965. Milton Friedman made the invitation. The work was concerned with the permanent income hypothesis and we had developed a time series spectral analysis formulation. After the talk, Friedman came up and said something like: "I didn't understand any of that but I am sure it was good!" [laughing] There is another paper with Michio [Brillinger & Hatanaka (1969)]. Data analyses were involved. My period at the LSE was by far the most theoretical in my career. I think because the time series data just weren't there. I was working as a consultant with the seismology group at Blackness. It was an offshoot of the Aldermaston Atomic Weapons Research Establishment outside that base. At one point, I provided an effective scheme for them to use with array data, but I guess that I wasn't able to explain it well enough. That's often been the story of my ideas. I don't know. Manny Parzen once quoted someone as saying: "First you have an idea and then you go out and sell it." But that was never me. I do try to ask myself: "Why am I writing this paper?" In the end I think that I am writing for John Tukey.

Victor: You've often mentioned the influence of scientific heroes.

David: Feynman would be one. I have read a lot by him and about him. I know that he enjoyed going to Brazil, as I have.

Victor: You didn't have a chance to meet him at Princeton, though.

David: No, he was long gone. He was there in the early war years, and left during them for Los Alamos. He ended up at Caltech . When I was asked to give a talk in Caltech once, he had died before. I might have been too intimidated to go talk to him anyway. Although I did talk to... Goodness, probably you know the name better than me. Who's the MIT linguist, who is in the news all the time?

Victor: Chomsky?

David: Yes Chomsky! I took Chomsky out for coffee once. It turned out that he and Tukey had organized a seminar on linguistics at the Institute for Advanced Study. This was when I was doing all these memorial articles about Tukey (Brillinger (2002)). I had noticed that Chomsky came to Berkeley regularly. So, I called a mutual friend and asked if they could arrange for a meeting next time Chomsky was in Berkeley. They did. Eventually, I met Chomsky at the linguistics department and took him over to this coffee place run by Palestinians. Victor, you have been there. While we were there, all these people were looking at Chomsky. One woman couldn't resist expressing her admiration for his work. He was such a humble sweet person. I asked him whether Tukey had any impact on the seminar. Chomsky said he sat there and grinned. I guess one takes that for what it is! So, being a Tukey student has given me entrée to countless situations. I'll tell you a story concerning that: just as I was finishing my studies at Princeton, I was invited to speak at the University of Michigan - I am sure due to Tukey interacting with Jimmy Savage. Jimmy Savage did a bit of political analysis of Lorie and me, and decided that our politics were on the left. He quickly organized for us to meet with Leslie Kish, sociologist in the Survey Research Center. That's when our close friendship started.

Victor: Leslie Kish had fought as a volunteer in the Spanish civil war.

David: That's right, and he was a leader of the Campaign for a Sane Nuclear Policy. So, Leslie had come to London and was giving a talk somewhere there. He later told me that he saw that I was in the last row doing something else. He said he got annoyed, but then immediately thought: "Oh no, he is a Tukey student, so that's all right!" [laughs] Now actually I was listening! Tukey could do three things at a time, I could maybe do two, sometimes.

Victor: Another name you often mentioned is David Cox.

David: Oh yes, he is another hero of mine. He too visited Bell Labs when I was working there. He was not a professor yet. He clearly had special things to say. Others might have done some of the things he did in a more mathematical way and subsequently gotten their names attached to them. I don't think he had a problem with that. I am thinking of things like getting approximate distributions of maximum likelihood estimators when the model is incorrect. He did that early on in a Berkeley Symposium paper [Cox (1961)]. Then, in another Berkeley Symposium, Huber came along and did it in a more formal way. Cox's paper has a wonderful statement, "Discussion of regularity conditions will not be attempted." There were very few, if any, of David's talks or papers that didn't have something clever in them. It's as if when he did something, if there wasn't anything clever in it [David thrusts his hand as if throwing away a piece of paper] then, no! Out of the window. He does it all in a very humble way. I have been on several committees with him and he would say few things for a while, but he would accumulate information and then he would come up with a proposition: "Well you could say ... maybe we could do ..." And everybody would agree. He could merge a lot of different opinions and information. He is one of my statistical heroes. He did reject a couple of papers that I submitted to Biometrika. I took that as saying, you can do better.

6. GOING TO CALIFORNIA

Victor: I understand that you would have been very happy to stay in London, but then things changed.

David: Yes, well my mother retired. She had had a hard life. She was a very bright woman, but because my maternal grandfather died in the great flu epedimic leaving my grandmother with five children, my mother had to go to typing school to help the family survive. Many years later, she went to adult school and got to be a country schoolteacher. We were sending her some money, but when she retired her pension was tiny. Even though I had become a Reader at LSE, there was just no way I made enough to make up what she needed. We had Jef and Matthew at that point, we were living quite happily, had a nice house a block away from Wimbledon Common. We were going to the theatre and concerts regularly. But there just was no way to be able to also support my mother. So I had to look for a higher income. Berkeley had already invited me several times. Actually, David Blackwell had called me just before I finished at Princeton. Now in the late sixties Berkeley was the place to be with the free speech movement, rock concerts, experimentation in the arts and all that. We had learned that when we were there on sabbatical in '67-'68. There were a growing number of protests against the Vietnam war, and Lorie was quite involved. So we knew Berkeley, and they knew me. And when Henry Scheffé asked me about moving there, we agreed. A person high in the academic totem pole told me once that a senior department member had said that I was the most influential appointment in the '70s. There were lots of mathematical things going on and I enjoyed that, but I was strongly interested in applications of mathematics. I immediately fell into place with Lucien Le Cam and Jerzy Neyman and all their visitors –they had a lot of important ones. So, we left London because we needed a higher income, but we landed in a very special place. Our older son, Jef, loved England. He was very sad about the move and that made Lorie and me sad. I think we expected that eventually he would move there.

Victor: So tell us a bit about your early Berkeley years.

David: The earliest years were '67-'68 when I was a visitor on leave from LSE and we have already talked about them. We moved to Berkeley permanently arriving by ship in January 1970 to be met by Erich Lehmann on one of the piers. At that time, there were a number of individuals who were then Assistant Professors but who did not get promoted to tenure, i.e. had to pack their bags and leave town. They were able academics so their non-retention was quite a shock for me. Actually, it seemed inhumane. Some of these people had children already at school. I was used to the English system where, if you were a Lecturer, and you had passed across the bar after three years, then you had tenure. You would hit the top salary of the lecturer scale but you might stay in your department the rest of your career – you had tenure. Some people did take advantage of that. We lost Berkeley friends that we had made and that was a great shock. Apart from that we were really enjoying the department, Berkeley and the Bay Area. The department seminars and the quality of the discussions in the lunch room was top notch. In these early years Kjell Doksum and his family became close friends.

Victor: Did you thus quickly forget about London?

David: No, not really. In fact, when in 1971 David Cox wrote that a professorial chair was available at Imperial College, and asked if I was interested, I was very interested! But going through the sums, with Alan Stuart's help, we just could not afford to return. Our old house was now worth more than twice as much as we had sold it for, within that short period. We couldn't afford to buy a comparable house.

I have sometimes wondered how things would have worked out with Jef's brain tumor had we returned. Cormack had just developed the first CT scanner at Atkinson Morley Hospital just down the

FIG 8. *David with Lorie along with David Blackwell and Maria Eulalia Vares.*

hill from our Wimbledon house. That technology wasn't yet available in the US, and might have helped.

Victor: But you found data at Berkeley.

David: Yes, I found data and fine applied scientists to work with at Berkeley. On reflection, I had reached the career that Tukey and Bell Labs had been training me for. Soon after arrival, I just wandered over to the seismographic station where I met this Australian fellow, Bruce Bolt. He and his family became dear friends. He was a sailor also, so we spent time on the Bay in his boat. Our families mingled. Bruce was religious, and I was no longer. However, we didn't seem to have the slightest difficulty talking about religion and other serious topics. He got me working on time series and other problems in seismology. We wrote several joint papers, but affected each other's research quite generally.

Victor: Was that around the time you wrote your invited paper on point process identification? [Brillinger (1975)]

David: There is a history to my work on point processes both in London and Berkeley. David Vere-Jones, another dear friend, another influence, presented an Invited Paper at a meeting of the Royal Statistical Society, [Vere-Jones (1970)]. I was asked to second the vote of thanks. When you are the seconder you are supposed to criticize the paper's content. Victor, you've probably been to these things. So I read David's very seriously. I don't think I had much in the way of criticizing, but it got me very interested in temporal point processes.

At Berkeley, Neyman and Scott had done path breaking work on spatial point processes, particularly in astronomy. Six months after my arrival in Berkeley in January the Sixth Berkeley Symposium took place. I presented a paper showing a way forward for making inferences based on data for processes with stationary increments. [Brillinger (1972)] This included stationary point processes. Around that time I also had a student, Tore Schweder, who was looking into that point process material when modeling whale tracks. To continue the story while Betty Scott was still department chair she asked me if there was anyone it would be good to invite to Berkeley for a term. I suggested David Vere-Jones. He and Daryl Daly came, and a whole world of point process work got started. In particular, David and Daryl organized a seminar series. Peter Lewis and "Pepe" Jose Segundo were important speakers. Peter's energy and enthusiasm and broad knowledge captivated the audience. Pepe came with specific problems and data concerning the firing of nerve cells. Pepe was a Professor in the Brain Research

Institute at UCLA. And he had all these wonderful data on nerve cells firing. And I just said, well this model that I have been fitting for earthquakes might be good. So then he sent me these massive piles of boxes of computer cards! They took up perhaps 10% of my office for many years! The thing that was interesting was that second-order spectral analysis seemed to be quite effective. So, I was working on point process data from seismology and point process data from neurophysiology at the same time. My students Rice and Akisik worked on these models/data also. The advantage of neurophysiology case was that it was a designed experiment situation, and thus you could repeat the experiment. So, that collaboration resulted because I was working on point processes from seismology. To my mind, one of the major successes was that the concept of partial coherency analysis could be extended quite directly to the point process case [Brillinger (1975)], and it let one infer the causal structure of networks of neurons, [Brillinger et al. (1976)].

Pepe had a daughter who died in a plane crash at Puerto Vallarta. At that time, I had a son with a brain tumour that could not be removed. These tragedies brought us very close together. Having a child die is pretty hard. Pepe and I had our scientific conversations to keep us focused on one good side of life.

Victor: Would you like to talk about Jef?

David: [David pauses and speaks with a broken voice] Well, yes. I mean it really affected Lorie, Matthew and me as well as Jef's and our friends. We have cared a lot about other people always. I don't believe that it is an accident that Lorie became a nurse midwife or that I started working with nerve cell spike trains. One works to fight for political ideals and to improve the system, but it is totally humbling to care so much about a child and not be able to help them in their time of greatest need.

Jef's illness went on many years. The first hint was in 1968 and he eventually died in 1988. It was not diagnosed as a brain tumor until 1973. He had three bouts of brain surgery and radiation between 1973 and 1988. In 1973 he was supposed to die within 6 months, but he just kept coming back. The night he died I didn't think he was going to die. He graduated from UC Santa Cruz in 1988, just two years behind his class. Everyone did everything imaginable. The doctors, his brother Matthew, Lorie and her nursing friends, our friends. The doctors made home visits. Nobody wants to see a child die. Many, many people attended the memorial.

Jef had a motorcycle, just as my mother and father had. I sometimes think about his motorcycle. I knew that I wasn't going to get on it but I knew about it. Jef rode it back and forth to Santa Cruz in part over a mountain. Once, there was a heavy rain storm and he thought that he might die. Another time, someone in the back of a pickup truck threw a bottle at him. He could have died on that motorcycle so easily. Then it would have been: if only, if only, if only... That's what our memories would have been. But our memory is that everybody did the best they could. Including Jef. Lorie has been really hard hit with death. She's had to nurse her dying parents, her son and her sister now.

Victor: Practically. everybody who's met you will attest to what an uplifting person you are; how it seems that you are always smiling.

David: Not always but most of the time. Probably my life was all fun until 1973 when Jef was diagnosed with the brain tumor. Science and researching kept me going through those times. Nowadays, I just have to think about my grandchildren and a smile surely appears on my face. Having gone through all this, I do go to a lot of effort to communicate with the Berkeley students about the importance of enjoying every day and realizing how lucky they are. In one of my classes in Berkeley, I realized that I was assigning a great number of problems. What I did at the spur of the moment was to say "OK, your problem assignment for this week is to go to a movie and then write on a piece of paper the name of the movie you've been to!" I think they just thought I was kidding. I wasn't. I have a hard time convincing today's students to put things into perspective. They seem quite terrified and not having all the fun that I had as a student. They are overly worried about getting registered in a class, about finding a thesis topic, about getting a post-doc, about getting a job, then about getting tenure, about getting a grant, getting to be a professor, getting to be invited to conferences. They have the problems of old people on their shoulders already! I am just sad for them. Things do work out. I hope you're trying to get your students to enjoy life, follow sports, things like that!

22

Victor: Well, I've had good advice, and try to pass on what I learned. Did research and sport help you at all during that difficult period?

David: When I was recently preparing an encyclopedia article on "soccer/world football" –that was the title I was given– and I was pulling out a lot of books, I found that there was a book by a couple of Russians on applications of mathematics to sports [Sadovskiĭ & Sadovskiĭ (1993)], because it has some material on soccer. When I read the introduction, I found them saying that to do mathematics well you want to be healthy and fit. I have known this for many years, but it was reassuring to see it in print. I think that participating in sports is important. You know, running around and interacting with others. I think of Shiryaev, since we're talking about the Russian point of view. He is a very good skier. He received a medal for it. There is something specific I'd like to feed into our conversation just now. I played a lot of intramural and informal soccer over the years. One year, two teams the Statistics Department was involved with met in the final. However, I stopped playing after Jef died. I wanted to be alone. Friends would come by my office to try to get me to play, but I just wanted to be alone. But my office looks over the Bay and much of the time I could see people sailing and windsurfing. I thought "Why don't I try windsurfing again?" I had tried once before and it hadn't really stuck. But when I tried again I got the basics. Windsurfing is one of those things where if you don't know what to try to do then you are in big trouble. What I found personally was that if I thought of anything else when I was windsurfing I would fall into the water. After I windsurfed for 2 hours I was just high. One day when I went back to Evans Hall, I saw Andrew Gelman and said something like: "I windsurfed all the way to Emeryville today!" Andrew said: "Well I climbed up the outside of Evans Hall today!" [laughs] It was that male thing, if someone is boasting too much, they get brought down. I do recommend to anyone who has some tragic situation to deal with, and they do like outdoor activity, that they take up windsurfing.

Victor: What was it like to arrive at Berkeley in the late 60's – early 70's?

David: Super. Rock concerts, progressive politics, long hair, hippies, tear gas. I was teaching once in a room in Wheeler Hall and all of a sudden there was some strange unfamiliar smell. I didn't know what was going on until someone in the class said: that's tear gas! It was really something. There had been "troubles" at LSE, but none with tear gas. I remember one friend I have, especially. When there was something radical going on I was out of there, headed away from the trouble. But I would invariably see him heading the opposite way i.e. in the direction of the trouble. I did see some bad things. Through my then office window on the third floor in the Physics building, I saw a sheriff's deputy club a young man who was just sitting under a tree reading a book. I think officers were totally frustrated because the demonstrators where leading them in a chase across campus. I do have to say that some were throwing rocks – and that's not cool. The deputies chased but they could not catch these guys. So, they just got more and more frustrated. Here's another story from that time period. Al Bowker had become Chancellor and joined our department. He had to deal with various ticklish situations during his tenure. Somehow, he always found a way. Evans was a new building and its inside walls were stark. One weekend some of the mathematicians came in and painted some murals. There was one of the death of Galois. The custodians cleaned them off. But the mathematicians re-painted the murals. A battle of wills was developing. Bowker said just leave them. Long after the murals were painted over when the building was refurbished and I don't know that there was any fuss.

Victor: Al [Bowker] told me a story about some students who were demonstrating. They came into his office wearing dark sunglasses – I suppose it was some sort of statement. But then Al caught them off guard: to their surprise he was already wearing dark sunglasses himself! [both laugh]

David: I had some fun like that too. When I was department chair, Lorie's brother was working for a video company that had produced a movie titled "Take This Job and Shove It". He mentioned that they were giving away hats with the movie title embossed. I asked if he could get me one of those, he did. One crisis that developed in my chairmanship occurred when the campus wished half of our space back – I confess that Betty Scott had been too effective in getting us space in the new Evans Hall. Anyway, when I went to see the Vice Chancellor I wore the hat and then passed it on to him! [both laugh] We ended up losing a quarter of our space.

Victor: What about departmental life? For example, Jerzy Neyman?

David: As far as I was concerned, being around him was a treat. One of Neyman's goals was "to find a model describing the data". In contrast, Tukey's goal was to "discover surprises in the data". Neyman was more for formalization, whereas Tukey was more for intuition. Surely, both are needed. I saw the two masters of these things at work. I attended the Neyman Seminar regularly and went for drinks afterwards. Neyman had a host of really wonderful visitors coming to Berkeley. I had total respect for that man.

Victor: And Neyman was one of the people you had gotten closer with along with Le Cam and Scott?

David: Yes. For one thing, they were always in the coffee room at lunch time, often with famous visitors eating Neyman's hard boiled eggs. The talk was lively, what with Neyman knowing so much about European history, all his languages and poems, and Betty being so full of heart and caring for people; Lucien being very French in such positive ways. The three cared so much about the students. Surely, the best part of Berkeley has always been the students. Once when I was in the coffee room, with Neyman and Le Cam, a student came in whose father was having a medical problem. Lucien and I were chipping in suggestions. After listening a while, Neyman remarked, "Isn't it wonderful that the professors are helping out the students with their personal problems?" All three would jump to help with student's personal difficulties. They were wonderful. I have been a bit unsatisfied with the Neyman biographies. They don't seem to bring out the essence of the man. I said this to Betty and Lucien once and they agreed. Biographies of scientists, by their nature, seem to focus on the science side. Setting down the human side is surely much harder.

I'll tell you one of the funny things that came to my head just now: somebody asked me once if I thought that Betty Scott and Jerzy Neyman were lovers. My immediate response was "I hope so!".

Victor: You had been exposed to two of three main schools of thought in statistics: Tukey-esque, British, and then came the third: Berkeley. What was that encounter like?

FIG 9. *David with John Tukey (left) and Jerzy Neyman (center).*

David: I would like to start by replacing "Tukey-esque" with Tukey-Bell-Labs-esque. That's the school that I learned EDA in. OK the encounter. I start by quoting Le Cam at this point. Once, at lunch, I told him about some research that I had just seen suggesting that cigarette smoking wasn't bad for one's health and at about the same time another report that suggested it was bad. What did he think about that? He replied: "They're both right!" The three schools are all right. We need each. I think it is important for people to travel and experience all three. The RSS meetings, for example, are a way to learn the British school. One meets these people and compares their discussions of the same paper. A lot of things exist in the scientific air, but are not written down, particularly heuristics. And it's very important to have heuristics along the way to nailing a problem down. Often, when you go to another center and are in a discussion, they quickly draw a little diagram and then you have picked that representation up. The thing is that you could go a whole career and never know that something could be simplified that much. As the years have passed, the British statistics school has become a lot more American. For example, consider measure theory and theorems. There have always been a lot of wonderful probabilists in England but they did not appear to have much influence on the statisticians until recently. One thing that I particularly respect about the English system, including people who aren't famous, is how well they can ask questions. There would be someone at a seminar, and then there would often be someone with a

British accent who would put their finger on a crucial point that's going on in the science. Not so much the mathematics, but the science of the situation. I have a lot of respect for that. What was the encounter like? I flitted amongst each of these schools. I am a scavenger. I have the luxury of trying a Tukey approach, trying a Cox approach, and trying a Neyman approach to problems. The Bell Labs group was influenced strongly by Cox, by Kempthorne and by Tukey. They weren't much influenced by Berkeley or Box.

Victor: 1975, Time Series: Data analysis and Theory [Brillinger (1975)].

David: Well, that book has got blood on every page! I wrote it when I was in England during the late sixties. It took too long to be published. I did enjoy working on it. I was going to LSE two days a week. We had a three-story townhouse. I would sit down on the top floor listening to the BBC's wonderful radio programs, working away on the book, while Lorie would be two floors down with Jef and Matthew. In the afternoon, I would be all involved with the kids. It was so enjoyable. The book started from my research, which got simplified for my lectures at LSE. Before reaching Berkeley in my 67-68 sabbatical, we spent the summer in Princeton. Tukey and I were supposed to be writing something up. But Tukey decided to go off somewhere, and there I was at Bell Labs. Ram Gnanadesikan asked me to give a course on time series. Luckily for me, somebody at the Labs was available to type up the notes. This provided a fine start to the book. There were all these wonderful computing facilities. The fast Fourier transform, a fast computer and graphics all came together there. Then I got back to England in the summer of 1968 and I guess that's when the serious filling in of material was done. The manuscript went to the publisher in '72 after I had made a serious attempt to have the references complete. It was printed in '74, but they put a date of '75 on it. It has now been with 4 publishers! That sounds amazing but Holt-Reinhart gave up their statistics list, Holden-Day went broke, and then it went to McGraw & Hill who put their binding on it but didn't do much else. It is now with SIAM and called a classic. How about that? There were some surprising benefits, like not having to do much preparation for lectures for many years. The thing that I enjoyed the very most was making up the problems at the ends of the chapters. Because I'd be thinking "Maybe there is a problem sort of like this", or "Maybe reasonable assumptions are something like these", and lastly "Maybe a solution could go as follows." The thing is one is negotiating with these three different vague items. It turned out that solving a problem was a lot easier than creating one! Victor, I did a vain thing the other day. I typed "Time Series: Data analysis and Theory" into Google. It claimed to have located 136,000 results!

Victor: You must have taught the time series graduate course "Stat 248" at Berkeley for many years.

David: I think every single year, except when I was on sabbatical. I believe Bob Shumway came then.

Victor: So did you change it quite a bit? I remember sitting in on three different versions.

David: Oh yes. I design it totally differently every year – and no one seems to notice! To allow variable content I call it "Random processes: data analysis and theory". A couple of students, not you of course, have said they should have come back. I try to tie it in to something I'm excited about at the time. Perhaps trajectories, perhaps point processes, perhaps spatial-temporal data and so on. I think if you are not excited about something, or if it is something you have done a long time ago it's boring. Nowadays, there are all these wonderful data sets and graphical devices to employ. It can take some time to prepare a display, but it would be a great shame not to.

Victor: You spent some time as a Visiting Professor of Mathematics in New Zealand. I know you are in love with New Zealand, is that when it started?

David: Yes. Alastair and Margaret Scott became dear friends in London. Alastair and I were Lecturers together. We had met at Bell Labs, and when I arrived in London he wrote me wondering if there were any jobs. So, I asked Jim Durbin, and there was a Lecturer position. Alastair stayed a couple of years longer than me. When Jef had the first surgery, he was really set back a long way. We wanted to go somewhere gentle, and that was New Zealand. There, his energy came back and he could do things like play basketball at a boys club Friday evenings and come home alone on the bus. He was about 12-13 years old then. It was the way things had been for me when I was that age. The Scott's friends became our friends right from the start. Alastair and I tried to collaborate on a paper once, but we never seemed to talk statistics. It wasn't that we didn't want to or couldn't, we just seemed to get talking about other things. But I do believe that we have influenced each other statistically a lot. So, New Zealand became our home away from home. NZ is where Lorie and I retreated to in 1988. That year was horrible. Lorie's father died, Jef died and my mother died. It has been important to Matthew, too. When Matthew decided he wanted to do a doctoral thesis in literature on Nabokov it turned out that the world's expert on Nabokov was in Auckland! To tie the knot even tighter we have three Kiwi grandchildren.

Another place I have a strong connection with is Brazil. It began in the context of graduate students. I had three Brazilian graduate students pretty early in my career. For many years, they were inviting me to come visit. I would tell them I was not going to any dictatorship. But eventually, the generals went away and luckily I was asked again. I went that time and had a wonderful visit. Brazilians and Canadians are very similar in many ways it turned out. In particular, they both have very high levels of teaching and research in statistics and of course sports are very important in both countries. Then, I got invited to another meeting and Pedro Morettin proposed that we apply for a joint NSF-CNPq (stet) grant. When the grant was funded for 3-4 years I decided it would be rude to have that grant and not make some attempt to learn Portuguese and took two courses. I have given talks in portuguese there and they have been very patient with me. One of the days that I was most proud of professionally was when I got elected to the Brazilian Academy of Sciences. That was quite a surprise!!

Victor: You also chaired the department at Berkeley for a couple of years. How was that?

David: I liked some parts of it, a lot. I got to know the staff very well, which I hadn't before. I got to know all the grad students very well, and many undergrads. I had many pleasant interactions with my colleagues also. But I couldn't do any research. Because whenever I tried to do research all of a sudden the day became too short or I was interrupted too often. I had agreed to do it for one year. The "candidates" had come down to David Freedman and me. David Blackwell said: "Well, it's you two. Time to chose." David and I each agreed to take it on for one year. I thought it was unfair that I was being expected to take it on then, because I had so many projects in process. David Freedman probably felt the same concerning himself. In the end I, did it for two years. David F. did it for five. As I just said, I did enjoy the job, but only after accepting not doing much research. The person whose model I followed in the job was Erich Lehmann. He had been chairman perhaps for four years and I just liked the way he did it. He would be in the coffee room at 10 am in case any of the students or faculty wanted to see him. One needs role models for how to do these different things, and Erich was my model for the chair position.

I just remembered a story. Actually during Erich's term I was (Acting) Chair for half a day. Erich had felt compelled to resign over some matter. I was Vice Chair which I guess made me Chair in a sense. However Erich didn't tell me that he had resigned until my "term" was virtually up.

Victor: So what is your opinion on leadership in academic departments? There's a sort of patriarchal paradigm with a dominant personality at the top and a democratic paradigm – e.g. Neyman years vs post-Neyman years. What's your take on that?

David: There is also an anarchist model. In fact when I first came to the Department there was something of an anarchist attitude - everything was being challenged, like language requirements. Barankin gave a stirring speech, which got rid of them. I believe that Neyman created some things that might never have existed without him. That was very special and what the right great leaders do. I don't feel that the faculty resented it too much, but I don't know. I liked being at the LSE rather than some other English university, because then there were something like 5 professors in the department [Figure 6]. Also mathematics was growing out of statistics there, not the other way around. The professors rotated the position around being chair for three years. What I tend to say when people tell me that they have been asked to be chair is: well, if you can do it, you have to. The thing is if the people who could do it manage to get out of doing so then the system of good governance collapses. Anyone who could do it has to take their turn. An advantage is that different things are emphasized depending on who is the chair. In my term, I put a lot of department resources into computing It seemed the time

Fig 10. *David and the Berkeley Statistics Soccer Crew. From left to right, starting at the top: Tom Permutt, Jan Bjornstad, Jim Veetch, ?, Annibal Parracho, David, Peter Guttorp, Kai(-squared), Eldar Straum, Albrecht Erle, Ken Saltrick.*

for that and I could handle the decisions. Incidentally, one of my students said that as soon as he learned I was going to be chair he worked very hard to get his thesis finished. So my taking the job on was good for him.

There are different attitudes concerning how to behave as chair. When I was doing it, the budgeting was actually very loose, but I didn't know that. A friend who was chair of another department heard me muttering about restrictions on money. And he said: "Oh just spend it! Let the dean find the money!" I guess there was no mechanism at the time to pick up on overspending. When I told the financial dean that I was spending money like it was my own he said "Good!" Many university things were much more casual back then.

Victor: By next year, you will have had 40 students, some very notable people amongst them.

David: Students have been one of my great joys at Berkeley. If for no other reason, they are a motivation for seeking a position here. There is a nice picture of me with many of "my" doctoral ones in Banff [Figure 11]. I sometimes wonder whether I could have supervised a student and not become friends with them. They certainly do become friends. As you point out, my rate is about one student a year, and that's probably a reasonable one because they take 2 to 3 years to complete the thesis. Nowadays, there are research groups or labs. I tried that in the mid-seventies, but it didn't seem to work well for me, or, more importantly, for the students. My goal is to have the students learn how to do independent research. This was Tukey's way. I sometimes see my ex-students treating their students the same way. I interact with a student to find a topic that they are really interested in. Nowadays, statistics is everywhere, so that hasn't be too hard. I think when you are interested in something you just find yourself progressing and the time flying by. I used to play a lot of intramural soccer. That's actually a good way to get to know students and visitors. When you kick them, accidentally of course, you see how they respond and when they kick you they see how you respond. You learn a lot about each other!

By the way, I will not sign off on a student's thesis until they have started arguing with me and are calling me David. For some students that can be hard, but they need to be toughened for the outside world.

7. "$2\pi \neq 1$"

Victor: I was wondering if we could go back to research a bit. The title you used for your 2005 Ney-

FIG 11. *David with a group of his PhD students in Banff, 2003. From left to right, starting at the top: Bruce Smith, Peter Guttorp, Tony Thrall, Knut Aase, Mark Rizzardi, Rick Schoenberg, Ed Ionides, Isao Miyaoka, Haiganoush Preisler, Jostein Lillestol, Tore Schweder, John Rice, Andrey Feuerverger, Alan Izenman, Raju Bhansali, David*

man Lecture [Brillinger (2008)] was "Dynamic Indeterminism in Science". Would you say this describes your scientific vita?

David: I like your question. In a word the answer is, maybe. That expression is to be found in a 1960 paper of Neyman's [Neyman (1960)]. He was encouraging people to learn about stochastic processes. I don't think many statisticians did back then. And then I was invited to give a talk [Brillinger (1984)] at the International Congress of Mathematicians in Poland in 1983. I talked about statistical inference for stochastic processes in a general way. There weren't many people doing that then. Murray Rosenblatt and Ulf Grenander were involved with it but the list of people working with a general process framework was short. One conceives a datum that is a realization of a process. That's what Neyman was encouraging people to work with. LeCam's approach was totally abstract, so everything was a particular case – but in a sensible way.

Victor: I recall you were mentioning in the doctoral course on applied statistics at Berkeley that "Any mathematical object can become a realization of a random variable".

David: For sure. You just put a collection of the objects in a hat. Then you find a sensible way to pick one of them at random and then you've got a realization. Think about the article I showed at my talk this morning about statisticians being the sexy thing to be for the next 10 years [Lohr (2009)]. The rest of the world has clued into that, finally! There are these wonderful data sets with people who care about them. And statistics has an immense amount to contribute to their study. Plus it's going to be a lot of fun to be doing it. You have music in your computer, videos in your computer you may even have a Bible in your computer – all this stuff is nowadays in a computer, just waiting for you to discover surprises in it! That's a Tukey attitude. I never saw Tukey doing any computer programming, but he could surely visualize it. And he was very much involved in the first Von Neumann computer [Brillinger & Tukey (1985)]. So, he knew about it in that sense. I did see him with coding sheets, but he was preparing things for cards to be punched for his citation indices [Brillinger & Tukey (1985)].

28

Victor: Some consider you as a theoretical statistician, others consider you as an applied statistician. Which one is it? Always learn new theory?

David: Oh yes? Where did you get that ?!? [both laugh]. That's my motto: always learn theory, for the theory becomes the practice. I can provide a lot of evidence about that and I think it is what places the Berkeley students in a good position when they finish. Because other places will create students who are really up to date the moment they finish, but not ready for new things that come along. It's harder for them to keep on top of things. They may well feel intimidated and struggling to keep up. I think the students coming to Berkeley get a lot of gifts from the people here. One can mention LeCam with his abstract approach to things and depth of thought. I had great respect for him for a lot of reasons. One of them is he could sit in his office and he could dream of these incredible mathematical problems, and dream up solutions. Whereas my thing to do is to find a parallel scientific situation where that problem exists. This can give important clues about how to approach the problem. Lucien always seemed able to generalize these things in such a way that he would encompass so many things. I would take some of his work and particularize it to a specific situation.

Victor: Is that your research strategy? How do you attack problems? How do you find or choose them?

David: I find them by people interacting with me, or by my asking them. As I mentioned earlier when I arrived in Berkeley, I went over to the Seismographic Station. They didn't come to me. I think that with a consulting service you don't really get the special people coming. You have to go over to them, to the scientists. You have to present yourself to them. Terry Speed and I agreed on this once. Terry was chasing across campus some time after he arrived, interacting with people, particularly in biology. When I think about my recent work: risk analysis was motivated by interactions with Bruce Bolt of the Seismographic Stations,the trajectory modelling was based on data collected by Brent Stewart of Hubbs Sea World,while both topics involved Alan Ager and Haiganoush Preisler of the US Forest Service.The work on sports statistics is based on data that I collected on my own. At a certain point you've got all the problems you can handle. It seems in any case that if you want to work with good people, then you have to go after them. So I've just come to know a lot of people. Various of my papers may be found in [Guttorp (2010)].

Now, I am a member of the scientific advisory panel this new center of excellence for evolutionary biology at the University of Oslo, and there is a flood of new problems coming into my head from that. It is just wonderful. But I was wondering: why me on this panel? And then I thought: oh evolution that is time-series, isn't it? It is just a totally different group of scientists from any I have been involved with before. Now I own a great thick book on evolutionary biology.

Victor: In a recent article [Dyson (2009)], Freeman Dyson classifies mathematicians as frogs and birds; or as Erich Lehmann put it [Lehman (2008)]: problem solvers and system builders. Where do you stand?

David: I like to be a bit of both. I like solving problems, but yet from my math background I like to abstract things. I like to transfer information between fields. So, I have worked at the same time with a seismologist, Bruce Bolt, and with a neuroscientist, Walter Freeman. Walter works with EEG [electroencephalogram] analysis. I would be telling Walter some of the clever things the seismologists were doing and I would be telling Bruce some of the clever things that the neuroscientists were doing. They each could then be thinking of applying these things to their own data. Abstraction was the route between the two fields. Transfer of knowledge is a topical goal and the politicians like it a lot. It probably makes sense because you can "start sooner" in a different field. Dyson by the way is another hero. I think I read various of his books and papers. I used to look a lot at the physics literature.

Victor: Do you have a favorite paper?

David: I believe that my favorite papers are the ones that I had to work the hardest to get the result. I believe I told you I had solved all the problems, except one, in Sam Wilk's book. The one which was about getting an asymptotic joint distribution of the median and the mean. I did not know how to get that and when I told Sam I don't think he knew how either. He said he had found the result in a paper by some Hungarians. I never found that paper either. Eventually, I ran into the notions of strong approximations, later called coupling, and read a report by

Ron Pyke -another role model of mine- and one of his students, on getting a strong approximation for the empirical CDF using tied down Brownian motion. But for the problem I was concerned with, I needed an error term. I think I was the first to set down that approximation with an error term. The Hungarians then referred to my work and generalized it to get a lot of wonderful results.

Victor: You're referring to your early Bulletin of the AMS paper on the representation of an empirical distribution function [Brillinger (1969)]?

David: That's right. That's one of my favorites. It just opened up a whole host of things. Then, of course, when you get such a result you can improve it a great deal. But this strong approximation just lets you write down results using standard calculus. That was an important one to me.

Victor: And what about a "favorite rejected paper", or to put it differently, is there an instance when you might have felt angry at a referee?

David: No, never anger at an academic referee, sometimes anger at a soccer referee [Victor laughs]. I had a paper once, that I thought was quite interesting, on a representation for polymeasures. So polymeasures do relate to polyspectra, but really it was more useful for non-linear operators. I mean there's this huge world of linear operators, but polymeasures provide you with representations for an important class of polynomial operators. And then, since I was just about to move to England, I thought it would make sense to send it to the Journal of the London Mathematical Society. To this day, I think that if I had actually been at LSE and sent it from there, they would have accepted it. But I just got a referee's report back saying that they were just not interested in that type of paper. I was young, I was learning. I still had the attitude that I'd rather be playing hockey than doing this stuff, and that stood me a good stead. Really, that's not made up. Plus I had Tukey telling me that he had many papers rejected. I think I read somewhere that Rob Tibshirani said that his first ten papers were rejected. Tukey's thing was resubmit somewhere else. I sent it to the Proceedings of the American Mathematical Society and they accepted it directly. [Brillinger (1967)].

Tukey and I had a paper rejected by two journals [Brillinger & Tukey (1985)]. He told me not to worry, it could appear in his Collected Works, and it did.

Victor: Going in the other direction, was there a paper that you found had much more impact than what you would have expected?

David: I just love to do math problems. All through High School and University, there were problems from the American Mathematical Monthly that I would try to solve. So, I was doing it for my amusement. You know, you could send a solution and sometimes they would publish it. So, I think in many cases that's why I was doing things: there was a problem, and I was there. So, the polyspectra paper [Brillinger (1965)], just started out from having fun. I found that cumulants were a way to go. They had this property that, if there was a multivariate variable, and if some set of its variables was independent of the rest, then the joint cumulant was zero. This takes one directly to a definition of mixing for general stationary processes. Perhaps the Russians knew that result, but anyway. But I was working on this for fun. At one point, Tukey mentioned the word, polyspectra, and I made the connection – and wrote that paper. That paper might have helped me get some invitations to speak and job offers and promotions. It surely led to my collaborating with Murray Rosenblatt.

Victor: Well, it's been cited over 200 times, I think!

David: I remember I gave a talk on that research at Cambridge. David Kendall, whose work you know well, had invited me. When I was done with the talk, I think he was as baffled as most other people were by what I was up to. Maybe I was just not good at explaining it. Hopefully, I eventually learned how to do so. Anyway, Kendall said something like "Now let's go have some poly-tea in our poly-cups". So that broke the ice [laughs]. Most of these great people have a sense of humor. They can seem pretty serious because one has to think hard to do the research. But you realize that basically they're people who have families, and have fun with their children at the playground. There is a human side to all of them. So, in the beginning, very few people would refer to that paper at all. I think Kolmogorov knew about it, and I had a bit of an interaction with Zurbenko about it. But that was pretty much it. But then, in the early '80's all of a sudden I get this flood of reprint requests! This was when people still used reprints, they didn't have things on the web. And so. all of a sudden I'm being invited to these conferences, some of them in exotic places, on "Higher Order Spectra" – that's what they called it. My preference is cumulant spectra. I remember saying things at some of these conferences, like "Nothing matters unless you show it used on a real data set". And I remember seeing some of the engineers looking at each other. Because in so many cases they would tend to use proof by simulation. That gave them the feeling they had done their duty in terms of a proof. I don't put them down, I have a huge amount of respect for engineers. My favorite committees are engineering committees because they have something better to do than being on the committees! And they have this attitude, that Allin Cornell, an earthquake engineer expressed to me once, the attitude that every engineering problem has a solution. And I think Tukey was showing me that many times over in the form that every statistics problem has a solution. And that it's the statistician's responsibility to find it. You can't just abandon a scientist and their data.

Victor: On your office door in Evans Hall there is a sticker: $2\pi \neq 1$. Would you care to elaborate on this for the uninitiated?

David: Oh well, yes, that's my logo! I usually like to make people figure it out. It goes back a long way. Here's one story, this student, Raffa [Irizarry] whom I have mentioned already, was just a joy. I would hear loud footsteps of someone running down the corridor towards my office. And then Raffa would appear, slide me off my chair, and open a window on my computer saying "You have got to see this!" One day he ran into my office saying: "I found it! 2π is not 1!". He had discovered what was going wrong in his computations by simulating the basic procedure countless times for a known case. His answer was out by a multiple of 2π. Rafa was already a modern statistician using Mathematica and simulation to deal with analytic problems. By the way, he just received COPSS' Young Statisticians Award. That made me very proud. Peter Guttorp just got an honorary degree from his home University of Lund. The grad students have been my great joy at Berkeley. Ross Ihaka received the Pickering Medal in New Zealand for his work in developing the statistical package R. Others too. I mean my students make me proud for their research and professional contributions. John Rice has excelled in those two areas and just completed a second successful term as our Department Chair. They are grandchildren of Tukey's, and a lot of what they are getting from me is what I learned from Tukey. For example, you've seen me filing papers with these plastic ziplock bags? Well this is a Tukey idea from many years ago! Victor, does Stephan [Morgenthaler] ever do that?

Victor: I don't recall, I'll make sure to check!

David: Well, you can tease him about it. If he says no, tell him that Brillinger says he would have a better career using these bags! He will have an answer to that, I'm sure! [both laugh]

Victor: Churchill [(Churchill, 1930, p. 17)] wrote something like "All students should learn English, and then the clever ones should take Latin as an honour and Greek as a treat". Translated into mathematical or statistical topics, what would be your pick?

David: You could probably ask me that 5 times and get 5 totally different answers! Because right now I think it's puzzles. As a youngster, I was always doing problems in the newspaper, you know: "three men are in a room and they can't see what's on their own head ...", and things like that. I had a lot of fun in doing that and a lot of good intellectual exercise. Perhaps the exercises in my book was the part I enjoyed most. It was the hardest part too. The things I had to work hardest on are the ones I respect the most. I developed an estimation method and a paper once, on my bike ride home. I had the idea, went to the typewriter upstairs, sat down, and typed it up. I sent it to Biological Cybernetics directly [Brillinger (1978)]. All done in a couple of hours! That didn't impress me. Then, there are some other things like how to handle the "integrate and fire" model in neuroscience [Brillinger & Segundo (1979)], which took quite a while to come along.

Victor: As we already mentioned, you will have supervised 40 PhD dissertations by next January. What would be your advice to the next generation?

David: It seems to me that learning mathematics is nowadays being replaced by learning computer science. I think it would be good for students to learn near equal amounts of each of these. Computer science lets one check out proposed methods, learn about data structures - after all the data are typically in a computer - and get approximate answers. But I am not sure it really takes you to the

FIG 12. *Dama and Victor with the Swiss Alps in the background. Photo taken during the interview session, September 2009. David is proudly wearing the Canadian Soccer team shirt.*

essence of a lot of situations. Think of the neural net models. They can be justified by the science, as in the threshold case mentioned above. However, I am uneasy about throwing everything in there and getting an answer without a scientific interpretation. I would rather use something that has scientifically interpretable parameters. Let me add though that I am certainly not averse to using some tool to see what it can do for me. I would like to see students come back to studying more serious mathematics. I'm astonished that some students in the computer science community don't know elementary trigonometric identities. For them, the Fourier transform is just the FFT: you put this in and you get this out. People learn a lot by just doing something and seeing what you get. That's a system identification approach where one inputs a signal and sees what comes out. I think it is a lot more rewarding to really get some understanding of *why* it is happening. Although in science it doesn't always work that way. I remember Fred Mosteller saying many years ago that nobody knew then why aspirin worked, but that of course we are going to use it because it appeared to work. But still I think learning what the thing was doing is fundamental, because then you can improve on it.

My bottom line is: have fun! That sounds trite but I'm serious. If you are worried about something consider what you can do about it. If there is something, do it. If not, what's the point of worrying? When you have a child die after a very long battle with cancer, as Lorie and I did, you simplify a lot of things. You take things to their essence. Don't be afraid to cry. It is another thing you learn going through a tragedy. Many say crying is hard sometimes. For me, it just happens.

Victor: David, thank you very much for sharing these memories of your remarkable life and career. But I have to ask one last question: would you still rather have been a hockey player?

David: Oh yes!!! [laughs out loud] There is noooooo doubt in that! I gave the after-dinner talk at one of the Canadian Statistical Society meetings and the title was: "Why I became a Statistician". You can guess what the punchline was!

Victor: Thanks again, David

David: Thank you Victor. You had some good questions. I mentioned only some of my students. I probably have an anecdote about each, but I'll save those for another time.

ACKNOWLEDGEMENTS

We would both like to thank Ms. Maroussia Schaffner for her patient work in transcribing our

32

nearly four-hour conversation. Parts have been lightly edited and reorganized for clarity. David thanks the many people who have helped him during his career.

REFERENCES

ARATO, M., KOLMOGOROV, A. N. & SINAI, YA. G. (1962). Estimation of the parameters of a complex Gaussian Markov process. Soviet Math. 1368-1371.

BARTON, D. K. ET AL. (1998). Final Report of the Haystack Orbital Debris Data Review Panel. Technical Memorandum 4809 NASA, Johnson Space Center.

BONDFIELD, G. H. (1912) *The Christian Movement in China.* Christian Literature Society for China, Shanghai.

BOURBAKI, N. (1951) *Algebra.* Hermann, Paris.

BRILLINGER, D.R. (1962). A note on the rate of convergence of a mean. *Biometrika, **49**:574–576.*

BRILLINGER, D.R. (1962). Examples bearing on the definition of fiducial probability with a bibliography. *Ann. Math. Statist., **33**: 1349–1355.*

BRILLINGER, D.R. (1963a). Necessary and sufficient conditions for a statistical problem to be invariant under the action of a Lie group. *Ann. Math. Statist., **34**: 492–500.*

BRILLINGER, D.R. (1963). A note on the re-use of samples. *Annals of Mathematical Statistics, **34**: 341–343.*

BRILLINGER, D.R. (1965). An introduction to polyspectra. Annals of Mathematical Statistics, **36**: 1351–1374.

BRILLINGER, D.R. (1967). Bounded polymeasures and associated translation commutative polynomial operators. *Proc. Am. Math. Soc., **18**: 487–491.*

BRILLINGER, D. R. (1969). An asymptotic representation of the sample distribution function, *Bulletin of the American Mathematical Society*, Vol. 75 (1969), pp. 545- 547

BRILLINGER, D. R. (1972). The speactral analysis of stationary interval functions. Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1, 483-513.

BRILLINGER, D.R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, San Francisco.

BRILLINGER, D. R. (1975). The identification of point process systems, Annals of Probability, Vol. 3, (1975), pp. 909-924.

BRILLINGER, D.R.(2002). John Wilder Tukey (1915-2000). *Notices American Mathematical Society, **49**: 193–201.*

BRILLINGER, D. R. (1978).A note on the estimation of evoked response, Biological Cybernetics, Vol. 31, 141-144.

BRILLINGER, D. R. (1984).Statistical inference for random processes, Proceedings 1982 International Congress of Mathematicians, Warsaw, 1049-1061.

BRILLINGER, D. R. (2002).John W. Tukey: the Life and Professional Contributions. Annals of Statistics 30, 1535-1575.

BRILLINGER, D.R. (2008). The 2005 Neyman Lecture: dynamic indeterminism in science. *Stat. Sci., **23**: 48–64.*

BRILLINGER, D. R., BRYANT H. L. & J.P. SEGUNDO (1976).D.R. Brillinger, Identification of synaptic interactions, Biological Cybernetics, Vol. 22, 213-228.

BRILLINGER, D.R. & DAVIS, R.A. (2009). A conversation with Murray Rosenblatt. *Statistical Science, **24**: 116–140.*

BRILLINGER, D.R. & HATANAKA, M. (1969). A harmonic analysis of non-stationary multivariate economic processes. *Econometrica, **7**: 131–141.*

BRILLINGER, D.R. & HATANAKA, M. (1970). A permanent income hypothesis relating to the aggregate demand for money. *Economic Studies Quarterly, **21**:44–71.*

BRILLINGER, D.R. & ROSENBLATT, M. (1967). Asymptotic theory of estimatesof k-th order spectra. *Proceedings on the National Academy of Sciences of the USA, **57**: 206–210.*

BRILLINGER, D.R. & ROSENBLATT, M. (1967). Asymptotic theory of estimates of k-th order spectra. *Advanced Seminar on Spectral Analysis (ed. B. Harris), New York, Wiley, 153–188.*

BRILLINGER, D.R. & ROSENBLATT, M. (1967). Computation and interpretation of k-th order spectra. *Advanced Seminar on Spectral Analysis (ed. B. Harris), New York, Wiley, 189–232.*

BRILLINGER, D. R. & J.P. SEGUNDO (1979). Empirical examination of the threshold model of neuron firing, Bio. Cybernetics, Vol. 35, 213-220.

BRILLINGER, D. R. & J.W. TUKEY (1985). Spectrum analysis in the presence of noise: some issues and examples, The Collected Works of J.W. Tukey, Vol. 2, (ed. D.R. Brillinger) Wadsworth. 1001-1141.

CHURCHILL, W.S. (1930). *My Early Life: A Roving Commission.* Charles Scribner's Sons, New York.

COX, D. R. (1961) Tests of separate familis of hypotheses. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 105-123

COXETER, H.S.M. (1961). *Introduction to Geometry.* John Wiley & Sons, New York.

DYSON, F. (2009). Birds and Frogs. *Notices of the American Mathematical Society, **56**: 212–223.*

FERNHOLZ, L.T. & MORGENTHALER, S. (2000). A Conversation with John W. Tukey and Elizabeth Tukey. *Statistical Science, **15**: 79–94.*

FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Cambridge University Press, Cambridge.

GUTTORP, P. (2010). *Selected Papers of David R. Brillinger* Institute of Mathematical Statistics.

KENDALL, M.G. & STUART, A. (1963). *Advanced Theory of Statistics, Volume I.* Griffin, London.

LEHMANN, E. L. (2008)] *Reminiscences of a Statistician* Springer, New York.

LONEY, S.L. (1930).*An Elementary Treatise on Statics.* Cambridge University Press, Cambridge.

LOHR, S. (2009). For Today's Graduate, Just One Word: Statistics. *The New York Times, August 6 2009, p. A1.*

NEYMAN, J. (1960). Indeterminism in Science and New Demands on Statisticians. *Journal of the American Statistical Association, **55**: 625–639*

PIAGGIO, H.T.H. (1920). *An Elementary Treatise on Differential Equations.* G. Bell & Sons.

SLEPIAN, D. & POLLACK, H.O. (1961). Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty – I. *Bell Syst. Tech. J., **40**:43–64.*

SADOVSKIĬ, L.E. & SADOVSKIĬ, A.L. (1993). *Mathematics and Sport.* American Mathematical Society, Providence.

TUKEY, J. W. (1963) Mathematics 569 - An introduction to the frequency analysis of time series, The Collected

Works of J.W. Tukey, Vol. 1 (1984), (ed. D.R. Brillinger) Wadsworth. 503-650.

VERE-JONES, D. (1970). Stochastic models for earthquake occurrence (with discussion). *J. R. Statist. Soc. B* ***32***: *1-62.*

WILKS, S.S. (1963). *Mathematical Statistics.* Princeton University Press.

Part I
Theoretical Statistics

Commentary by Victor M. Panaretos

Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group [1963]

David began working on this paper while he held appointments at Princeton and Bell Labs, and completed it at the London School of Economics. He recalls (Panaretos [16]) that his motivation to consider this problem came from Don Fraser's program of *structural probability*, and in particular from the issue of formalising aspects of Fisher's *fiducial probability*. A particular example that David had in mind was that of the correlation coefficient: could Fraser's results be used to show that Fisher's fiducial distribution (Fisher [7]) can be obtained as a Bayesian posterior for some prior - as is often the case when a unique sufficient statistic exists? Lindley [15] had proved that, in the real case, a fiducial distribution would arise as a posterior if and only if the statistical problem were invariant, and so David set out to find conditions for invariance. To study the general case, and since most of Fisher's examples seemed to essentially involve a Haar measure, David was naturally led to consider the invariance of statistical decision problems with manifold parameter/sample spaces being acted upon by a suitable Lie group. In the general case, David's necessary conditions amounted to requiring that: (a) the loss function vanish under the action of the infinitesimal generator of the product Lie group (i.e. the Lie group simultaneously acting on the sample space, action space and parameter space), and (b) that the image of the log-likelihood function under the infinitesimal generator of the product Lie group acting on the parameter and sample space be equal to minus the derivative of the action of the Lie group on the sample space. In short, the sufficient conditions (which are of interest for the application that David had in mind) essentially require that there exist linear differential operators (understood as infinitesimal generators of the candidate Lie group) such that the necessary conditions hold true plus the additional requirement that these generate a global group. Conditions for the generation of a local group (Lie's second Fundamental Theorem) were given, and then two recipes for attempting to establish that the local group is in fact a global group were described. When the parameter and sample spaces are Euclidean, the conditions simplify significantly. Applying his results to the special case of the correlation coefficient of bivariate Gaussian data, David was then able to conclude that there exists no prior distribution such that Fisher's fiducial distribution be interpreted as a Bayesian posterior. Once in a talk of his at a Bell Labs seminar in the '60's, David recalls someone chasing him around a bit during the talk, and Martin Wilk saying something in the lines of "*watch out or David will trap you in a Lie group*". "*That turned out not to be necessary*" David remarked when telling me this story.

An asymptotic representation of the sample distribution function [1969]

This paper was published in 1969, communicated by David Blackwell, when David Brillinger was a lecturer at the London School of Economics. Its story, however, can be traced ten years back to an anecdote dating just before David started his PhD at Princeton. David had just

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_1,

arrived in New Jersey and Sam Wilks asked him to go through all the problems in the book he was preparing on Mathematical Statistics, and come up with solutions. David managed to finish them all except for one: proving that the mean and median where jointly asymptotically Gaussian. This problem was left out of the book, and as David later recalled *"it took me a while until I found a neat way to do that"* (Panaretos [16]). This neat approach is the topic of this paper. It consists in the constructing a coupling between the uniform empirical process and a Brownian bridge such that, with probability 1, their uniform distance decreases like $n^{-1/4}\sqrt{\log n \sqrt{\log\log n}}$. Obviously, application of the probability integral transform for a general distribution function would then allow the application of this result to the study of the asymptotic behavior of functionals of arbitrary continuous distributions. The coupling relies on a so-called renewal construction: the empirical process is represented through the partial sums of independent exponential random variables. This type of construction can be found in Breiman's book [5, p. 285], but it was Brillinger who independently realised that such a representation could be employed for the purpose of proving uniform central limit rates. The coupling allows the use of the Skorokhod-Strassen approximation theorem (e.g. Shorack & Wellner [21, p. 60]) to obtain the desired rate. This rate was later improved through a different technique known as the *Hungarian construction* (see Gaenssler and Stute [8, par. 3.5]). A side remark is that the Brownian motion, say $\{B(x)\}$, in David's coupling is dependent on n through $B(x) \equiv B(n,x) = n^{-1/2}W(nx)$, where W is standard Brownian motion. The Brownian bridge transformation of this process, $\mathbb{B}(n,x) = n^{-1/2}[B(n,x) - xB(n,1)] = n^{-1/2}[W(nx) - xW(n)]$, allowing for n to take real values, has become known as the *Brillinger process* (Shorack & Wellner [21, p. 33]). In our recent conversation (Panaretos [16]), David mentioned that this was one of his favourite papers.

The spectral analysis of stationary interval functions [1972]

In the summer of 1970, six months after David's arrival at Berkeley as a permanent faculty member, the Sixth (and last) Berkeley Symposium took place, with a glittering line-up of participants, including Anderson, Bartlett, Burkholder, Birnbaum, Chernoff, Cochran, D.R. Cox, Daniels, Doob, Dvoretzky, Gani, Gnedenko, Hàjek, Hammersley, Itô, Kac, Kakutani, Kingman, Lévy, Lipster, Meyer, C.R. Rao, Rosenblatt, Savage, Shiryaev, Stein and Varadhan, to mention a few – and of course the all-star Berkeley faculty, including the later Nobel Prize winner in economics Gérard Debreu. David's contribution to the symposium was "*a way forward for making inferences based on data for processes with stationary increments*" (Panaretos [16]). Much of the work had been carried out in London, where David was motivated by the work of Bartlett, Cox, Lewis and Vere-Jones. David in fact recalls David Cox, Maurice Bartlett, Toby Lewis, John Kingman and Henry Daniels being in the audience for his talk. His study is carried out in the (equivalent) framework of stationary interval processes, that is, vector-valued random processes defined on bounded sub-intervals of the real line, $\{\boldsymbol{X}(\Delta)\}$ with $\Delta = (\alpha, \beta]$, $\alpha, \beta \in \mathbb{R}$ – David was influenced in his approach by Bochner's [4] book on harmonic analysis and probability, and Kolmogorov's [14] work on curves in Hilbert space. Under moment assumptions, he demonstrated that stationary interval processes admit a spectral representation analogous to that of Kolmogorov's for real processes with stationary increments. This representation was then used in conjunction with mixing assumptions (for well-separated intervals) to extend distributional asymptotics on finite Fourier transforms of continuous stationary processes to the interval function case. The exploitation of such results for statistical inference was then considered. Bartlett [1, 2, 3] had fairly recently initiated the frequency approach to inference on stationary point processes, and David's results appear to be the first instance of an asymptotic theory for spectral inference on general second-order stationary point processes (Bartlett [2] had only briefly described periodogram asymptotics in the special case of Cox processes on the real line). Thus David's work has been fundamental for inference in point processes – a search in the literature on formal inference for point processes (statistical tests, for example) reveals the significant extent to which his results have underpinned many methods that are being put to use. However, David's results apply more generally to other situations – for example, integrals of continous-time time series $\boldsymbol{X}(\Delta) = \int_\Delta \boldsymbol{Y}(t)dt$, and hybrids of continuous processes and point processes, such as $\boldsymbol{X}(\Delta) = \int_\Delta \boldsymbol{Y}(t)\Pi(dt)$. These special cases, including their regression-like manifestations (time-invariant filtering), are treated in some depth as particular examples of application of the general results. At the end of his talk, David made the throw-away remark that in much of what he was doing he was in essence replacing the empirical Fourier transform $\int_0^T e^{-i\lambda t} X_t dt$ by $\int_0^T e^{-i\lambda t} dX(t)$. Herman Rubin then approached him and asked "*why didn't you say that from the beginning?*".

On the number of solutions of systems of random equations [1972]

This is one of the first few papers that David wrote after having moved from the LSE to Berkeley in the late '60s. The problem under consideration was determining expressions for the expectation and factorial moments of the number of solutions of n real equations in n unknowns. For random fields that are almost everywhere differentiable, such expressions can be exploited in order to study the behavior of their critical points, or indeed provide bounds for crossing probabilities. Within the latter context, the results in David's paper can be interpreted as a generalisation of *Rice's formula* [19], applicable to the non-stationary and non-Gaussian vector field case. David was well aware of the work of a number of probabilists on crossings of stochastic processes at the time - Cyril Offord had become the first Professor of Mathematics at the LSE shortly before David's departure, and he had read his papers, including a famous one with Littlewood. David had also attended Leadbetter's Berkeley Symposium talk. Although he recalls that much of this he had done "*for amusement*", he then came across one of the new acquisitions in the Berkeley Maths library on Geometric Measure Theory, by H. Federer [6]. Skimming through the book, he realised that one of Federer's results on counting zeroes could be adapted into a stochastic variant and also be applied to level curves. For a random system $\boldsymbol{Y}(\boldsymbol{x}) = \boldsymbol{y}_0$, where $\boldsymbol{Y}$ is an $\mathbb{R}^n$–valued field and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, one may define a spatial point process $N(A; \boldsymbol{y}_0)$ over the Borel subsets of $\mathbb{R}^n$, which counts the number of solutions of the system falling in A. In this context, David generalised Kac's [11] approach (via Federer) to counting roots and showed that the mean measure, $\mathbb{E}[N(A)]$, can be expressed as $\int_{\mathbb{R}^n} \int_A |t| f(\boldsymbol{y}_0, t; \boldsymbol{x}) dt d\boldsymbol{x}$, where $f(\boldsymbol{y}, t; \boldsymbol{x})$ is the joint density function of $\boldsymbol{Y}(\boldsymbol{x})$ and $T(\boldsymbol{x}) = [J\boldsymbol{Y}](\boldsymbol{x})$, the latter denoting the Jacobian of $\boldsymbol{Y}$ at $\boldsymbol{x}$. Notice that, although $\mathbb{E}[N(A)]$ reflects an essentially global property of $\boldsymbol{Y}$ on A, this expression requires only the *marginal distribution* of the random $\mathbb{R}^{n+1}$–vector $(\boldsymbol{Y}(\boldsymbol{x}), T(\boldsymbol{x}))$ (which, of course, is then integrated w.r.t. $\boldsymbol{x}$). The expression for the factorial moments is of similar "philosophy". In the stationary non-Gaussian one-dimensional case, and taking $A = [0, 1]$, Rice's formula becomes $\mathbb{E}[N([0, 1])] = \int_0^\infty t f(y_0, t) dt$, where $f(y, t)$ is the joint density of Y and its derivative Y' – a special case of David's formula. Conditions for such a formula to hold true can be rather intricate and hard to verify, and David's paper appears to be the first to consider the validity of such types of formulae for almost all y_0 rather than for each specific value – allowing for great simplification of the sufficient conditions for validity (the requirement being Lipschitz continuity and existence of the joint density; similar approaches were subsequently considered by Geman & Horowitz [9] and later Zähle [22]). That being said, David also provided sufficient conditions in this paper for his formula to be valid for *any particular value* y_0. An interesting remark is that his acquaintance with Moe Hirsch, a Berkeley mathematician whom David acknowledges at the end of the paper for pointing out a useful lemma, was made during an anti-war demonstration at Berkeley.

Asymptotic Normality of Finite Fourier Transforms of Stationary Generalized Processes.[1982]

In one of our first meetings at Berkeley, during a general discussion on potential PhD thesis topics, I recall David stressing: *"you want to be thinking of data as anything that can be mathematically expressed – thinking of a stochastic process X_t as X and t being in arbitrary spaces with some general structure that can be taken advantage of"*. David points out (Panaretos [16]) that he was influenced to think this way by reading Bourbaki already as a first year undergraduate at Toronto, on the encouragement of A.J. Coleman. In many ways, this philosophy exemplifies a theme that characterises much of David's research, namely using mathematical machinery that might not be considered part of the standard statistical toolbox with a concrete data–analytic goal in mind. This paper –communicated by Murray Rosenblat– is an instance of this approach in one of David's main fields of expertise: the spectral analysis of stationary processes. Fourier transforming a process $\{X_t\}$ requires in principle X to have realisations lying in a suitable function space (ideally a Schwartz space), and t belonging in a space with a group structure (e.g. Rudin [20]). The latter also allowing to make sense of the notion of stationarity, as invariance under the action of the group onto itself). David considers stationary processes with a "time index" belonging in a locally compact Abelian topological group – quite a general setting extending the usual choices of $\mathbb{R}$ or $\mathbb{Z}$. The "paths" of the process are consequently assumed to belong to the class of Schwartz-Bruhat distributions, the generalisation of the notion of Schwartz distributions in the case where the domain is a locally compact Abelian group instead of a real vector space (Osborne [17]). In the more traditional setting of t in $\mathbb{R}^d$ or $\mathbb{Z}$, one has the celebrated result that the finite Fourier transform of a stationary mixing process is asymptotically (complex) Gaussian and independent for different frequencies – opening the door for the use of "iid technology" for statistical inferences based on observed sample paths. David had developed similar types of results for various types of stationary processes and was seeking in this paper to find a general unifying framework. He thus provided a mixing condition in this more abstract setting for analogous results to hold. The conditions and results require quite some technical prerequisites to be exposed, but can be seen to be similar in nature with the more "usual cases", once the concepts involved have been suitably generalised. An interesting aspect is how to make sense of the statement $t \to \infty$, which is formalised by David through suitable *nets* of tapers (as David notes in the paper, a Fourier transform not based on a net might fail to be Gaussian). When I asked David in the end of that meeting which courses he thought I should take, he replied "*sit in as many math courses as you can, and also take the applied statistics course*".

A particle migrating randomly on a sphere [1997]

In the mid '90's, David began a collaboration with Brent Stewart of the Hubbs-Sea World Research Institute on the modeling and analysis of the movement of elephant seals. Every year, elephant seals travel enormous distances migrating from the southern Californian coast to the northeastern Pacific. With the exception of minor perturbations due to foraging, their course seems to lie on great circles connecting their departure point and destination. Modeling their trajectories is of interest since it provides insight into whether the seals have the ability to somehow continually move "straight ahead" on a geodesic path – a ship, for instance, would need to iteratively correct its course to stay close to a geodesic path. In this paper, David set down the probabilistic framework for the analyses later conducted considering stochastic differential equations on the sphere attracted to a cap of the sphere. In a sense, his work can be seen to be the "spherical complement" of that of D.G. Kendall [12] who considered the scenario of an attractive pole on the plane in order to model bird navigation (indeed, David demonstrated how his model approximated that of Kendall's when the radius of the sphere grows to infinity). Diffusions on spheres and homogeneous spaces have a long history starting with Perrin's [18] investigation of *rotational Brownian motion* and have been the subject of much study (also see Kendall [13] and Hsu [10]). David first provided an overview of the approaches to Brownian motion on the sphere: intrinsic definition, definition through embedding in $\mathbb{R}^3$, definition through setting the spherical Laplacian as the generator, and definition through invariance arguments. He then proceeded to develop the Itô equations for a spherical diffusion drifting towards the north pole and subject to Brownian disturbances. From these, expressions for the generator, the invariant density and the likelihood ratio were determined. From the latter, David also proposed estimators for the parameters of interest when a complete path is observed and suggested approaches to estimation when only partial observation of the path is feasible (as would be the case in practice). This paper marks the beginning of a long research program of David's involving the use of stochastic differential equation-based models for animal movement – a program that he is still actively pursuing to date.

Some statistical methods for random process data from seismology and neurophysiology [1988]

Soon after arriving at Berkeley, David begun long and fruitful collaborations with a seismologist, Bruce Bolt, and a neurophysiologist, Jose Segundo, on statistical problems arising in their work. When he was asked to give the 1983 Wald Memorial Lectures, he decided to present an overview of the output of these collaborations with the aim of bringing out the unifying features of statistics, and the transfer of technology it enables. A central theme that permeates the paper is the use of spectral techniques. David placed significant emphasis on the need for the statistician to have a serious understanding of and genuine interest in the applied matter under study. The actual lectures were given in Toronto, and David recalls many old teachers as well as colleagues from Bell Labs and Princeton attending. It is worthwhile mentioning that, although the lectures were given in 1983, they start off by a phrase that one would usually expect to find in papers written about fifteen years later: "*the basic data are curves and surfaces. If* n *denotes the sample size and* p *denotes the dimension, then the concern is with the case of* n *much less than* p". The extent of his work in the areas is reflected in the staggering 54 page length of the overview – upon reading the print version Terry Speed remarked that never before had he seen so much material go into such lectures. Among the topics that David considers are: the use of asymptotic properties of Fourier transforms of stationary processes to estimate the frequencies in the Earth's free oscillations using nonlinear least squares, given knowledge of basic physical parameters describing the structure of the earth; the balancing of the bias/variance tradeoff as this manifests itself through regularisation in the corresponding inverse problem of estimating the perturbations of the previously known physical parameters given a perturbed estimate of the frequency of the free oscillation; the use of probit analysis in the estimation of fault plane parameters, with special interest in determining whether the first seismic motion was compression or dilation; the employment of non-linear probit models as a means to formally (as opposed to graphically) estimate basic parameters corresponding to an earthquake fault-plane given the direction of first motion of the earthquake and related quantities; the estimation of "size" related seismic quantities such as the seismic moment and stress drop using asymptotic results in spectral analysis, when noise in the spectral domain is heteroskedastic even asymptotically; the frequency-wavenumber analysis of seismic velocity vectors from non-stationary array data; the use of spectral deconvolution methods to attack the inverse problem of exploration seismology, i.e. the probing of important geological characteristics of a region (e.g. presence of gas or oil) given seismic data related to that region; the investigation of interactions within small networks of neurons by means of analysis of the cross intensity of the multidimensional point process of their firing times, and of regression modeling of the conditional intensity functions of these times, fitted by spectral methods; the elucidation of the direction of influence in such small neuronal networks by focusing on the partial coherency functions stemming from the joint point process model; maximum likelihood inference on the neuronal firing threshold and the firing filter in an integrate-and-fire setting, where current flow and firing pattern are recorded; and the use of filter identification to the aim of answering fundamental questions on the neuronal responses evoked by a variety of external stimuli. The paper concludes with the visionary statement that "*it seems that the future will see many of the traditional statistical techniques extended to apply to datum of more complicated forms – specifically to curves, moving surfaces, point clouds and the like.*"

The paper is dedicated to Jeff, David's elder son. Jeff died in 1988 after a fifteen year struggle with brain cancer. He was able to see the print copy before he passed away.

References

[1] Bartlett, M.S. (1963). Statistical estimation of density functions. *Sankhyā, Series A,* ***25****: 245–254.*

[2] Bartlett, M.S. (1963). The spectral analysis of point processes (with discussion). *Journal of the Royal Statistical Society, Series B,* ***25****: 264-296*

[3] Bartlett, M.S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika,* ***51****: 299–311.*

[4] Bochner, S. (1960). *Harmonic Analysis and the Theory of Probability.* University of California Press, Berkeley and Los Angeles.

[5] Breiman, L. (1968). *Probability.* Addison-Wesley, Reading.

[6] Federer, H. (1969). *Geometric Measure Theory.* Springer-Verlag, Berlin.

[7] Fisher, R.A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society,* ***26****: 528-535.*

[8] Gaenssler, P. & Stute, W. (1979). Empirical Processes: A Survey of Results for Independent and Identically Distributed Random Variables. *The Annals of Probability,* ***7****: 193–243.*

[9] Geman, D. & Horowitz, J. (1973). Occupation times for smooth stationary processes. *Annals of Probability,* ***1****: 131 137.*

[10] Hsu, E.P. (2002). *Stochastic Analysis on Manifolds.* AMS Graduate Series in Mathematics, Providence.

[11] Kac, M. (1943). On the average number of real roots of a random algebraic equation. *Bulletin of the American Mathematical Society,* ***49****: 314–320.*

[12] Kendall, D.G. (1974). Pole-seeking Brownian motion and bird navigation (with discussion). *Journal of the Royal Statistical Society, Series B, **36**: 365–417.*

[13] Kendall, W.S. (1987). Stochastic differential geometry: an introduction. *Acta Applicandae Mathematicae **9**: 29–60.*

[14] Kolmogorov, A.N. (1940). Curves in Hilbert space invariant with regard to a one parameter group of motions. *Dokl. Akad. Nauk SSSR **26**: 6–9.*

[15] Lindley, D.V. (1959). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society, Series B, **21**: 102-107.*

[16] Panaretos, V.M. (2010). A conversation with D.R. Brillinger. *Statistical Science (to appear).*

[17] Osborne, M.S. (1975). On the Schwartz-Bruhat space and the Paley-Wiener theorem for locally compact Abelian groups. *Journal of Functional Analysis **19**: 40–49.*

[18] Perrin, M.F. (1928). Mouvement Brownien de rotation. *Annales de l'Ecole Normale Supérieure **45**: 1–51.*

[19] Rice, S.O. (1945). Mathematical analysis of random noise. *Bell Syst. Tech. J. **24**: 46–156.*

[20] Rudin, W. (1962). *Fourier Analysis on Groups.* Wiley, New York.

[21] Shorack, G.R. & Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons, New York.

[22] Zähle, U. (1984). A general Rice formula, Palm measures and horizontal window conditioning for random fields. *Stochastic Processes and their Applications, **17**: 265–283.*

NECESSARY AND SUFFICIENT CONDITIONS FOR A STATISTICAL PROBLEM TO BE INVARIANT UNDER A LIE GROUP[1]

By David R. Brillinger

Bell Telephone Laboratories and Princeton University

1. Introduction and summary. Although a great deal has been written concerning the theory of tests, decisions and inference for statistical problems invariant under the action of some group, (see for example [4]–[7], [9], [12]–[14], [16]), no great amount of literature exists concerning the problem of discerning whether or not a given problem is actually invariant under some group. In fact the literature seems to consist of one abstract [8] and one paper [15].

In this paper necessary and sufficient conditions are developed that a statistical problem must satisfy in order that it be invariant, in a precise sense to be defined later, under a fairly general class of transformation groups, Lie transformation groups. It must be added, however, that the sufficient conditions are to some extent tautological. In addition two methods of actually constructing the group, if it can be shown to exist, are given, and the main theorem is illustrated by a variety of examples.

One of the examples yields the interesting result that the fiducial distribution of the correlation coefficient derived from a sample from a bivariate normal distribution by R. A. Fisher is not a Bayes' distribution for any prior distribution.

2. The definition of a Lie group. The following definitions are essential to what follows.

A group G is said to be a *transformation group* on the set E if G is a subgroup of the group of all $1 - 1$ mappings of E onto itself.

Let F be any subset of E, then the set of all elements gx for $g \varepsilon G$, $x \varepsilon F$, is called the *orbit* of F under G.

A *topological group* is a group which is also a Hausdorff space and the maps,

(i) $g, h \rightarrow gh: G \times G \rightarrow G$

(ii) $g \rightarrow g^{-1}: G \rightarrow G$

are continuous.

Let G be a topological group and X a Hausdorff space. Assume that for each $g \varepsilon G$ there exists a homeomorphism of X onto X, $T_g: X \rightarrow X: x \rightarrow \varphi(x, g) = T_g x$ such that

(i) $T_e = \text{identity} = I$, e the identity of G

(ii) $T_{g_1} T_{g_2} = T_{g_1 g_2}$

(iii) The function $(g, x) \rightarrow \varphi(x, g) : G \times X \rightarrow X$ is continuous, then $\varphi: G \times X$

Received November 3, 1961; revised December 27, 1962.

[1] Part of this research was carried out while the author was at the London School of Economics with the support of a Research Training Fellowship of the Social Science Research Council.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_2, © Springer Science+Business Media, LLC 2012

$\to X$ is called a *topological transformation group* of G acting on X by the function φ.

If in addition,

(iv) $T_g = I \Leftrightarrow g = e$, then G acts *effectively* on X.

A topological transformation group G is *transitive* on a space X if for every $x, y \in X$ there is a $g \in G$ such that $gx = y$. The space X is then called a *homogeneous* space.

In what follows, differentiable should always be understood to refer to that of class C^∞.

An *n-dimensional manifold* M^n, is a Hausdorff space which is locally n-dimensional Euclidian at each point. Because M^n is locally Euclidian, every point p has a neighborhood with a system of coordinates $x_1^p, \cdots, x_n^p$.

A manifold, together with a set of overlapping coordinate systems, which cover the entire manifold and has the property that the transformation between any two overlapping coordinate systems is differentiable is called a *differentiable manifold.*

A *Lie group* is a topological group which is also a differentiable manifold and such that,

(i) $g, h \to gh: G \times G \to G$

(ii) $g \to g^{-1}: G \to G$

are differentiable maps.

Let G be a Lie group and M^n a differentiable manifold. Assume that for each $g \in G$ there exists a diffeomorphism of M^n onto M^n, $T_g : M^n \to M^n: x \to \varphi(x, g)$ such that:

(i) $\varphi: G \times M^n \to M^n$ is a topological transformation group,

(ii) The function $(g, x) \to \varphi(x, g): G \times M^n \to M^n$ is differentiable, then $\varphi: G \times M^n \to M^n$ is a *Lie transformation group.*

Examples of Lie transformation groups include all affine transformations of R^n and all orthogonal transformations of R^n.

Associated with any Lie transformation group $G: M \to M$, is a set of infinitesimal generators defined as follows: the mapping $(g, x) \to gx = y$, $g \in G$, $x, y \in M$ is differentiable. Let $(g) = (g^1, g^2, \cdots, g^r)$ denote a set of coordinates at $e \in G$ and $(x) = (x^1, x^2, \cdots, x^n)$ be a set of coordinates at some point x_0 of M. The mapping can now be expressed as follows, $y^\alpha = \varphi^\alpha(g, x)$ $\alpha = 1, \cdots, n$, with φ^α differentiable. Put

$$\psi_i^\alpha(x) = [\partial\varphi^\alpha(g, x)/\partial g^i]_{g=e}$$

$X_i = \sum_\alpha \psi_i^\alpha(x)(\partial/\partial x^\alpha)$ is called an *infinitesimal generator* of the Lie transformation group G over the manifold M. It is a differential operator acting on the space of differentiable functions of x.

Lie groups appear to be natural groups for a statistician to be concerned with. They are locally compact and hence possess the Haar measure required in the Hunt-Stein Theorem [9] or in Fraser's work on fiducial probability, [4], [5] for example. The transformations they induce are defined on a manifold, the type of space, the sample space or parameter space usually is in a statistical problem. Finally the transformations induced by a Lie group are continuous, a property that seems sensible for transformations applied to random variables or parameters in a real problem.

3. An invariant statistical problem. In what follows the essential components of a statistical decision problem will be defined and necessary and sufficient conditions that such a problem be invariant given, for most statistical problems are generally made up of some of these basic components.

Let $\mathfrak{X}$ be the set of all experimental outcomes, $B_{\mathfrak{X}}$ a σ-algebra of subsets of $\mathfrak{X}$, Θ the set of states of nature, and $P(\cdot,\cdot)$, the specification, a real-valued function defined on $B_{\mathfrak{X}} \times \Theta$, such that for each $\theta \varepsilon \Theta$, $P(\cdot, \theta)$ is a probability measure on $B_{\mathfrak{X}}$. Let A be the set of actions available to the statistician, B_A a σ-algebra of subsets of A, and L, the loss function, a real-valued function on $\Theta \times A \times \mathfrak{X}$ such that $L(\theta, a, x)$ is the loss to the statistician when he takes action a, after observing x, and θ is the true state of nature. L is assumed jointly measurable in a and x. Let D be the class of randomized decision functions from $\mathfrak{X} \times B_A$ into the unit interval, such that for each x, δ_x is a probability measure on (A, B_A) and such that for fixed a, δ_a is a measurable function of x. The associated risk $R(\theta, \delta)$ for $\theta \varepsilon \Theta$ is defined by,

$$R(\theta, \delta) = \iint L(\theta, a, x)\, d(x, a)\, dP(x, \theta).$$

The above statistical decision problem is said to be invariant under the group G if,

(i) G is a transformation group on each of $\mathfrak{X}$, Θ, A,

(ii) $g: \mathfrak{X} \to \mathfrak{X}$ and $g: A \to A$ are each measurable transformations on the respective spaces,

(iii) $P(gB, g\theta) = P(B, \theta), B \varepsilon B_{\mathfrak{X}}, \theta \varepsilon \Theta$ where g belongs to the transformation group G on $\Theta \times A \times \mathfrak{X}$ obtained by defining $g(\theta, a, x) = (g\theta, ga, gx)$ and

(iv) $L(g\theta, ga, gx) = L(\theta, a, x)$.

If in addition $g\delta = \delta$ for all $g \varepsilon G$ the decision procedure δ will be said to be invariant under G.

The goal of this paper is to find necessary and sufficient conditions that a statistical decision problem, in which Θ, A, and $\mathfrak{X}$ are differentiable manifolds, must satisfy, in order that it be invariant under the action of some Lie transformation group.

The following lemmas will be required later.

LEMMA 1. *If G is a Lie transformation group over the manifolds, Θ, A, $\mathfrak{X}$ its infinitesimal generators, when it is considered a Lie transformation group over $\Theta \times A \times \mathfrak{X}$, are of the form,*

$$X_i = \sum_\alpha \psi_i^\alpha(\theta) \frac{\partial}{\partial \theta^\alpha} + \sum_\beta \rho_i^\beta(a) \frac{\partial}{\partial a^\beta} + \sum_\gamma \sigma_i^\gamma(x) \frac{\partial}{\partial x^\gamma},$$

i.e., X_i is the sum of the infinitesimal generators of G over Θ, A, $\mathfrak{X}$ respectively.

PROOF. This is an immediate result of the particular nature of,

$$G: \Theta \times A \times \mathfrak{X} \to \Theta \times A \times \mathfrak{X}.$$

LEMMA 2. *Let F be any real-valued differentiable function defined on $\Theta \times A \times \mathfrak{X}$. If F is invariant under the Lie group G, i.e., $F(g\theta, ga, gx) = F(\theta, a, x)$ for all $g \varepsilon G$, then $X_i F \equiv 0$, where the X_i are the infinitesimal generators of G. If $X_i F \equiv 0$ and G is connected, then the converse is true, namely F is invariant under G.*

PROOF. If $F(\theta, a, x)$ is invariant under G then $F(g\theta, ga, gx)$ is independent of g. That is, $\partial F(g\theta, ga, gx)/\partial g^i = 0$. But $[\partial F(g\theta, ga, gx)/\partial g^i]_{g=e} = X_i F$, thus $X_i F \equiv 0$ for all i.

Conversely suppose $X_i F \equiv 0$ for all i. Now,

$$F(gh\theta_0, gha_0, ghx_0) = F(g\theta_1, ga_1, gx_1),$$

where $\theta_1 = h\theta_0$, $a_1 = ha_0$, $x_1 = hx_0$. Thus

$$[\partial F(gh\theta_0, gha_0, ghx_0)/\partial g^j]_{g=e} = [X_j F]_{\theta=\theta_1, a=a_1, x=x_1} = 0.$$

Now

$$\begin{aligned} \partial F(h\theta_0, ha_0, hx_0)/\partial h^j &= [\partial F(gh\theta_0, gha_0, ghx_0)/\partial (gh)^j]_{g=e} \\ &= \sum_i \{[\partial F(gh\theta_0, gha_0, ghx_0)/\partial g^i] \cdot [\partial g^i/\partial (gh)^j]\}_{g=e}. \end{aligned}$$

Thus $\partial F(h\theta_0, ha_0, hx_0)/\partial h^j = 0$ for all j, and since G is connected F is invariant.

The main theorem of the paper is the following.

THEOREM. *Let the random variable X have a density $f(x, \theta)$ with respect to Lebesgue measure, and let f and the loss function L be differentiable functions. If the statistical decision problem is invariant under a Lie transformation group G then,*

$$(1) \quad \sum_\alpha \psi_i^\alpha(\theta) \frac{\partial L}{\partial \theta^\alpha} + \sum_\beta \rho_i^\beta(a) \frac{\partial L}{\partial a^\beta} + \sum_\gamma \sigma_i^\gamma(x) \frac{\partial L}{\partial x^\gamma} = 0 \qquad \textit{for all } i,$$

$$(2) \qquad \sum_\alpha \psi_i^\alpha(\theta) \frac{\partial \log f}{\partial \theta^\alpha} + \sum_\gamma \sigma_i^\gamma(x) \frac{\partial \log f}{\partial x^\gamma} = -\sum_\gamma \frac{\partial \sigma_i^\gamma(x)}{\partial x^\gamma} \qquad \textit{for all } i,$$

where the linear differential operators

$$(3) \qquad X_i = \sum_\alpha \psi_i^\alpha(\theta) \frac{\partial}{\partial \theta^\alpha} + \sum_\beta \rho_i^\beta(a) \frac{\partial}{\partial a^\beta} + \sum_\gamma \sigma_i^\gamma(x) \frac{\partial}{\partial x^\gamma}$$

$\alpha = 1, \cdots, s$; $\beta = 1, \cdots, t$; $\gamma = 1, \cdots, u$; $i = 1, \cdots, v$ are the infinitesimal generators of G.

Conversely, if there exist differentiable functions, $\psi_i^\alpha(\theta)$, $\rho_i^\beta(a)$, $\sigma_i^\gamma(x)$; $\alpha = 1, \cdots, s$; $\beta = 1, \cdots, t$; $\gamma = 1, \cdots, u$; $i = 1, \cdots, v$ such that

(1′) *(1) and (2) above are satisfied*

(2′) *there exist constants c_{jk}^i $i, j, k = 1, \cdots, r$ such that $X_k X_j - X_j X_k = \sum_i c_{jk}^i X_i$ the X_i being given by (3)*

(3′) *the X_i generate a connected group G, then, the statistical problem defined previously is invariant under G.*

PROOF. The measure induced by the density function $f(x, \theta)$ is invariant under the action of G, if and only if $f(x, \theta) = f(gx, g\theta)$ Det $(\partial\varphi^\alpha(g, x)/\partial x^\gamma)$.

Differentiating this relation with respect to g^i and setting $g = e$ one obtains,

$$0 = X_i f + f[\sum_\gamma (\partial\sigma_i^\gamma(x)/\partial x^\gamma)]$$

This is equivalent to (2). (1) follows immediately from Lemma 2, completing the proof of the necessary part of the theorem.

It was stated in the introduction of the paper that the sufficient conditions that would be developed were somewhat tautological. The following discussion should indicate the reason for this statement.

In the second lemma it was shown that if the differentiable function $F(\theta, a, x)$ is invariant under the Lie group G, then it must be such that $X_i F = 0$ for X_i any infinitesimal generator of G. It is not however true that if a function F is such that $X_i F \equiv 0$ for a set of linear differential operators of the form under consideration, then there is a Lie group of transformations G under which F is invariant. The linear operators may not generate a (global) group of transformations, but only a local one (i.e., one for which inverses etc. are only defined in a neighborhood), or perhaps nothing significant at all. The condition (2′) above is the necessary and sufficient condition that a local group be generated (this is Lie's Second Fundamental Theorem). Necessary and sufficient conditions have been given for a global group to be generated (see [11]), however these conditions do not seem to be easy to apply. If one does have a particular set of linear operators satisfying (2′) the most efficient method of finding out if they generate a global group appears to be to generate the local group and then to check to see if it is actually global. Two methods of doing this follow later.

The sufficiency part of the theorem relative to the loss function now follows from the above considerations and Lemma 2. To complete the proof of the theorem it must now be shown that the probability measure induced by $f(x, \theta)$ is invariant or that, $f(x, \theta) = f(hx, h\theta)$ Det $(\partial\varphi^\alpha(h, x)/\partial x^\gamma)$ for all h.

Now,

$$\begin{aligned}\partial f(hx, h\theta)/\partial h^j &= [\partial f(ghx, gh\theta)/\partial(gh)^j]_{g=e} \\ &= \sum_i \{[\partial f(ghx, gh\theta)/\partial g^i]\cdot[\partial g^i/\partial(gh)^j]\}_{g=e} = \sum_i [\partial g^i/\partial(gh)^j]_{g=e}[X_i f]_{x=x_1, \theta=\theta_1}\end{aligned}$$

where $x_1 = hx$, $\theta_1 = h\theta$.

Similarly,

$$\begin{aligned}&(\partial/\partial h^j) \text{ Det } (\partial\varphi^\alpha(h, x)/\partial x^\gamma) \\ &\qquad = \sum_i [\partial g^i/\partial(gh)^j]_{g=e}\big(\sum_\gamma \sigma_i^\gamma(hx)\big) \text{ Det } (\partial\varphi^\alpha(h, x)/\partial x^\gamma).\end{aligned}$$

Using the above relations and (2),

$$\frac{\partial}{\partial h^j}\{f(hx, h\theta)\ \mathrm{Det}\,(\partial\varphi^\alpha(h, x)/\partial x^\gamma)\} = 0.$$

Therefore $f(hx, h\theta)$ Det $(\partial\varphi^\alpha(h, x)/\partial x^\gamma)$ is a constant with respect to h, and setting $h = e$, and using the fact that G is connected, it equals $f(x, \theta)$. The proof of the theorem is now completed.

The theorem has the following corollary for the case in which X is a real random variable with cdf $F(x)$.

Corollary. *Let X be a real random variable with* c.d.f. $F(x, \theta)$. *The statistical problem is invariant under the connected group G if and only if $F_x/F_\theta = a(\theta)/b(x)$* i.e., *$F_x/F_\theta$ factors into a function of x times a function of θ and $b(x)(\partial/\partial x) + a(\theta)(\partial/\partial\theta)$ generates G.*

Proof. This corollary follows immediately from Lemma 2 and the fact that for a real random variable invariance of the probability measure and invariance of the c.d.f. are equivalent. This corollary is inherent in [10].

The reader will have noted that derivatives of all orders have been assumed to exist in the definition of a Lie transformation group. One can in fact proceed with fewer derivatives than this. Minimal conditions are a topic of current research in the theory of Lie groups.

Two methods of actually generating the local group from a given set of linear operators now follow:

Method 1. Let the local group G be generated by the r linear operators X_1, $\cdots$, X_r. If $\bar{x} = g(x)$, then,

$$\bar{x}^i = \exp\,(g^1X_1 + g^2X_2 + \cdots + g^rX_r)x^i \quad \text{for} \quad i = 1, \cdots, n,$$

i.e.,

$$\bar{x}^i = \sum_m [(g^1X_1 + \cdots + g^rX_r)^m/m!]\, x^i$$

Method 2. Let the local group G be generated by $X_1, \cdots, X_r$ once again. Find the integrals $\varphi_1, \cdots, \varphi_n$ of $(g^1X_1 + \cdots + g^rX_r)u = 1$ then solve

$$\varphi_i(\bar{x}^1, \cdots, \bar{x}^n) = \varphi(x^1, \cdots, x^n) + 1 \qquad i = 1, \cdots, n$$

for $\bar{x}$.

The following example illustrates the first method and the corollary.

Example 1. Suppose one is concerned with the Pareto distribution $F(x) = 1 - (1/x^\alpha)$ $x \geqq 1$ $\alpha > 0$, i.e., one wishes to find if it is invariant under some group. Now $F_x/F_\alpha = (\alpha/x) \ln x$, i.e., F_x/F_α factors as required; therefore $F(x)$ may be invariant under some local Lie transformation group. Let us try to generate this group by Method 1 given above.

Its infinitesimal generator is $X = -\alpha(\partial/\partial\alpha) + x \ln x(\partial/\partial x)$, $\bar{\alpha} = \exp\{-g[\alpha\ \partial/\partial\alpha - x \ln x\ (\partial/\partial x)\ ]\alpha\} = \exp[-g\alpha\ \partial/\partial\alpha]\alpha = \alpha - g\alpha + g^2\alpha/2! - g^3\alpha/3! \cdots = \alpha e^{-g}$, $\bar{x} = \exp[gx \ln x\ (\partial/\partial x]x = \sum_m(1/m!)[gx \ln x(\partial/\partial x)]^m x$.

Let $\ln \ln x = y$, $\bar{x} = \sum_m (1/m!)(g\, \partial/\partial y)^m \exp(e^y) = \exp(e^{y+g}) = x^{e^g}$, i.e., the local group is given by $\bar{\alpha} = c\alpha$, $\bar{x} = x^{1/c}$, $c > 0$, which is actually a group.

EXAMPLE 2. A problem that is of interest in the field of statistical inference is to find out if the c.d.f. of the correlation coefficient r, estimated from a sample of size N from a bivariate normal distribution with correlation coefficient ρ is invariant under some group of transformations. This problem is of interest for two reasons, first if such a group exists and it is locally compact the Haar measure on it provides a particularly appealing prior measure to use in the application of Bayes' theorem. Secondly if the group exists and it satisfies certain properties a fiducial distribution for ρ may be constructed following Fraser [4] [5] and it would be of interest to compare this fiducial distribution with the one given by Fisher in his original paper on fiducial probability [3].

The density of the correlation coefficient may be written as,

$$(4) \qquad f_n(r, \rho) = \frac{2^{n-2}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{(n-2)!\pi} \sum_{\alpha=0}^{\infty} \frac{(2\rho r)^\alpha}{\alpha!} \Gamma^2[\tfrac{1}{2}(n+\alpha)]$$

where $n = N - 1$. (See [1].)

The c.d.f. will be invariant under a Lie transformation group only if there exist differentiable functions $a(r)$, $b(\rho)$ such that,

$$\frac{d}{dr}[a(r) f_n(r, \rho)] = b(\rho) \frac{d}{d\rho} f_n(r, \rho).$$

Write f_n as, $K_n(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}C(2\rho r)$. Therefore

$$(5) \qquad \begin{aligned} &(1-\rho^2)^{\frac{1}{2}n} \frac{d}{dr}[a(r)(1-r^2)^{\frac{1}{2}(n-3)}C(2\rho r)] \\ &\quad = b(\rho)(1-r^2)^{\frac{1}{2}(n-3)}[(1-\rho^2)^{\frac{1}{2}n}C'(2\rho r)2r - (1-\rho^2)^{\frac{1}{2}n-1}n\rho C(2\rho r)]. \end{aligned}$$

Set $\rho = 0$

$$C(0)\frac{d}{dr}[a(r)(1-r^2)^{\frac{1}{2}(n-3)}] = b(0)(1-r^2)^{\frac{1}{2}(n-3)}C'(0)2r.$$

Therefore,

$$C(0)a(r)(1-r^2)^{\frac{1}{2}(n-3)} = -b(0)(1-r^2)^{\frac{1}{2}(n-1)}C'(0)2/(n-1) + K.$$

Equation (5) becomes

$$(6) \quad \begin{aligned} &(1-\rho^2)[b(0)C'(0)2rC(2\rho r)/C(0) \\ &\quad - b(0)(1-r^2)C'(0)4\rho C'(2\rho r)/C(0)(n-1) + 2\rho K C'(2\rho r)/C(0)] \\ &\qquad = b(\rho)[(1-\rho^2)C'(2\rho r)2r - n\rho C(2\rho r)]. \end{aligned}$$

Set $r = 0$, yielding $b(\rho) = (1-\rho^2) \times$ a constant.

Setting $\rho = 0$ the constant $= b(0)$, i.e., $b(\rho) = (1-\rho^2)b(0)$. Substitute in (6),

(7) $$b(0)C'(0)2rC'(2\rho r)/C(0) - b(0)\ (1 - r^2)C'(0)4\rho C'(2\rho r)/C(0)(n - 1) + 2\rho KC'(2\rho r) = (1 - \rho^2)b(0)[(1 - \rho^2)C'(2\rho r)2r - n\rho C(2\rho r)].$$

Equating the coefficient of ρr^2 on both sides of (7),

$$b(0)C'(0)\Gamma^2[\tfrac{1}{2}(n + 1)]/C(0) + b(0)C'(0)\Gamma^2[\tfrac{1}{2}(n + 1)]/C(0)(n - 1) = b(0)\Gamma^2[\tfrac{1}{2}(n + 2)].$$

Therefore $b(0) = 0$ or,

$$\Gamma^4[\tfrac{1}{2}(n + 1)]n/(n - 1) = \Gamma^2[\tfrac{1}{2}(n + 2)]\Gamma^2[\tfrac{1}{2}n].$$

The latter is easily seen to be impossible, consequently $b(0) = 0$ implying $b(\rho) = 0$. It therefore follows that the c.d.f. of the correlation coefficient is invariant under no Lie transformation group.

Lindley in [10] proved a theorem to the effect that for real random variables a fiducial distribution is a Bayes' posterior distribution if and only if the problem is invariant.

Applying this theorem here one can now say that the fiducial distribution for ρ is not a Bayes' posterior distribution for any prior distribution.

This example also demonstrates that a possibility suggested to me by Dr. J. Berkson is not true in general. Namely, that when the fiducial distribution provides a frequency interpretable probability, that probability is the one given by the Bayes' formula with a uniform distribution of the prior probabilities.

Example 3. In [2] an example is given to demonstrate that the above mentioned theorem of Lindley does not extend to spaces of dimension higher than one. Unfortunately the theorem quoted in [2], from which it follows that the given example is in fact a counterexample, is not general enough. The theorem proved above is of sufficient generality. Doubtlessly the given example still provides a counterexample; however it seems useful to give an example to which the above theorem may be applied more easily.

Consider

$$f(x, y, \alpha, \beta) = (2\pi)^{-1}\sigma(\alpha)\ \exp[-\tfrac{1}{2}(x - \alpha)^2]\ \exp[-\tfrac{1}{2}\sigma^2(\alpha)(y - \beta)^2]$$

where $\sigma(\alpha)$ is selected as in [2]. One may easily verify that Condition (1) of the theorem above leads to a contradiction.

I would like to thank David Lowdenslager for looking over a section of this paper, and the referee for a number of suggestions.

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] Brillinger, D. R. (1962). Examples bearing on the definition of fiducial probability with a bibliography. *Ann. Math. Statist.* **33** 1349–1355.

[3] Fisher, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535.

[4] FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.
[5] FRASER, D. A. S. (1961). On fiducial inference. *Ann. Math. Statist.* **32** 661–676.
[6] KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.
[7] KUDO, H. (1955). On minimax invariant estimates of the transformations parameter. *Nat. Sci. Rep. Ochanomizu Univ.* **6** 31–73.
[8] LEHMANN, E. L. and STEIN, C. (1953). The totality of transformations leaving a family of normal distributions invariant (abstract). *Ann. Math. Statist.* **24** 142.
[9] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
[10] LINDLEY, D. V. (1959). Fiducial distributions and Bayes' theorem. *J. Roy. Stat. Soc. Ser. B* **21** 102–107.
[11] PALAIS, R. S. (1957). *Global Formulation of the Lie Theory of Transformation Groups*. Amer. Math. Soc. Memoir.
[12] PEISAKOFF, M. (1951). Transformation parameters. Unpublished Princeton thesis.
[13] PITMAN, E. J. G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215.
[14] PITMAN, E. J. G. (1939). Location and scale parameters. *Biometrika* **30** 391–421.
[15] WENDEL, J. G. (1957). Invariance of normal distributions. *Michigan Math. J.* **4** 173–174.
[16] WESLER, O. (1959). Invariance theory and a modified minimax principle. *Ann. Math. Statist.* **30** 1–20.

AN ASYMPTOTIC REPRESENTATION OF THE SAMPLE DISTRIBUTION FUNCTION

BY DAVID R. BRILLINGER

Communicated by David Blackwell, January 10, 1969

1. Let $X_1, \cdots, X_n$ be independent observations from the uniform distribution on $[0, 1]$. Let $F_n(x) =$ the proportion of the $X_j \leqq x$. We will prove

THEOREM. *There is a random function* $\{G_n(x);\ 0 \leqq x \leqq 1\}$, *with the same distribution as* $\{F_n(x);\ 0 \leqq x \leqq 1\}$ *for each* n, *and there is a Brownian motion* W, *such that for the Brownian* $B(x) = n^{-1/2}W(nx)$

$$\sup_{0 \leqq x \leqq 1} \left| n^{1/2}[G_n(x) - x] - [B(x) - xB(1)] \right| = O[n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}] \tag{1}$$

almost surely as $n \to \infty$.

This theorem is of use in the investigation of the asymptotic behavior of functionals of $\{F_n(x);\ 0 \leqq x \leqq 1\}$, especially functionals dependent on n.

2. We construct $G_n(x)$ as follows; let $Y_1, Y_2, \cdots$ be independent exponential variables with mean 1. Let $S(k) = Y_1 + \cdots + Y_k$, $k = 1, 2, \cdots$ and let $S(0) = 0$. Set

$$G_n(x) = k/n \qquad \text{if } S(k)/S(n+1) \leqq x < S(k+1)/S(n+1).$$

This $\{G_n(x);\ 0 \leqq x \leqq 1\}$ has the same distribution as $\{F_n(x);\ 0 \leqq x \leqq 1\}$ for each n. We now record a series of lemmas.

LEMMA 1. *There is a Brownian motion* W *such that*

$$\sup_{1 \leqq k \leqq n} \left| k - S(k) - W(k) \right| = O[n^{1/4}(\log n)^{1/2}(\log \log n)^{1/4}] \tag{2}$$

almost surely as $n \to \infty$.

PROOF. This result is deducible from Theorem 1.5 of Strassen [8].

LEMMA 2. *Almost surely as* $n \to \infty$

$$\sup_{0 \leqq x \leqq 1} \left| S(nG_n(x)) - xS(n+1) \right| = O[n^{1/4}]. \tag{3}$$

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_3, © Springer Science+Business Media, LLC 2012

PROOF.

$$\begin{aligned}
| S(nG_n(x)) - xS(n+1) | & \\
&= | S(k) - xS(n+1) | \qquad \text{if } S(k) \leqq xS(n+1) < S(k+1) \\
&\leqq S(k+1) - S(k) \qquad \text{if } S(k) \leqq xS(n+1) < S(k+1). \\
&\leqq \max_{1 \leqq k \leqq n} Y_k
\end{aligned}$$

and one sees, by elementary calculations, that this last $= O[n^{1/4}]$ almost surely as $n \to \infty$.

LEMMA 3. *Almost surely as* $n \to \infty$

$$\sup_{0 \leqq x \leqq 1} | nG_n(x) - S(nG_n(x)) - W(nG_n(x)) | = O[n^{1/4}(\log n)^{1/2}(\log \log n)^{1/4}]. \tag{4}$$

PROOF.

$$\begin{aligned}
| nG_n(x) - S(nG_n(x)) - W(nG_n(x)) | & \\
&= | k - S(k) - W(k) | \qquad \text{if } S(k) \leqq xS(n+1) < S(k+1) \\
&\leqq \sup_{1 \leqq k \leqq n} | k - S(k) - W(k) |
\end{aligned}$$

and (4) follows from (2).

LEMMA 4. *Almost surely as* $n \to \infty$

$$\sup_{0 \leqq x \leqq 1} | G_n(x) - x | = O[n^{-1/2}(\log \log n)^{1/2}]. \tag{5}$$

PROOF. See Theorem 2* in Chung [3].

We next define the Brownian motion B by $B(x) = n^{-1/2}W(nx)$ and then have

LEMMA 5. *Almost surely as* $n \to \infty$

$$\sup_{0 \leqq x \leqq 1} | B(G_n(x)) - B(x) | = O[n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}]. \tag{6}$$

PROOF. (6) follows from (5) and Lévy's Hölder condition for Brownian motion (see Itô and McKean [4]) extended to apply to the interval $[0, n]$.

PROOF OF THEOREM. Up to an error term

$$O[n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4}],$$

that is uniform in x, almost surely as $n \to \infty$

$$\begin{aligned} n^{1/2}G_n(x) &= n^{-1/2}S(nG_n(x)) + n^{-1/2}W(nG_n(x)) \qquad \text{from (4)},\\ &= n^{-1/2}xS(n+1) + B(G_n(x)) \qquad \text{from (3)},\\ &= n^{-1/2}x[(n+1) - W(n+1)] + B(x) \qquad \text{from (2) and (6)},\\ &= n^{1/2}x - xB(1) + B(x), \end{aligned}$$

giving (1).

3. We may use the probability integral transformation to deduce a representation of the sample distribution function of observations from any continuous distribution. The results of Rosenkrantz [7] may be adapted to obtain rates of convergence in distribution for certain functionals of $F_n(x)$. The announcement of Kiefer [5] suggests that the error term in (1) may be best possible.

Bickel [1] and Billingsley [2] consider the weak convergence of the process $n^{1/2}[F_n(x)-x]$ to $W(x)-xW(1)$. Pyke and Root [6] let the distribution of Y depend on n and then prove

$$\sup_{0 \leqq x \leqq 1} \left| n^{1/2}[G_n(x) - x] - [W(x) - xW(1)] \right| = o(1)$$

almost surely as $n \to \infty$. I would like to thank Professor Pyke for the remark that B, as constructed above, depends on n.

References

1. P. J. Bickel, *Some contributions to the theory of order statistics*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), vol. I: *Statistics*, Univ. of California Press, Berkeley, Calif., 1967, pp. 575–591.

2. P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1968.

3. K. L. Chung, *An estimate concerning the Kolmogorov limit distribution*, Trans. Amer. Math. Soc. 67 (1949), 36–50.

4. K. Itô and H. P. McKean, Jr., *Diffusion processes and their sample paths*, Springer-Verlag, Berlin, 1965.

5. J. Kiefer, *The deviations in Skorokhod-Strassen approximation*, Notices Amer. Math. Soc. 15 (1968), 936–937.

6. R. Pyke and D. Root, *An application of stopping times to obtain weak convergence*, Technical Report No. 16, University of Washington, Seattle, Wash., 1968.

7. W. A. Rosenkrantz, *On rates of convergence for the invariance principle*, Trans. Amer. Math. Soc. 129 (1967), 542–552.

8. V. Strassen, *Almost sure behavior of sums of independent random variables and martingales*, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), vol. II: *Contributions to probability theory*, Part 1, Univ. of California Press, Berkeley, Calif., 1967, pp. 315–343.

The London School of Economics and Political Science,
Aldwych, England

THE SPECTRAL ANALYSIS OF STATIONARY INTERVAL FUNCTIONS

DAVID R. BRILLINGER
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction and summary

We consider stationary, additive, interval functions $\mathbf{X}(\Delta)$. These are vector valued stochastic processes having real intervals $\Delta = (\alpha, \beta]$ as domain, having finite dimensional distributions invariant under time translation and satisfying

$$\mathbf{X}(\Delta_1 \cup \Delta_2) = \mathbf{X}(\Delta_1) + \mathbf{X}(\Delta_2), \tag{1.1}$$

for disjoint intervals Δ_1, Δ_2. Such processes are considered in some detail in Bochner [5]. Setting

$$\mathbf{X}(t) = \mathbf{X}(0, t], \tag{1.2}$$

$-\infty < t < \infty$, and in the reverse direction setting

$$\mathbf{X}(\alpha, \beta] = \mathbf{X}(\beta) - \mathbf{X}(\alpha), \tag{1.3}$$

we see that a consideration of stationary interval functions is equivalent with a consideration of processes $\mathbf{X}(t)$, $-\infty < t < \infty$, having stationary increments. These last are discussed in Yaglom [24] for example. Important examples of processes of the type under consideration are provided by the point processes. Here the components of $\mathbf{X}(\Delta)$ give the number of events of various sorts that occur in the interval Δ. A variety of properties and applications of point processes may be found in Cox and Lewis [11], Bartlett [4], and Srinivasan [21].

The paper is divided into various sections. In Section 2 we introduce a key assumption for the processes; specifically we require that all the moments of $\mathbf{X}(\Delta)$ exist and have particular integral representations. We are then able to define

$$f_{a_1, \cdots, a_k}(\lambda_1, \cdots, \lambda_k), \tag{1.4}$$

$-\infty < \lambda_j < \infty, a_1, \cdots, a_k = 1, \cdots, r$, the cumulant spectra of order k of the r vector valued $\mathbf{X}(\Delta)$. These turn out to be generalizations of the cumulant spectra of order k of a continuous time series discussed in Brillinger and Rosenblatt [9]. We then present a spectral representation for $\mathbf{X}(\Delta)$. This representation was introduced in Kolmogorov [17] for real valued processes with stationary increments. It takes the form

$$\mathbf{X}(0, t] = \int_{-\infty}^{\infty} \left[\frac{\exp\{i\lambda t\} - 1}{i\lambda}\right] d\mathbf{Z}_X(\lambda), \tag{1.5}$$

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_4, © Springer Science+Business Media, LLC 2012

with $\mathbf{Z}_X(\lambda)$, $-\infty < \lambda < \infty$, a vector valued process. The process $\mathbf{Z}_X(\lambda)$ relates to the cumulant spectrum (1.4) through the expression

$$\text{cum}\,\{dZ_{a_1}(\lambda_1), \cdots, dZ_{a_k}(\lambda_k)\} = \delta\Big(\sum_1^k \lambda_j\Big) f_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_k)\, d\lambda_1, \cdots, d\lambda_k \tag{1.6}$$

with $\delta(\lambda)$ denoting the Dirac delta function.

In Section 3 of the paper we indicate how the theory developed applies to integrated continuous time series, to point processes, and to processes that are hybrids of these last two. In the case of point processes we relate the cumulant spectra to important parameters that have been introduced by Bahba [1] and by Kuznetsov and Stratonovich [18].

Section 4 of the paper discusses various asymptotic properties of the statistic

$$\mathbf{d}_X^{(T)}(\lambda) = \int_0^T \exp\{-i\lambda t\}\, d\mathbf{X}(t) \tag{1.7}$$

based on an observed stretch of an $\mathbf{X}(\Delta)$ process. It will be seen to behave in a similar manner to the finite Fourier transform of a stretch of a continuous stationary series. It follows that the estimates of the various cumulant spectra of $\mathbf{X}(\Delta)$ may be formed in the manner of Brillinger and Rosenblatt [9] and that the properties developed in that paper, such as asymptotic normality, will continue to hold. A selection of results that therefore become available is provided. In particular results relating to the linear time invariant regression of one stationary interval function on another are given. Because point processes are particular cases of the processes under consideration it follows that an asymptotic theory for the spectral estimates of order two of point processes has now been provided.

In Section 5 we apply the previously mentioned asymptotic results to develop estimates of the parameters suggested by Bahba and by Kuznetsov and Stratonovich for point processes. In Section 6 we consider the problem of the estimation of the second order spectra of a continuous time series when its values are available only for random times that are the occasions of events of an independent point process.

Section 7 discusses briefly some practical implications and extensions of the results of the paper. The proofs of the various lemmas and theorems of the paper are given in Section 8.

I would like to thank P. A. W. Lewis for a variety of helpful comments on the point process sections of the paper.

2. Random interval functions

Let $\boldsymbol{\Delta}$ denote the collection of finite intervals of the form $\Delta = (\alpha, \beta]$. We consider r vector valued stochastic processes $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$ with the additivity property

$$\mathbf{X}(\Delta) = \mathbf{X}(\Delta_1) + \mathbf{X}(\Delta_2), \tag{2.1}$$

for $\Delta, \Delta_1, \Delta_2 \in \boldsymbol{\Delta}$ with $\Delta = \Delta_1 \cup \Delta_2$, $\Delta_1 \cap \Delta_2 = \varnothing$. Such a process will be called an *r vector valued additive stochastic interval function.* As one example we mention

$$\mathbf{X}(\Delta) = \int_\Delta \mathbf{Y}(t)\, dt, \tag{2.2}$$

where $\mathbf{Y}(t)$, $-\infty < t < \infty$, is a continuous r vector valued time series. As a second example we consider $\mathbf{X}(\Delta) = \mathbf{N}(\Delta)$ where $\mathbf{N}(\Delta)$ is an r vector valued point process with $N_a(\Delta)$ giving the number of events of the ath component of the process that occur in the interval Δ. As a final example we mention

$$\mathbf{X}(\Delta) = \int_\Delta \mathbf{Y}(t)\, dN(t), \tag{2.3}$$

where $\mathbf{Y}(t)$, $-\infty < t < \infty$, is an r vector valued continuous time series and $N(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, is a point process. If $\tau_1, \cdots, \tau_n$ denote the times of events of the process $N(\Delta)$ in the interval Δ, then $\mathbf{X}(\Delta)$ equals

$$\mathbf{Y}(\tau_1) + \cdots + \mathbf{Y}(\tau_n), \tag{2.4}$$

in this case.

In connection with the process $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, we set down

ASSUMPTION 2.1. *The process* $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *is an r vector valued stochastic interval function possessing moments of all orders such that for* $\Delta_1, \cdots, \Delta_k \in \boldsymbol{\Delta}$; $a_1, \cdots, a_k = 1, \cdots, r$; $k = 1, 2, \cdots$,

$$E\{X_{a_1}(\Delta_1) \cdots X_{a_k}(\Delta_k)\} = \int_{\Delta_1} \cdots \int_{\Delta_k} dM_{a_1, \cdots, a_k}(t_1, \cdots, t_k) \tag{2.5}$$

for some function $M_{a_1, \cdots, a_k}(t_1, \cdots, t_k)$, $-\infty < t_j < \infty$, *of bounded variation in finite intervals.*

In the case that $\mathbf{X}(\Delta)$ satisfies this assumption and $\phi_a(t)$ is bounded and continuous for t in some interval of $\boldsymbol{\Delta}$ and 0 outside the interval, we may define stochastic integrals of the form

$$\int \phi_a(t)\, dX_a(t) \tag{2.6}$$

as the limit in mean (of any order $v > 0$) of approximating Riemann sums

$$\sum_{j=1}^{n} \phi_a(t_j)\, X_a(\Delta_j), \tag{2.7}$$

where $t_j \in \Delta_j$ and $\Delta_1 \cup \cdots \cup \Delta_n$ is a partition of the support of $\phi_a(t)$, $a = 1, \cdots, r$. (See Cramér and Leadbetter [12], p. 86 for the case $v = 2$.) These integrals have the property

$$(2.8)\qquad E\left\{\int \phi_{a_1}(t_1)\, dX_{a_1}(t_1) \cdots \int \phi_{a_k}(t_k)\, dX_{a_k}(t_k)\right\}$$
$$= \int \cdots \int \phi_c\, (t_1) \cdots \phi_{a_k}(t_k)\, dM_{a1,\cdots,a_k}(t_1, \cdots, t_k),$$

for $a_1, \cdots, a_k = 1, \cdots, r$.

For $\Delta = (\alpha, \beta]$ in $\boldsymbol{\Delta}$ we denote the translated interval $(\alpha + t, \beta + t]$ by $\Delta + t$ for $-\infty < t < \infty$. We will now say that an r vector valued additive stochastic interval function is *stationary* if the joint distributions of all finite collections of variates

$$(2.9)\qquad \mathbf{X}(\Delta_1 + t), \cdots, \mathbf{X}(\Delta_k + t),$$

$\Delta_1, \cdots, \Delta_k \in \boldsymbol{\Delta}$, $-\infty < t < \infty$, and $k = 1, 2, \cdots$, do not depend on t. In this connection we have

LEMMA 2.1. *If* $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *is a stationary r vector valued additive interval function satisfying Assumption* 2.1, *then for* $a_1, \cdots, a_k = 1, \cdots, r; k = 1, 2, \cdots,$

$$(2.10)\qquad E\{X_{a_1}(\Delta_1) \cdots X_{a_k}(\Delta_k)\}$$
$$= \int_{\Delta_1} \cdots \int_{\Delta_k} dM'_{a_1,\cdots,a_k}(t_1 - t_k, \cdots, t_{k-1} - t_k)\, dt_k,$$

for some function $M'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})$, $-\infty < u_j < \infty$, *of bounded variation in finite intervals.*

In the case $k = 1$, the lemma indicates that

$$(2.11)\qquad EX_a(\Delta) = C'_a|\Delta|,$$

for some constant C'_a, $a = 1, \cdots, r$ with $|\Delta|$ denoting the length of the interval Δ.

It follows from this lemma that one can write

$$(2.12)\qquad \operatorname{cum}\{X_{a_1}(\Delta_1), \cdots, X_{a_k}(\Delta_k)\}$$
$$= \int_{\Delta_1} \cdots \int_{\Delta_k} dC'_{a_1,\cdots,a_k}(t_1 - t_k, \cdots, t_{k-1} - t_k)\, dt_k,$$

for a function $C'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})$, of bounded variation in finite intervals. In differential notation we may write this last as

$$(2.13)\qquad \operatorname{cum}\{dX_{a_1}(u_1 + t), \cdots, dX_{a_{k-1}}(u_{k-1} + t), dX_{a_k}(t)\}$$
$$= dC'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})\, dt.$$

Taking note of Lemma 2.1 and (2.12) we set down the key assumption of our work. It is

ASSUMPTION 2.2. *The process* $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *is a stationary r vector valued additive interval function satisfying Assumption* 2.1 *and such that* $C'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})$ *of* (2.12) *satisfies*

$$(2.14)\qquad \int_{-\infty}^{\infty} \cdots \int |u_j|\, d|C'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})| < \infty,$$

for $j = 1, \cdots, k;\quad a_1, \cdots, a_k = 1, \cdots, r;\quad k = 2, 3, \cdots.$

This assumption has the nature of a mixing condition on the increments of $\mathbf{X}(t)$, that is, increments that are well separated in time are only weakly dependent.

In view of condition (2.13) we can define the Fourier transforms

$$f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = (2\pi)^{-k+1}\int_{-\infty}^{\infty}\cdots\int \exp\left\{-i\sum_1^{k-1}\lambda_j u_j\right\} dC'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}), \tag{2.15}$$

for $-\infty < \lambda_1,\cdots,\lambda_k < \infty$, where we understand $\lambda_1 + \cdots + \lambda_k = 0$ in the definition. For completeness we set

$$f_a(\lambda) = C'_a, \tag{2.16}$$

where C'_a was defined in (2.11), $a = 1,\cdots,r$. The transform $f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k)$ is called a *cumulant spectrum of order k* of the process $\mathbf{X}(\Delta)$, $\Delta \in \mathbf{\Delta}$. We will sometimes find it convenient to adopt the unsymmetric notation

$$f'_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_{k-1}) = f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k). \tag{2.17}$$

The second order cumulant spectra, $f'_{a,b}(\lambda)$, $-\infty < \lambda < \infty$, are of particular importance. It is convenient to collect them together into the $r \times r$ *spectral density matrix*

$$\mathbf{f}'_{X,X}(\lambda) = [f'_{a,b}(\lambda)]. \tag{2.18}$$

We also collect the first order spectra together into the r vector

$$\mathbf{f}'_X = [f'_a]. \tag{2.19}$$

There is an intimate connection between stationary interval functions and stationary series. Suppose that, $\mathbf{X}(\Delta)$, $\Delta \in \mathbf{\Delta}$, satisfies Assumption 2.2. and has cumulant spectra

$$f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k). \tag{2.20}$$

Suppose the real valued, $\phi_a(t)$, $-\infty < t < \infty$, satisfies

$$\int |t||\phi_a(t)|\,dt < \infty, \tag{2.21}$$

for $a = 1,\cdots,r$. Then the r vector valued times series, $\mathbf{Y}(t)$, $-\infty < t < \infty$, with components

$$Y_a(t) = \int \phi_a(t-u)\,dX_a(u), \tag{2.22}$$

$a = 1,\cdots,r$ may be seen to be stationary and such that

$$\operatorname{cum}\{Y_{a_1}(t+t_1),\cdots,Y_{a_{k-1}}(t+t_{k-1}),Y_{a_k}(t)\} = \int\cdots\int \phi_{a_1}(t_1-u_1)\cdots\phi_{a_{k-1}}(t_{k-1}-u_{k-1})\,\phi_{a_k}(u_k)\, dC'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})\,du_k. \tag{2.23}$$

Taking a Fourier transform, we see that the cumulant spectra of $\mathbf{Y}(t)$, in the sense of Brillinger and Rosenblatt [9], are given by

$$\Phi_{a_1}(\lambda_1) \cdots \Phi_{a_k}(\lambda_k) f_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_k), \tag{2.24}$$

where

$$\Phi_a(\lambda) = \int \exp\{-i\lambda t\} \phi_a(t)\, dt, \tag{2.25}$$

for $a = 1, \cdots, r$.

A variety of authors (including Kolmogorov [17], Doob [14] p. 551, Ito [16], Yaglom [24], [25] p. 86, Bochner [5] p. 159) have given spectral representations for stationary interval functions (or processes with stationary increments). In this connection we mention

THEOREM 2.1. *Let the process* $X(\Delta)$, $\Delta \in \Delta$, *satisfy Assumption* 2.2. *Let*

$$\mathbf{Z}_X^{(T)}(\lambda) = (2\pi)^{-1} \int_{-T}^{T} \left[\frac{1 - \exp\{-i\lambda t\}}{-it}\right] d\mathbf{X}(t), \tag{2.26}$$

for $-\infty < \lambda < \infty$. *Then there exists,* $\mathbf{Z}_X(\lambda)$, $-\infty < \lambda < \infty$, *such that* $\mathbf{Z}_X^{(T)}(\lambda)$ *tends to* $\mathbf{Z}_X(\lambda)$ *in mean order* v, *for any* $v > 0$. $\mathbf{Z}_X(\lambda)$ *satisfies*

$$\begin{aligned} &\operatorname{cum}\{Z_{a_1}(\lambda_1), \cdots, Z_{a_k}(\lambda_k)\} \\ &\quad = \int_0^{\lambda_1} \cdots \int_0^{\lambda_k} \delta\left(\sum_1^k \alpha_j\right) f_{a_1,\cdots,a_k}(\alpha_1, \cdots, \alpha_k)\, d\alpha_1 \cdots d\alpha_k, \end{aligned} \tag{2.27}$$

for $a_1, \cdots, a_k = 1, \cdots, r$; $k = 1, 2, \cdots$. *Also*

$$\mathbf{X}(\Delta) = \int_{-\infty}^{\infty} \left[\int_\Delta \exp\{i\lambda t\}\, dt\right] d\mathbf{Z}_X(\lambda), \tag{2.28}$$

with probability one.

In differential notation particular cases of (2.27) include:

$$E\, d\mathbf{Z}_X(\lambda) = \delta(\lambda)\mathbf{f}_X\, d\lambda; \tag{2.29}$$

$$\operatorname{Cov}\{d\mathbf{Z}_X(\lambda), d\mathbf{Z}_X(\mu)\} = \delta(\lambda - \mu)\mathbf{f}'_{X,X}(\lambda)\, d\lambda\, d\mu; \tag{2.30}$$

$$\begin{aligned} &\operatorname{cum}\{dZ_{a_1}(\lambda_1), \cdots, dZ_{a_k}(\lambda_k)\} \\ &\quad = \delta(\lambda_1 + \cdots + \lambda_k) f_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_{k-1})\, d\lambda_1 \cdots d\lambda_k. \end{aligned} \tag{2.31}$$

Also if we set $\mathbf{X}(t) = \mathbf{X}(0, t]$, then (2.28) takes the form

$$\mathbf{X}(t) = \int \left[\frac{\exp\{i\lambda t\} - 1}{i\lambda}\right] d\mathbf{Z}_X(\lambda), \tag{2.32}$$

for $-\infty < t < \infty$.

The representation (2.28) is useful for displaying the effect of linear time invariant operations on the process, $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$. Suppose $\mathbf{a}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, is an $s \times r$ matrix valued interval function of bounded variation satisfying

$$\int |t|\, d|\mathbf{a}(t)| < \infty. \tag{2.33}$$

Set

$$\mathbf{A}(\lambda) = \int \exp\{-i\lambda t\}\, d\mathbf{a}(t), \tag{2.34}$$

for $-\infty < \lambda < \infty$. The s vector valued interval function

$$\begin{aligned} \mathbf{Y}(\Delta) &= \mathbf{a} * \mathbf{X}(\Delta) \\ &= \int \mathbf{a}(\Delta - u)\, d\mathbf{X}(u), \end{aligned} \tag{2.35}$$

$\Delta \in \boldsymbol{\Delta}$, may be seen to satisfy Assumption 2.2. Also the process $\mathbf{Z}_Y(\lambda)$, $-\infty < \lambda < \infty$, of its spectral representation may be seen to satisfy,

$$d\mathbf{Z}_Y(\lambda) = \mathbf{A}(\lambda)\, d\mathbf{Z}_X(\lambda). \tag{2.36}$$

We may infer from this last that the spectral density matrices of $\mathbf{X}(\Delta)$ and $\mathbf{Y}(\Delta)$ are related by

$$\mathbf{f}'_{Y,Y}(\lambda) = \mathbf{A}(\lambda)\mathbf{f}'_{X,X}(\lambda)\overline{\mathbf{A}(\lambda)}^{\tau}. \tag{2.37}$$

This last relation has the identical form with that giving the effect of linear time invariant operations on the spectral density matrices of time series.

We conclude this section by remarking that the function $M'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})$ of (2.10) may be determined as

$$\begin{aligned} &M'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1}) \\ &\quad = E\left\{T^{-1} \int_0^T X_{a_1}(t, t + u_1] \cdots X_{a_{k-1}}(t, t + u_{k-1}]\, dX_{a_k}(t)\right\}. \end{aligned} \tag{2.38}$$

3. Some examples

EXAMPLE 3.1. Suppose that $\mathbf{Y}(t)$, $-\infty < t < \infty$, is an r vector valued stationary time series possessing moments of all orders. If

$$c'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1}) = \text{cum}\,\{Y_{a_1}(t + u_1), \cdots, Y_{a_{k-1}}(t + u_{k-1}), Y_{a_k}(t)\}, \tag{3.1}$$

satisfies

$$\int \cdots \int |u_j| |c'_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})|\, du_1 \cdots du_{k-1} < \infty. \tag{3.2}$$

the cumulant spectra of the series, $\mathbf{Y}(t)$, $-\infty < t < \infty$, are given by

$$(3.3)\qquad f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = (2\pi)^{-k+1}\int\cdots\int \exp\left\{-i\sum_1^{k-1}\lambda_j u_j\right\} c'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})\, du_1\cdots du_{k-1},$$

understanding $\lambda_1 + \cdots + \lambda_k = 0$ (see Brillinger and Rosenblatt [9]). Also the Cramér representation of $\mathbf{Y}(t)$ is given by

$$(3.4)\qquad \mathbf{Y}(t) = \int \exp\{i\lambda t\}\, d\mathbf{Z}_Y(\lambda),$$

where

$$(3.5)\qquad \mathbf{Z}_Y(\lambda) = \underset{T\to\infty}{\text{l.i.m.}}\ (2\pi)^{-1}\int_{-T}^{T}\left[\frac{\exp\{-i\lambda t\} - 1}{-it}\right]\mathbf{Y}(t)\, dt.$$

Suppose we construct the interval process

$$(3.6)\qquad \mathbf{X}(\Delta) = \int_\Delta \mathbf{Y}(t)\, dt,$$

then we quickly see that this process satisfies Assumption 2.2 with

$$(3.7)\qquad C'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}) = \int_0^{u_1}\cdots\int_0^{u_{k-1}} c'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})\, du_1\cdots du_{k-1}.$$

The cumulant spectra of the interval process, $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, are therefore the same as the cumulant spectra of the time series $\mathbf{Y}(t)$, $-\infty < t < \infty$.

A comparison of expressions (3.5) and (2.26) indicates that $\mathbf{Z}_X(\lambda)$ of the spectral representation of $\mathbf{X}(\Delta)$ is equivalent with $\mathbf{Z}_Y(\lambda)$ of the Cramér representation of $\mathbf{Y}(t)$.

In a later section we will see that our proposed empirical analysis of the process, $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, reduces to the usual empirical analysis of the continuous series $\mathbf{Y}(t)$, $-\infty < t < \infty$.

EXAMPLE 3.2. Consider an r vector valued point process, $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$. Here $N_a(\Delta)$ represents the number of events of the ath sort that occur in the interval Δ. If we let $\mathbf{1}_a$ denote a vector with 1 as its ath component and 0 elsewhere, then we may set down

ASSUMPTION 3.1. *The point process,* $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *possesses moments of all orders and is such that if* $\Delta_1, \cdots, \Delta_k$ *are disjoint intervals with* $|\Delta_1|, \cdots |\Delta_k| \leqq \delta < \infty$,

$$(3.8)\qquad P\{\mathbf{N}(\Delta_1) = \mathbf{n}_1, \cdots, \mathbf{N}(\Delta_k) = \mathbf{n}_k\} < K_\delta |\Delta_1|^{|\mathbf{n}_1|}\cdots|\Delta_k|^{|\mathbf{n}_k|}$$

for some finite K_δ *and for* $\mathbf{n}_1, \cdots, \mathbf{n}_k$ *having nonnegative integral coordinates.*

Also if $t_j \in \Delta_j$, for such $\Delta_1, \cdots, \Delta_k$, there is a function $p_{a_1,\cdots,a_k}(t_1, \cdots, t_k)$, bounded in finite intervals, such that

$$\lim_{|\Delta_j|\to 0} |\Delta_1|^{-1} \cdots |\Delta_k|^{-1} P\{\mathbf{N}(\Delta_1) = \mathbf{1}_{a_1}, \cdots, \mathbf{N}(\Delta_k) = \mathbf{1}_{a_k}\} = p_{a_1,\cdots,a_k}(t_1, \cdots, t_k), \tag{3.9}$$

uniformly in $t_1, \cdots, t_k$.

The functions $p_{a_1,\cdots,a_k}(t_1, \cdots . t_k)$ have been called *product density functions*, see Srinivasan [21]. The function $p_a(t)$, $-\infty < t < \infty$, is called the *density of events* of the ath sort at time t, $a = 1, \cdots, r$.

We note that the process satisfies

$$P\left\{\sum_{a=1}^{r} N_a(\Delta) > 1\right\} = O(|\Delta|^2), \tag{3.10}$$

and is therefore *orderly* (events tend not to happen simultaneously). Also we have

$$P\{N_a(\Delta) = 1\} = p_a(t)|\Delta| + o(|\Delta|), \tag{3.11}$$

and

$$P\{N_a(\Delta) = 0\} = 1 - p_a(t)|\Delta| + o(|\Delta|). \tag{3.12}$$

In the theorem below we let $\delta\{x\}$ denote the Kronecker delta $\delta\{x\} = 1$ if $x = 0$ and $\delta\{x\} = 0$ otherwise. We let $\chi_\Delta(\tau)$ denote the indicator function $\chi_\Delta(\tau) = 1$ if $\tau \in \Delta$, $\chi_\Delta(\tau) = 0$ otherwise. We have

THEOREM 3.1. *Let the r vector valued point process $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, satisfy Assumption 3.1. Then*

$$E\{N_{a_1}(\Delta_1) \cdots N_{a_k}(\Delta_k)\} = \sum_{\ell=1}^{k} \sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r} [\prod_{j\in\nu_1} \delta\{\alpha_1 - a_j\}] \cdots [\prod_{j\in\nu_\ell} \delta\{\alpha_\ell - a_j\}]$$
$$\int \cdots \int [\prod_{j\in\nu_1} \chi_{\Delta_j}(\tau_1)] \cdots [\prod_{j\in\nu_\ell} \chi_{\Delta_j}(\tau_\ell)]\, p_{\alpha_1,\cdots,\alpha_\ell}(\tau_1, \cdots, \tau_\ell)\, d\tau_1 \cdots d\tau_\ell \tag{3.13}$$

with the sum extending over all partitions $(\nu_1, \cdots, \nu_\ell)$ of the set $(1, \cdots, k)$.

We see that the moments of $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, have the integral representation required in Assumption 2.1. Particular cases of this theorem include

$$EN_a(\Delta) = \int_\Delta p_a(\tau)\, d\tau, \tag{3.14}$$

$$E\{N_a(\Delta_1)N_a(\Delta_2)\} = \int_{\Delta_1}\int_{\Delta_2} p_{a,a}(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 + \int_{\Delta_1\cap\Delta_2} p_a(\tau)\, d\tau, \tag{3.15}$$

$$E\{N_a(\Delta_1)N_b(\Delta_2)\} = \int_{\Delta_1}\int_{\Delta_2} p_{a,b}(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 \qquad \text{if } a \neq b, \tag{3.16}$$

and

$$E\{N_a(\Delta)^k\} = \sum_{\ell=1}^{k} \mathscr{S}_\ell^{(k)} \int_\Delta \cdots \int_\Delta p_{a,\cdots,a}(\tau_1, \cdots, \tau_\ell)\, d\tau_1 \cdots d\tau_\ell, \tag{3.17}$$

where $\mathscr{S}_\ell^{(k)}$ denotes a Stirling number of the second kind. Expression (3.17) was set down by Ramakrishnan [20] and Kuznetsov and Stratonovich [18].

Kuznetsov and Stratonovich [18] remarked that it might prove more useful to consider the cumulant functions

$$\begin{aligned} &q_{a_1,\cdots,a_k}(t_1, \cdots, t_k) \\ &\quad = \sum_{\ell=1}^{k} (-1)^{\ell-1}(\ell - 1)!\, p_{a_j;\, j\in v_1}(t_j; j \in v_1) \cdots p_{a_j;\, j\in v_\ell}(t_j; j \in v_\ell), \end{aligned} \tag{3.18}$$

where the summation extends over all partitions $(v_1, \cdots, v_\ell)$ of $(1, \cdots, k)$. These functions have the property of tending to 0 as $|t_i - t_j| \to \infty$ in the case that the increments of $\mathbf{N}(t)$ are tending to become independent as they separate in time. Particular cases of the functions include

$$q_a(t) = p_a(t), \tag{3.19}$$

$$q_{a,b}(t_1, t_2) = p_{a,b}(t_1, t_2) - p_a(t_1)p_b(t_2). \tag{3.20}$$

The inverse relation to (3.18) is

$$p_{a_1,\cdots,a_k}(t_1, \cdots, t_k) = \sum_{\ell=1}^{k} q_{a_j;\, j\in v_1}(t_j; j \in v_1) \cdots q_{a_j;\, j\in v_\ell}(t_j; j \in v_\ell). \tag{3.21}$$

We have

THEOREM 3.2. *Let the r vector valued point process* $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *satisfy Assumption* 3.1. *Then*

$$\begin{aligned} &\operatorname{cum}\{N_{a_1}(\Delta_1), \cdots, N_{a_k}(\Delta_k)\} = \sum_{\ell=1}^{k} \sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r} [\prod_{j\in v_1} \delta\{\alpha_1 - a_j\}] \cdots [\prod_{j\in v_\ell} \delta\{\alpha_\ell - a_j\}] \\ &\qquad \int \cdots \int [\prod_{j\in v_1} \chi_{\Delta_j}(\tau_1)] \cdots [\prod_{j\in v_\ell} \chi_{\Delta_j}(\tau_\ell)]\, q_{\alpha_1,\cdots,\alpha_\ell}(\tau_1, \cdots, \tau_\ell)\, d\tau_1 \cdots d\tau_\ell \end{aligned} \tag{3.22}$$

where the summation extends over all partitions $(v_1, \cdots, v_\ell)$ *of the set* $(1, \cdots, k)$.

The relation (3.22) has the same form as the relation (3.13). As particular cases we mention

$$EN_a(\Delta) = \int_\Delta q_a(\tau)\, d\tau = \int_\Delta p_a(\tau)\, d\tau, \tag{3.23}$$

$$\operatorname{Cov}\{N_a(\Delta_1), N_a(\Delta_2)\} = \int_{\Delta_1} \int_{\Delta_2} q_{a,a}(\tau_1, \tau_2)\, d\tau_1\, d\tau_2 + \int_{\Delta_1\cap\Delta_2} q_a(\tau)\, d\tau, \tag{3.24}$$

$$\operatorname{Cov}\{N_a(\Delta_1), N_b(\Delta_2)\} = \int_{\Delta_1} \int_{\Delta_2} q_{a,b}(\tau_1, \tau_2)\, d\tau_1\, d\tau_2, \tag{3.25}$$

and

$$\text{(3.26)} \qquad \text{cum}_k \{N_a(\Delta)\} = \sum_{\ell=1}^{k} \mathscr{S}_\ell^{(k)} \int_\Delta \cdots \int_\Delta q_{a,\cdots,a}(\tau_1, \cdots, \tau_\ell)\, d\tau_1 \cdots d\tau_\ell.$$

This last expression was given by Kuznetsov and Stratonovich [18]. We remark that in differential notation, (3.22) has the form

$$\text{(3.27)} \qquad \text{cum}\, \{dN_{a_1}(t_1), \cdots, dN_{a_k}(t_k)\} = q_{a_1,\cdots,a_k}(t_1, \cdots, t_k)\, dt_1 \cdots dt_k,$$

if the t_j are distinct. As a further implication of the theorem we have

COROLLARY 3.1. *Under the conditions of the theorem and if* $\phi_a(t)$ *is continuous with finite support,* $a = 1, \cdots, k$, *then*

$$\text{(3.28)} \qquad \text{cum}\left\{\int \phi_1(t)\, dN_{a_1}(t), \cdots, \int \phi_k(t)\, dN_{a_k}(t)\right\}$$

$$= \sum_{\ell=1}^{k} \sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r} [\prod_{j\in\nu_1} \delta\{\alpha_1 - a_j\}] \cdots [\prod_{j\in\nu_\ell} \delta\{\alpha_\ell - a_j\}] \int \cdots \int [\prod_{j\in\nu_1} \phi_j(\tau_1)] \cdots [\prod_{j\in\nu_\ell} \phi_j(\tau_\ell)]\, q_{\alpha_1,\cdots,\alpha_\ell}(\tau_1, \cdots, \tau_\ell)\, d\tau_1 \cdots d\tau_\ell$$

where the summation is over all partitions $(\nu_1, \cdots, \nu_\ell)$ *of* $(1, \cdots, k)$.

If the point process $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, is stationary, then

$$\text{(3.29)} \qquad p_{a_1,\cdots,a_k}(t + t_1, \cdots, t + t_k) = p_{a_1,\cdots,a_k}(t_1, \cdots, t_k),$$

and

$$\text{(3.30)} \qquad q_{a_1,\cdots,a_k}(t + t_1, \cdots, t + t_k) = q_{a_1,\cdots,a_k}(t_1, \cdots, t_k),$$

for all real t, $t_1, \cdots, t_k$. In this case we set

$$\text{(3.31)} \qquad r_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1}) = q_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1}, 0).$$

The parameter r_a is called the *mean intensity* of the process $N_a(\Delta)$, $\Delta \in \boldsymbol{\Delta}$: $r_{a,a}(u)$ is called the *covariance density* of the process $N_a(\Delta)$, $\Delta \in \boldsymbol{\Delta}$; and $r_{a,b}(u)$, for $a \neq b$, is called the *cross covariance density* of the component $N_a(\Delta)$ with the component $N_b(\Delta)$.

We now set down

ASSUMPTION 3.2. $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *is an* r *vector valued stationary point process satisfying Assumption* 3.1 *and such that*

$$\text{(3.32)} \qquad \int \cdots \int |u_j| |r_{a_1,\cdots,a_k}(u_1, \cdots, u_{k-1})|\, du_1 \cdots du_{k-1} < \infty$$

for $a_1, \cdots, a_k = 1, \cdots, r$; $k = 2, 3, \cdots$.

If the process $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, satisfies this assumption, then we may define the

Fourier transforms

$$(3.33)\quad g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = \int\cdots\int \exp\left\{-i\sum_1^{k-1}\lambda_j u_j\right\} r_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})\,du_1\cdots du_{k-1},$$

understanding $\lambda_1 + \cdots + \lambda_k = 0$. For completeness we set

$$(3.34)\quad g_a(\lambda) = r_a = q_a(t) = p_a(t),$$

in the case $k = 1$. We now have

THEOREM 3.3. *Let the point process* $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *satisfy Assumption* 3.2. *Then the process satisfies Assumption* 2.2. *Its cumulant spectra are given by*

$$(3.35)\quad f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = (2\pi)^{-k+1}\sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}\Big[\prod_{j\in\nu_1}\delta\{\alpha_1 - a_j\}\Big]\cdots\Big[\prod_{j\in\nu_\ell}\delta\{\alpha_\ell - a_j\}\Big]\, g_{\alpha_1,\cdots,\alpha_\ell}\Big[\sum_{j\in\nu_1}\lambda_j,\cdots,\sum_{j\in\nu_\ell}\lambda_j\Big],$$

with the summation extending over all partitions $(\nu_1,\cdots,\nu_\ell)$ *of* $(1,\cdots,k)$.

As particular cases of the cumulant spectra we mention

$$(3.36)\quad f'_a = r_a,$$

$$(3.37)\quad f'_{a,a}(\lambda) = (2\pi)^{-1}[g'_{a,a}(\lambda) + g_a] = (2\pi)^{-1}\left[\int \exp\{-i\lambda t\}\, r_{a,a}(t)\,dt + r_a\right],$$

in agreement with Bartlett [4], p. 183. Also

$$(3.38)\quad f'_{a,b}(\lambda) = (2\pi)^{-1}\int \exp\{-i\lambda t)\, r_{a,b}(t)\,dt \qquad \text{if} \quad a \neq b,$$

and

$$(3.39)\quad f'_{a,a,a}(\lambda_1,\lambda_2) = (2\pi)^{-2}\,[g'_{a,a,a}(\lambda_1,\lambda_2) + g'_{a,a}(\lambda_1) + g'_{a,a}(\lambda_2) + g'_{a,a}(-\lambda_1 - \lambda_2) + g'_a],$$

We have the following relation, inverse to (3.35),

$$(3.40)\quad g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = \sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}(-1)^{\ell-1}(\ell-1)!\,(2\pi)^{k-\ell}\Big[\prod_{j\in\nu_1}\delta\{\alpha_1 - a_j\}\Big]\cdots\Big[\prod_{j\in\nu_\ell}\delta\{\alpha_\ell - a_j\}\Big]\cdot f_{\alpha_1,\cdots,\alpha_\ell}\Big(\sum_{j\in\nu_1}\lambda_j,\cdots,\sum_{j\in\nu_\ell}\lambda_j\Big),$$

where the summation is again over all partitions $(\nu_1,\cdots,\nu_\ell)$ of the integers $(1,\cdots,k)$.

In Section 2 we discussed a class linear time invariant operations on stationary interval processes. It may be of interest to indicate a subclass of these operations which carry point processes over into point processes. Let $\sigma_j, j = 0, \pm 1, \cdots$, be a sequence of real numbers. Let

$$a(\Delta) = \text{the number of } \sigma_j \in \Delta, \tag{3.41}$$

then.

$$\begin{aligned} Y(\Delta) &= a * N(\Delta) \\ &= \int a(\Delta - u)\, dN(u) \end{aligned} \tag{3.42}$$

will be a real valued point process in the case that $N(\Delta)$, $\Delta \in \mathbf{\Delta}$, is one. If τ_j, $j = 0, \pm 1, \cdots$, denote the times of events of a realization of $N(\Delta)$, then events of this $Y(\Delta)$ occur at the times $\tau_j + \sigma_k, j, k = 0, \pm 1, \cdots$.

Daley [13] discusses the second order spectral theory of point processes, considers operations on point processes, and presents a variety of examples.

Example 3.3. Suppose that $\mathbf{Y}(t)$, $-\infty < t < \infty$, is an r vector valued stationary time series satisfying the conditions of Example 3.1 and having Cramér representation

$$\mathbf{Y}(t) = \int \exp\{i\lambda t\}\, d\mathbf{Z}_Y(\lambda). \tag{3.43}$$

Suppose $N(\Delta)$, $\Delta \in \mathbf{\Delta}$, is an independent stationary point process satisfying Assumption 3.2 and having spectral representation

$$N(\Delta) = \int \left[\int_\Delta \exp\{i\lambda t\}\, dt\right] dZ_N(\lambda). \tag{3.44}$$

In Section 6 of the paper we will consider the process

$$\begin{aligned} \mathbf{X}(\Delta) &= \int_\Delta \mathbf{Y}(t)\, dN(t) \\ &= \mathbf{Y}(\tau_1) + \cdots + \mathbf{Y}(\tau_n), \end{aligned} \tag{3.45}$$

if $\tau_1, \cdots, \tau_n$ are the events of $N(\Delta)$ in the interval Δ. One can check that this process satisfies Assumption 2.2. If its spectral representation is

$$\mathbf{X}(\Delta) = \int \left[\int_\Delta \exp\{i\lambda t\}\, dt\right] d\mathbf{Z}_X(\lambda), \tag{3.46}$$

then we see directly that

$$d\mathbf{Z}_X(\lambda) = \int [d\mathbf{Z}_Y(\lambda - \alpha)]\, dZ_N(\alpha), \tag{3.47}$$

for $-\infty < \lambda < \infty$. Expression (3.47) may be used to determine the cumulant spectra of $\mathbf{X}(\Delta)$ in terms of those of $\mathbf{Y}(t)$ and $N(\Delta)$.

We mention that Walker suggested the consideration of real valued processes of the form (3.45) in the discussion of Bartlett [3].

4. Stochastic properties of finite Fourier transforms

We now turn to an investigation of certain statistics useful in the estimation of the cumulant spectra of a stationary interval function $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$. We will suppose that the values of $\mathbf{X}(\Delta)$ are available for Δ contained in the support of a function $h(t/T)$, $T = 1, 2, \cdots$. We set down,

ASSUMPTION 4.1. *The function* $h(t)$, $-\infty < t < \infty$, *is measurable in* t, *bounded, zero for* $|t| > 1$ *and there exists a finite* K *such that*

$$\int |h(t + u) - h(t)|\, dt < K|u| \tag{4.1}$$

for all real u.

The inequality (4.1) will be satisfied if $h(t)$ is of bounded variation, for example. For given T, the function $h(t/T)$ has been called a *taper* by Tukey [22]. It has also been called a *data window*.

The principal statistics of our analysis of interval processes are the finite Fourier transforms,

$$d_a^{(T)}(\lambda) = \int h_a(t/T) \exp\{-i\lambda t\}\, dX_a(t), \tag{4.2}$$

$a = 1, \cdots, r$, $-\infty < \lambda < \infty$. In the case of Example 3.1, the statistic (4.2) takes the form

$$d_a^{(T)}(\lambda) = \int h_a(t/T) \exp\{-i\lambda t\}\, Y_a(t)\, dt, \tag{4.3}$$

that is, it is the Fourier transform of the tapered values that was considered in Brillinger and Rosenblatt [9]. In the case of Example 3.2, if we let $\tau_a(1), \cdots, \tau_a(n_a)$ denote the times of events of the ath sort that occur in the support of $h_a(t/T)$, then the statistic (4.2) has the form

$$\sum_{j=1}^{n_a} h_a(\tau_a(j)/T) \exp\{-i\lambda\tau_a(j)\}. \tag{4.4}$$

This statistic, excluding the taper, was considered in Bartlett [3] for the case $r = 1$ and suggested for the case of general r by Jenkins in the discussion of that paper. In the case of Example 3.3, the statistic has the form

$$\sum_{j=1}^{n} h_a(\tau_j/T) \exp\{-i\lambda\tau_j\}\, Y_a(\tau_j), \tag{4.5}$$

if $\tau_1, \cdots, \tau_n$ denote the times of events of the process $N(\Delta)$ in the support of $h_a(t/T)$.

We next present a basic theorem indicating the asymptotic joint cumulants of the Fourier transform (4.2). In the theorem we let

$$H_{a_1,\cdots,a_k}(\lambda) = \int h_{a_1}(t) \cdots h_{a_k}(t) \exp\{-i\lambda t\}\, dt. \tag{4.6}$$

THEOREM 4.1. *Let the process* $X(\Delta)$, $\Delta \in \Delta$, *satisfy Assumption* 2.2. *Let* $h_a(t)$, $a = 1, \cdots, r$, $-\infty < t < \infty$, *satisfy Assumption* 4.1. *Then as* $T \to \infty$

$$(4.7) \qquad \operatorname{cum}\{d_{a_1}^{(T)}(\lambda_1), \cdots, d_{a_k}^{(T)}(\lambda_k)\} = TH_{a_1,\cdots,a_k}\left(T\sum_1^k \lambda_j\right)(2\pi)^{k-1}f_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_{k-1}) + O(1)$$

for $a_1, \cdots, a_k = 1, \cdots, r$; $k = 1, 2, \cdots$. *The* $O(1)$ *term is uniformly bounded in* $\lambda_1, \cdots, \lambda_k$.

We see that the joint cumulants are of reduced order unless $\Sigma_1^k \lambda_j$ is near zero. We see from (4.6) and (4.7) that the joint cumulants based on disjoint stretches of data are of reduced order as well.

If $h_a(t) = 1$ for $0 \leqq t \leqq 1$ and $h_a(t) = 0$ otherwise, then this theorem has identical nature with the key theorem used in Brillinger and Rosenblatt [9], Brillinger [6], Brillinger [7] to develop properties of spectral estimates. The results of these papers therefore become directly available. We indicate a selection of results that now hold.

We begin by considering the asymptotic distribution of the finite Fourier transform. Let $N_r^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the complex r variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We have

THEOREM 4.2. *Let* $\mathbf{X}(\Delta)$, $\Delta \in \Delta$, *be an r vector valued interval process satisfying Assumption* 2.2. *Let* $s_j(T)$ *be an integer with* $\lambda_j(T) = 2\pi s_j(T)/T \to \lambda_j$ *as* $T \to \infty$ *for* $j = 1, \cdots, J$. *Suppose* $\lambda_j(T) \pm \lambda_k(T) \neq 0$ *for* $j, k = 1, \cdots, J$. *Let*

$$(4.8) \qquad \mathbf{d}_X^{(T)}(\lambda) = \int_0^T \exp\{-i\lambda t\}\, d\mathbf{X}(t)$$

for $-\infty < \lambda < \infty$. *Then* $\mathbf{d}_X^{(T)}(\lambda_j(T))$, $j = 1, \cdots, J$ *are asymptotically independent* $N_r^C(\mathbf{0}, 2\pi T\mathbf{f}'_{X,X}(\lambda))$ *variates, respectively. Also* $\mathbf{d}_X^{(T)}(0) = \mathbf{X}(0, T]$ *is asymptotically* $N_r(T\mathbf{f}'_X, 2\pi T\mathbf{f}'_{X,X}(0))$ *independently of the previous variates.*

This theorem has the nature of a central limit theorem. Let $W_r^C(n, \boldsymbol{\Sigma})$ denote the complex Wishart distribution of dimensions $r \times r$, degrees of freedom n and covariance matrix $\boldsymbol{\Sigma}$. Define the matrix of periodograms

$$(4.9) \qquad \mathbf{I}_{X,X}^{(T)}(\lambda) = (2\pi T)^{-1}\, \mathbf{d}_X^{(T)}(\lambda)\, \overline{\mathbf{d}_X^{(T)}(\lambda)}^{\tau}.$$

We have the following corollary.

COROLLARY 4.1. *Under the conditions of Theorem* 4.2, *if* $\lambda_1 = \cdots = \lambda_J = \lambda$ *and if*

$$(4.10) \qquad \mathbf{f}_{X,X}^{(T)}(\lambda) = J^{-1}\sum_{j=1}^J \mathbf{I}_{X,X}^{(T)}(\lambda_j(T)),$$

$\mathbf{f}_{X,X}^{(T)}(\lambda)$ *is asymptotically* $J^{-1}W_r^C(J, \mathbf{f}'_{X,X}(\lambda))$ *as* $T \to \infty$.

This corollary makes precise the chi square approximation for the distribution of second order spectral densities of point processes suggested by Bartlett [3].

We next construct consistent asymptotically normal estimates of the cumulant spectra of different orders of an interval process $\mathbf{X}(\triangle)$, $\triangle \in \Delta$. We begin by letting $W(u_1, \cdots, u_k)$ be a weight function satisfying

ASSUMPTION 4.2. *The function* $W(u_1, \cdots, u_k)$, $-\infty < u_j < \infty$, *is symmetric in* $u_1, \cdots, u_k$, *is concentrated on the plane* $\Sigma_1^k u_j = 0$, *and is such that*

$$\int_{-\infty}^{\infty} \cdots \int W(u_1, \cdots, u_k)\, \delta\left(\sum_1^k u_j\right) du_1 \cdots du_k = 1 \tag{4.11}$$

and

$$\left| W\left(u_1, \cdots, u_{k-1}, -\sum_1^{k-1} u_j\right)\right|, \left|\frac{\partial}{\partial u_\ell} W\left(u_1, \cdots, u_{k-1}, -\sum_1^{k-1} u_j\right)\right| \leqq A\left(1 + \left[\sum_1^{k-1} u_j^2\right]^{1/2}\right)^{-k-\varepsilon+1} \tag{4.12}$$

for some $A, \varepsilon > 0$, $\ell = 1, \cdots, k$.

Given the sequence of nonnegative numbers $B_T^{(k)}$, $T = 2, 3, \cdots$, we set

$$W_T(u_1, \cdots, u_k) = \left(B_T^{(k)}\right)^{-k+1} W\left(B_T^{(k)^{-1}} u_1, \cdots, B_T^{(k)^{-1}} u_k\right). \tag{4.13}$$

We suppose $B_T^{(2)} \leqq B_T^{(3)} \leqq \cdots$. Next we set $\Psi(u_1, \cdots, u_k) = 1$ if $\Sigma_1^k u_j = 0$ but no proper subset of the u_j has sum 0, and set it $= 0$ otherwise. Let

$$\mathbf{d}_X^{(T)}(\lambda) = \int_0^T \exp\{-i\lambda t\}\, d\mathbf{X}(t). \tag{4.14}$$

Finally set

$$I_{a_1,\cdots,a_k}^{(T)}(\lambda_1, \cdots, \lambda_k) = (2\pi)^{-k+1} T^{-1} \prod_{j=1}^k d_{a_j}^{(T)}(\lambda_j). \tag{4.15}$$

As an estimate of $f_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_k)$ we now take

$$f_{a_1,\cdots,a_k}^{(T)}(\lambda_1, \cdots, \lambda_k) = \left(\frac{2\pi}{T}\right)^{k-1} \sum_{s_1} \cdots \sum_{s_k} W_T\left(\lambda_1 - \frac{2\pi s_1}{T}, \cdots, \lambda_k - \frac{2\pi s_k}{T}\right) \cdot \Psi(s_1, \cdots, s_k) I_{a_1,\cdots,a_k}^{(T)}\left(\frac{2\pi s_1}{T}, \cdots, \frac{2\pi s_k}{T}\right). \tag{4.16}$$

In connection with this estimate we have the theorem,

THEOREM 4.3. *Let* $\mathbf{X}(\triangle)$, $\triangle \in \Delta$, *satisfy Assumption* 2.2. *Let* $W(u_1, \cdots, u_k)$ *satisfy Assumption* 4.2. *Let* $f_{a_1,\cdots,a_k}^{(T)}(\lambda_1, \cdots, \lambda_k)$ *be given by* (4.16). *Let* $B_T^{(k)} \to 0$, $(B_T^{(k)})^{k-1} T \to \infty$ *as* $T \to \infty$, *then*

(4.17) $\qquad Ef_{a_1,\cdots,a_k}^{(T)}(\lambda_1, \cdots, \lambda_k)$

$$= \int \cdots \int W_T(\lambda_1 - \alpha_1, \cdots, \lambda_k - \alpha_k) f_{a_1, \cdots, a_k}(\alpha_1, \cdots, \alpha_k)$$

$$\cdot \delta(\alpha_1 + \cdots + \alpha_k)\, d\alpha_1 \cdots d\alpha_k + O(B_T^{(k)^{-1}} T)$$

$$= f_{a_1, \cdots, a_k}(\lambda_1, \cdots, \lambda_k) + O(B_T^{(k)}) + O(B_T^{(k)^{-1}} T),$$

(4.18) $$\lim_{T \to \infty} (B_T^{(k)})^{k-1} T \operatorname{Cov} \{f^{(T)}_{a_1, \cdots, a_k}(\lambda_1, \cdots, \lambda_k), f^{(T)}_{a_1', \cdots, a_k'}(\mu_1, \cdots, \mu_k)\}$$

$$= 2\pi \sum_P \delta\{\lambda_1 - \mu_{P,1}\} \cdots \delta\{\lambda_k - \mu_{P,k}\} f'_{a_1 a_{P,1}'}(\lambda_1) \cdots f'_{a_k a_{P,k}'}(\lambda_k)$$

$$\int \cdots \int W(u_1, \cdots, u_k)^2 \delta\left(\sum_1^k u_j\right) du_1 \cdots du_k,$$

where the summation is over all permutations P of the integers $1, \cdots, k$. Collections of spectral estimates are asymptotically jointly normally distributed as $T \to \infty$ with estimates of different orders asymptotically independent and estimates of the same order having covariance structure given by (4.18).

We next turn to the development of an empirical analysis of the linear time invariant model,

(4.19) $$\mathbf{Y}(\Delta) = \mathbf{a} * \mathbf{X}(\Delta) + \boldsymbol{\varepsilon}(\Delta) = \int \mathbf{a}(\Delta - u)\, d\mathbf{X}(u) + \boldsymbol{\varepsilon}(\Delta),$$

with $\mathbf{X}(\Delta)$, $\boldsymbol{\varepsilon}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, independent stationary interval processes and

(4.20) $$\int |u|\, d|\mathbf{a}(u)| < \infty.$$

In differential notation we may write (4.19) as

(4.21) $$d\mathbf{Y}(t) = \int [d\mathbf{a}(t - u)]\, d\mathbf{X}(u) + d\boldsymbol{\varepsilon}(t).$$

Denote the cross spectral density matrix of the process

(4.22) $$\begin{bmatrix} \mathbf{X}(\Delta) \\ \mathbf{Y}(\Delta) \end{bmatrix},$$

$\Delta \in \boldsymbol{\Delta}$, by

(4.23) $$\begin{bmatrix} \mathbf{f}'_{X,X}(\lambda) & \mathbf{f}'_{X,Y}(\lambda) \\ \mathbf{f}'_{Y,X}(\lambda) & \mathbf{f}'_{Y,Y}(\lambda) \end{bmatrix},$$

and that of $\boldsymbol{\varepsilon}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, by $\mathbf{f}'_{\varepsilon,\varepsilon}(\lambda)$. Set

(4.24) $$\mathbf{A}(\lambda) = \int \exp\{-i\lambda t\}\, d\mathbf{a}(t).$$

Then (4.19) gives

$$\mathbf{f}'_{Y,X}(\lambda) = \mathbf{A}(\lambda)\mathbf{f}'_{X,X}(\lambda), \tag{4.25}$$

$$\mathbf{f}'_{Y,Y}(\lambda) = \mathbf{A}(\lambda)\mathbf{f}'_{X,X}(\lambda)\overline{\mathbf{A}(\lambda)}^{\tau} + \mathbf{f}'_{\varepsilon,\varepsilon}(\lambda). \tag{4.26}$$

These last suggest that we may base estimates of $\mathbf{A}(\lambda)$ and $\mathbf{f}'_{\varepsilon,\varepsilon}(\lambda)$ on an estimate of the spectral density matrix (4.23). We could construct an estimate of this last in the manner of (4.16); however, in order to display an alternate form of spectral estimate of order two we proceed slightly differently.

In constructing this alternate estimate we let $h(t)$, $-\infty < t < \infty$, be a tapering function satisfying Assumption 4.1. We then set

$$H_k(\lambda) = \int h(t)^k \exp\{-i\lambda t\}\, dt, \tag{4.27}$$

for $-\infty < \lambda < \infty$. We next set

$$\begin{aligned} \mathbf{d}_X^{(T)}(\lambda) &= \int h(t/T) \exp\{-i\lambda t\}\, d\mathbf{X}(t), \\ \mathbf{d}_Y^{(T)}(\lambda) &= \int h(t/T) \exp\{-i\lambda t\}\, d\mathbf{Y}(t); \end{aligned} \tag{4.28}$$

and we let $W(\alpha)$ be a weight function satisfying

ASSUMPTION 4.3. *$W(\alpha)$, $-\infty < \alpha < \infty$, is real valued, even, absolutely integrable, has an absolutely integrable first derivative, and*

$$\int_{-\infty}^{\infty} W(\alpha)\, d\alpha = 1. \tag{4.29}$$

The variate (4.22) has mean

$$\begin{bmatrix} \mathbf{f}'_X \\ \mathbf{f}'_Y \end{bmatrix} |\Delta|. \tag{4.30}$$

Estimates of $\mathbf{f}'_X$, $\mathbf{f}'_Y$ based on tapered values are provided by

$$\begin{aligned} \mathbf{f}_X^{(T)} &= \int h(t/T)\, d\mathbf{X}(t) \Big/ \int h(t/T)\, dt = \mathbf{d}_X^{(T)}(0)/[TH_1(0)], \\ \mathbf{f}_Y^{(T)} &= \int h(t/T)\, d\mathbf{Y}(t) \Big/ \int h(t/T)\, dt = \mathbf{d}_Y^{(T)}(0)/[TH_1(0)], \end{aligned} \tag{4.31}$$

respectively. The Fourier transform of the process (4.22) corrected for its sample mean is then given by

$$\begin{aligned} \mathbf{e}_X^{(T)}(\lambda) &= \int \exp\{-i\lambda t\} h(t/T)[d\mathbf{X}(t) - \mathbf{f}_X^{(T)}\, dt] \\ &= \mathbf{d}_X^{(T)}(\lambda) - \mathbf{d}_X^{(T)}(0) H_1(T\lambda)/H_1(0), \end{aligned} \tag{4.32}$$

$$\mathbf{e}_Y^{(T)}(\lambda) = \int \exp\{-i\lambda t\} h(t/T)[d\mathbf{Y}(t) - \mathbf{f}_Y^{(T)}\, dt]$$
$$= \mathbf{d}_Y^{(T)}(\lambda) - \mathbf{d}_Y^{(T)}(0) H_1(T\lambda)/H_1(0).$$

Let B_T be a sequence of nonnegative numbers tending to 0 as $T \to \infty$. Set

(4.33) $$W^{(T)}(\alpha) = B_T^{-1} W(B_T^{-1}\alpha).$$

As an estimate of the cross spectral density matrix (4.23) we now propose

(4.34) $$\begin{bmatrix} \mathbf{f}_{X,X}^{(T)}(\lambda) & \mathbf{f}_{X,Y}^{(T)}(\lambda) \\ \mathbf{f}_{Y,X}^{(T)}(\lambda) & \mathbf{f}_{Y,Y}^{(T)}(\lambda) \end{bmatrix} = \int W^{(T)}(\lambda - \alpha)(2\pi T)^{-1} \begin{bmatrix} \mathbf{e}_X^{(T)}(\alpha) \\ \mathbf{e}_Y^{(T)}(\alpha) \end{bmatrix} \overline{\begin{bmatrix} \mathbf{e}_X^{(T)}(\alpha) \\ \mathbf{e}_Y^{(T)}(\alpha) \end{bmatrix}}^{\tau} d\alpha.$$

As estimates of $\mathbf{A}(\lambda)$, $\mathbf{f}'_{\varepsilon,\varepsilon}(\lambda)$, we then take

(4.35) $$\mathbf{A}^{(T)}(\lambda) = \mathbf{f}_{Y,X}^{(T)}(\lambda)\mathbf{f}_{X,X}^{(T)}(\lambda)^{-1},$$

(4.36) $$\mathbf{g}_{\varepsilon,\varepsilon}^{(T)}(\lambda) = \mathbf{f}_{Y,Y}^{(T)}(\lambda) - \mathbf{f}_{Y,X}^{(T)}(\lambda)\mathbf{f}_{X,X}^{(T)}(\lambda)^{-1}\mathbf{f}_{X,Y}^{(T)}(\lambda).$$

We can now state the following theorem.

THEOREM 4.4. *Let the process* $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *satisfy Assumption* 2.2. *Suppose* $\mathbf{f}'_{X,X}(\lambda)$ *is nonsingular. Let the process* $\boldsymbol{\varepsilon}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, *satisfy Assumption* 2.2., *have mean* $\mathbf{0}$ *and be statistically independent of the process* $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$. *Let* $\mathbf{a}(\Delta)$ *satisfy* (4.20). *Let* $\mathbf{Y}(\Delta)$ *be given by* (4.19). *Let* $W(\alpha)$ *satisfy Assumption* 4.3 *and* $h(t)$ *satisfy Assumption* 4.1. *Then if* $B_T \to 0$, $B_T T \to \infty$ *as* $T \to \infty$,

(4.37) $$\lim_{T\to\infty} \overrightarrow{\text{ave}}\, \mathbf{A}^{(T)}(\lambda) = \mathbf{A}(\lambda),$$

(4.38) $$\lim_{T\to\infty} B_T T\, \overrightarrow{\text{Cov}}\, \{\text{vec}\, \mathbf{A}^{(T)}(\lambda), \text{vec}\, \mathbf{A}^{(T)}(\mu)\}$$
$$= 2\pi H_4(0) H_2(0)^{-2} \delta\{\lambda - \mu\} \mathbf{f}'_{\varepsilon,\varepsilon}(\lambda) \otimes \mathbf{f}'_{X,X}(\lambda)^{-1} \int W(\alpha)^2\, d\alpha,$$

(4.39) $$\lim_{T\to\infty} \overrightarrow{\text{ave}}\, \mathbf{g}_{\varepsilon,\varepsilon}^{(T)}(\lambda) = \mathbf{f}'_{\varepsilon,\varepsilon}(\lambda),$$

(4.40) $$\lim_{T\to\infty} B_T T\, \overrightarrow{\text{Cov}}\, \{g_{j,k}^{(T)}(\lambda), g_{m,n}^{(T)}(\mu)\}$$
$$= 2\pi H_4(0) H_2(0)^{-2} (\delta\{\lambda - \mu\} [\mathbf{f}'_{\varepsilon,\varepsilon}(\lambda)]_{j,m} [\mathbf{f}'_{\varepsilon,\varepsilon}(-\lambda)]_{k,n}$$
$$+ \delta\{\lambda + \mu\} [\mathbf{f}'_{\varepsilon,\varepsilon}(\lambda)]_{j,n} [f'_{\varepsilon,\varepsilon}(-\lambda)]_{k,m}) \int W(\alpha)^2\, d\alpha,$$

(4.41) $$\lim_{T\to\infty} B_T T\, \overrightarrow{\text{Cov}}\, \{\text{vec}\, \mathbf{A}^{(T)}(\lambda), g_{j,k}^{(T)}(\mu)\} = 0,$$

for $j, k, m, n = 1, \cdots, s$. *Also the variates* $\mathbf{A}^{(T)}(\lambda)$, $g_{\varepsilon,\varepsilon}^{(T)}(\mu)$ *are asymptotically jointly normal with the above covariance structure.*

(In this theorem $\overrightarrow{\text{ave}}$, $\overrightarrow{\text{Cov}}$ have technical definitions allowing the use of Taylor series expansions in determining asymptotic moments. See Brillinger and Tukey [10].)

In the case $r = s = 1$, we may define the *coherency* of $X(\Delta)$ with $Y(\Delta)$ at frequency λ by

$$|R_{Y,X}(\lambda)|^2 = \frac{|f'_{Y,X}(\lambda)|^2}{[f'_{X,X}(\lambda)f'_{Y,Y}(\lambda)]}. \tag{4.42}$$

As an estimate of the coherency we consider the statistic

$$|R^{(T)}_{Y,X}(\lambda)|^2 = \frac{|f^{(T)}_{Y,X}(\lambda)|^2}{[f^{(T)}_{X,X}(\lambda)f^{(T)}_{Y,Y}(\lambda)]}. \tag{4.43}$$

We then have from the theorem

COROLLARY 4.2. *Under the conditions of the theorem and if* $f'_{X,X}(\lambda).f'_{Y,Y}(\lambda) \neq 0$, $|R^{(T)}_{Y,X}(\lambda)|^2$ *is asymptotically normal with*

$$\lim_{T\to\infty} \overrightarrow{\text{ave}}\, |R^{(T)}_{Y,X}(\lambda)|^2 = |R_{Y,X}(\lambda)|^2 \tag{4.44}$$

and

$$\begin{aligned}\lim_{T\to\infty} B_T T\, \overrightarrow{\text{Cov}}\, \{|R^{(T)}_{Y,X}(\lambda)|^2, |R^{(T)}_{Y,X}(\mu)|^2\} \\ = 4\pi H_4(0)H_2(0)^{-2}[\delta\{\lambda - \mu\} \\ + \delta\{\lambda + \mu\}]|R_{Y,X}(\lambda)|^2[1 - |R_{Y,X}(\lambda)|^2]^2 \int W(\alpha)^2\, d\alpha.\end{aligned} \tag{4.45}$$

A comparison of the results of this theorem and its corollary, with the corresponding results for the regression of one vector valued stationary time series on another, shows that they are identical. This will also be the case for the interval process extension of many of the asymptotic results of the analysis of stationary time series.

5. Estimation of product densities

Let $\mathbf{N}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, be a stationary point process satisfying Assumption 3.2. We have defined various characteristics of such a process. These may be summarized as follows:

$$\begin{aligned}p_{a_1,\cdots,a_k}(t_1, \cdots, t_k) \\ = \lim_{dt_j\to 0} p\{dN_{a_1}(t_1) = 1, \cdots, dN_{a_k}(t_k) = 1\}/(dt_1 \cdots dt_k)\end{aligned} \tag{5.1}$$

for $t_1, \cdots, t_k$ distinct;

$$\begin{aligned}q_{a_1,\cdots,a_k}(t_1, \cdots, t_k) \\ = \sum_{\ell=1}^{k} (-1)^{\ell-1}(\ell - 1)!p_{a_j;j\in\nu_1}(t_j; j \in \nu_1) \cdots p_{a_j;j\in\nu_\ell}(t_j; j \in \nu_\ell);\end{aligned} \tag{5.2}$$

(5.3) $$p_{a_1,\cdots,a_k}(t_1,\cdots,t_k) = \sum_{\ell=1}^{k} q_{a_j;\,j\in\nu_1}(t_j;j\in\nu_1)\cdots q_{a_j;\,j\in\nu_\ell}(t_j;j\in\nu_\ell);$$

(5.4) $$r_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}) = q_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},0);$$

(5.5) $$g'_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_{k-1}) = \int\cdots\int \exp\left\{-i\sum_1^{k-1}\lambda_j u_j\right\} r_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})\,du_1\cdots du_{k-1};$$

and if $g'_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_{k-1})$ is integrable

(5.6)
$$r_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}) = (2\pi)^{-k+1}\int\cdots\int \exp\left\{i\sum_1^{k-1}\lambda_j u_j\right\} g'_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_{k-1})\,d\lambda_1\cdots d\lambda_{k-1}.$$

Also

(5.7) $$g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = g'_{a_1,\cdots,a_k}(\lambda_1\cdots,\lambda_{k-1})$$

understanding $\Sigma_1^k \lambda_j = 0$. Continuing

(5.8) $$f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = (2\pi)^{-k+1}\sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}\left[\prod_{j\in\nu_1}\delta\{\alpha_1-a_j\}\right]\cdots\left[\prod_{j\in\nu_\ell}\delta\{\alpha_\ell-a_j\}\right] \cdot g_{\alpha_1,\cdots,\alpha_\ell}\left[\sum_{j\in\nu_1}\lambda_j,\cdots,\sum_{j\in\nu_1}\lambda_j\right];$$

(5.9) $$g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = \sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}(-1)^{\ell-1}(\ell-1)!\left[\prod_{j\in\nu_1}\delta\{\alpha_1-a_j\}\right]\cdots\left[\prod_{j\in\nu_\ell}\delta\{\alpha_\ell-a_j\}\right] \cdot(2\pi)^{k-\ell} f_{\alpha_1,\cdots,\alpha_\ell}\left[\sum_{j\in\nu_1}\lambda_j,\cdots,\sum_{j\in\nu_\ell}\lambda_j\right].$$

The summations in (5.2), (5.3), (5.8), (5.9) are over all partitions $(\nu_1,\cdots,\nu_\ell)$ of the integers $(1,\cdots,k)$.

In the previous section we developed an estimate of $f_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k)$. Let us now put this work to use in developing estimates of the various functions listed above. As an estimate of $g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k)$, in the light of (5.9), we may consider

(5.10) $$g^{(T)}_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = \sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}(-1)^{\ell-1}(\ell-1)!\left[\prod_{j\in\nu_1}\delta\{\alpha_1-a_j\}\right]\cdots\left[\prod_{j\in\nu_\ell}\delta\{\alpha_\ell-a_j\}\right] \cdot(2\pi)^{k-\ell} f^{(T)}_{\alpha_1,\cdots,\alpha_\ell}\left[\sum_{j\in\nu_1}\lambda_j,\cdots,\sum_{j\in\nu_\ell}\lambda_j\right],$$

where the $f^{(T)}_{a_1,\cdots,a_\ell}(\mu_1,\cdots,\mu_\ell)$, $\ell = 1,\cdots,k$, are formed in the manner of Theorem 4.3. From that theorem we see that

$$(5.11)\qquad Eg^{(T)}_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) = g_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k) + O(B_T^{(k)}) + O(B_T^{(k)^{-1}}T^{-1});$$

and because estimates of order less than k have asymptotic variance of smaller order than that of estimates of order k, the covariance of $g^{(T)}_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k)$ with $g^{(T)}_{b_1,\cdots,b_\ell}(\mu_1,\cdots,\mu_\ell)$ will be asymptotically equivalent to that of $f^{(T)}_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_k)$ with $f^{(T)}_{b_1,\cdots,b_\ell}(\mu_1,\cdots,\mu_\ell)$ as given in Theorem 4.3. Also the estimates will be asymptotically normal and estimates of different orders will be asymptotically independent.

Suppose next that $g'_{a_1,\cdots,a_k}(\lambda_1,\cdots,\lambda_{k-1})$ vanishes for $|\lambda_j| > \Lambda$. As an estimate of $r_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1})$ we can then consider

$$(5.12)\qquad r^{(T)}_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}) = (2\pi)^{-k+1}\int\cdots\int_{|\lambda_j|\leqq\Lambda} \exp\left\{i\sum_1^{k-1}\lambda_j u_j\right\} g^{(T)}_{a_1,\cdots,a_k}\left(\lambda_1,\cdots,\lambda_{k-1},-\sum_1^{k-1}\lambda_j\right) d\lambda_1\cdots d\lambda_{k-1}.$$

From (5.11) this estimate will be asymptotically unbiased.

By analogy with Theorem 5.2 of Brillinger [6], we would expect, for example, that

$$(5.13)\qquad \lim_{T\to\infty} T\,\mathrm{Cov}\{r^{(T)}_{a_1,b_1}(u_1), r^{(T)}_{a_2,b_2}(u_2)\}$$
$$= \int_{-\Lambda}^{\Lambda} \exp\{i\alpha(u_1-u_2)\} f'_{a_1,a_2}(\alpha) f'_{b_1,b_2}(-\alpha)\,d\alpha$$
$$+ \int_{-\Lambda}^{\Lambda} \exp\{i\alpha(u_1+u_2)\} f'_{a_1,b_2}(\alpha) f'_{b_1,a_2}(-\alpha)\,d\alpha$$
$$+ 2\pi \iint_{-\Lambda}^{\Lambda} \exp\{i(\alpha_1u_1+\alpha_2u_2)\} f'_{a_1,b_1,a_2,b_2}(\alpha_1,-\alpha_1,\alpha_2)\,d\alpha_1 d\alpha_2,$$

in the case $k = 2$ and $a_j \neq b_j$.

Next one can take

$$(5.14)\qquad q^{(T)}_{a_1,\cdots,a_k}(t_1,\cdots,t_k) = r^{(T)}_{a_1,\cdots,a_k}(t_1-t_k,\cdots,t_{k-1}-t_k)$$

as an estimate of $q_{a_1,\cdots,a_k}(t_1,\cdots,t_k)$ and

$$(5.15)\qquad p^{(T)}_{a_1,\cdots,a_k}(t_1,\cdots,t_k) = \sum_{\ell=1}^{k} q^{(T)}_{a_j;\,j\in\nu_1}(t_j;j\in\nu_1)\cdots q^{(T)}_{a_j;\,j\in\nu_\ell}(t_j;j\in\nu_\ell)$$

as an estimate of $p_{a_1,\cdots,a_k}(t_1,\cdots,t_k)$.

In the case $k = 1$, we would estimate r_a by $N_a(0,T]/T$. In Theorem 4.2 we saw that this statistic was asymptotically normal with mean r_a and variance $2\pi T^{-1}f'_{a,a}(0)$.

6. Estimation of second order spectra from sampled values

Let $Y(t)$, $-\infty < t < \infty$, be a real valued time series satisfying the conditions of Example 3.1, having mean c'_Y and autocovariance function $c'_{Y,Y}(u)$, $-\infty < u < \infty$. Let $N(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, be an independent real valued point process satisfying Assumption 3.2, having mean intensity r_N and autocovariance density $r_{N,N}(u)$, $-\infty < u < \infty$. Suppose that events of a realization of the process $N(\Delta)$ occur at the times $\tau_1, \cdots, \tau_n$ in the interval $(0, T]$. Consider the problem of estimating the autocovariance $c'_{Y,Y}(u)$, $-\infty < u < \infty$, and power spectrum $f'_{Y,Y}(\lambda)$, $-\infty < \lambda < \infty$, of the series $Y(t)$ from the values

$$\tau_1, \cdots, \tau_n \tag{6.1}$$

and

$$Y(\tau_1), \cdots, Y(\tau_n). \tag{6.2}$$

We can construct a stationary interval process $X(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, in the manner of Example 3.3 by setting

$$X(\Delta) = \int_\Delta Y(t)\, dN(t), \tag{6.3}$$

or, in differential notation, by setting

$$dX(t) = Y(t)\, dN(t). \tag{6.4}$$

The first and second order measures of this process satisfy

$$C'_X\, dt = c'_Y r_N\, dt, \tag{6.5}$$

and

$$dC'_{X,X}(u)\, dt = \big(c'_{Y,Y}(u) r_{N,N}(u) + c'_{Y,Y}(u) r_N \delta(u) + c'_{Y,Y}(u) r_N^2 + (c'_Y)^2 r_{N,N}(u) + (c'_Y)^2 r_N \delta(u)\big)\, du\, dt. \tag{6.6}$$

The measure $C'_{X,X}(u)$ is seen to have absolutely continuous part and an atom of mass $c'_{Y,Y}(0) r_N + (c'_Y)^2 r_N$ at $u = 0$. If we let $r_{X,X}(u)$ denote the derivative of the absolutely continuous part of $C'_{X,X}(u)$ then, from (6.6),

$$r_{X,X}(u) = c'_{Y,Y}(u) r_{N,N}(u) + c'_{Y,Y}(u) r_N^2 + (c'_Y)^2 r_{N,N}(u) \tag{6.7}$$

for $-\infty < u < \infty$. For convenience set

$$h(u) = r_{X,X}(u) - (c'_Y)^2 r_{N,N}(u). \tag{6.8}$$

If

$$r_N^2 + r_{N,N}(u) \neq 0, \tag{6.9}$$

then, from (6.7),

$$c'_{Y,Y}(u) = \frac{h(u)}{[r_{N,N}(u) + r_N^2]}. \tag{6.10}$$

We see, from (6.8) and (6.10), that an estimate of $c'_{Y,Y}(u)$ may be constructed from estimates of $r_{X,X}(u)$, c'_Y, $r_{N,N}(u)$, r_N. One can then proceed to form an estimate of $f'_{Y,Y}(\lambda)$.

Alternatively we could proceed directly to the frequency domain and note that the power spectrum of the process $X(\Delta)$ is given by

$$(6.11)\qquad f'_{X,X}(\lambda) = (2\pi)^{-1}\left[\int r_{X,X}(u)\exp\{-i\lambda u\}\,du + c'_{Y,Y}(0)r_N + (c'_Y)^2 r_N\right]$$
$$= \int f'_{Y,Y}(\lambda-\alpha)g'_{N,N}(\alpha)\,d\alpha + f'_{Y,Y}(\lambda)r_N^2 + (c'_Y)^2 f'_{N,N}(\lambda) + (2\pi)^{-1}c'_{Y,Y}(0)r_N,$$

for $-\infty < \lambda < \infty$. If we rewrite this in the form

$$(6.12)\qquad f'_{Y,Y}(\lambda)r_N^2 + \int f'_{Y,Y}(\alpha)g'_{N,N}(\lambda-\alpha)\,d\alpha$$
$$= f'_{X,X}(\lambda) - (c'_Y)^2 f'_{N,N}(\lambda) - (2\pi)^{-1}c'_{Y,Y}(0)r_N$$
$$= H(\lambda),$$

then we have an integral equation for $f'_{Y,Y}(\lambda)$. This equation may be solved for $f'_{Y,Y}(\lambda)$, under the condition (6.9), as follows: set

$$(6.13)\qquad P(\lambda) = (2\pi)^{-1}\int \exp\{-i\lambda u\} r_{N,N}(u)/[r_N^2 + r_{N,N}(u)]\,du,$$

then

$$(6.14)\qquad f'_{Y,Y}(\lambda) = r_N^{-2}H(\lambda) - 2\pi r_N^{-2}\int P(\lambda-\alpha)H(\alpha)\,d\alpha.$$

Once estimates of r_N, $r_{N,N}(u)$, c'_Y, $c'_{Y,Y}(0)$, $f'_{N,N}(\lambda)$, $f'_{X,X}(\lambda)$, are available an estimate of $f'_{Y,Y}(\lambda)$ may be constructed from (6.14). The estimates may be determined as follows:

$$(6.15)\qquad r_N^{(T)} = n/T;$$

$$(6.16)\qquad c_Y^{(T)} = [Y(\tau_1) + \cdots + Y(\tau_n)]/n;$$

$$(6.17)\qquad m_{Y,Y}^{(T)}(0) = [Y(\tau_1)^2 + \cdots + Y(\tau_n)^2]/n;$$

$$(6.18)\qquad c_{Y,Y}^{(T)}(0) = m_{Y,Y}^{(T)}(0) - c_Y^{(T)2};$$

and finally estimates $f_{N,N}^{(T)}(\lambda)$, $f_{X,X}^{(T)}(\lambda)$ may be constructed in the manner of (4.16) or (4.34).

A problem related to the one just considered is that of obtaining as estimate of the cross spectrum $f'_{Y_1,Y_2}(\lambda)$ of a series $Y_1(t)$ with a series $Y_2(t)$ from the values

$$(6.19)\qquad \tau_1, \cdots, \tau_n,$$

$$(6.20)\qquad Y_1(\tau_1), \cdots, Y_1(\tau_n),$$

and

$$Y_2(\tau_1), \cdots, Y_2(\tau_n). \tag{6.21}$$

In this case the expression (6.11) is replaced by

$$f'_{X_1,X_2}(\lambda) = \int f'_{Y_1,Y_2}(\lambda - \alpha) g'_{N,N}(\alpha)\, d\alpha + f'_{Y_1,Y_2}(\lambda) r_N^2 + (c'_{Y_1})(c'_{Y_2}) f'_{N,N}(\lambda) + (2\pi)^{-1} c'_{Y_1,Y_2}(0) r_N. \tag{6.22}$$

A second related problem would be to construct an estimate of $f'_{Y_1,Y_2}(\lambda)$ from the values

$$\sigma_1, \cdots, \sigma_m, \tag{6.23}$$

$$\tau_1, \cdots, \tau_n, \tag{6.24}$$

$$Y_1(\sigma_1), \cdots, Y_1(\sigma_m), \tag{6.25}$$

$$Y_2(\tau_1), \cdots, Y_2(\tau_n), \tag{6.26}$$

where $\sigma_1, \cdots, \sigma_m$ are the times of events in $(0, T]$ of a point process $N_1(\Delta)$ and $\tau_1, \cdots, \tau_n$ are the times of events in $(0, T]$ of a related point process $N_2(\Delta)$ with the bivariate point process satisfying Assumption 3.2. In this case expression, (6.11) is replaced by the simpler expression

$$f'_{X_1,X_2}(\lambda) = \int f'_{Y_1,Y_2}(\lambda - \alpha) g_{N_1,N_2}(\alpha)\, d\alpha + f'_{Y_1,Y_2}(\lambda) r_{N_1} r_{N_2} + c_{Y_1} c_{Y_2} f'_{N_1,N_2}(\lambda). \tag{6.27}$$

7. Further considerations

We next discuss briefly some practical implications and extensions of the previous results. We saw, in Section 2, that if $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, was a stationary interval process with cumulant spectra

$$f_{X_{a_1},\cdots,X_{a_k}}(\lambda_1, \cdots, \lambda_k), \tag{7.1}$$

then

$$Y_a(t) = \int \phi_a(t - u)\, dX_a(u), \tag{7.2}$$

$a = 1, \cdots, r$, $-\infty < t < \infty$, was a stationary time series with cumulant spectra

$$f_{Y_{a_1},\cdots,Y_{a_k}}(\lambda_1, \cdots, \lambda_k) = \Phi_{a_1}(\lambda_1) \cdots \Phi_{a_k}(\lambda_k) f_{X_{a_1},\cdots,X_{a_k}}(\lambda_1, \cdots, \lambda_k). \tag{7.3}$$

This suggests that one might estimate the spectrum (7.1) by a statistic of the form

$$f^{(T)}_{Y_{a_1},\cdots,Y_{a_k}}(\lambda_1, \cdots, \lambda_k)/[\Phi_{a_1}(\lambda_1) \cdots \Phi_{a_k}(\lambda_k)], \tag{7.4}$$

having formed

$$f^{(T)}_{Y_a,\cdots,Y_a}(\lambda_1, \cdots, \lambda_k) \tag{7.5}$$

in the manner of Brillinger and Rosenblatt [9]. (In the case $k = 2$ this suggestion was made by Priestly in the discussion of Bartlett [3].) This procedure is seen to be analogous with the technique of prewhitening a time series prior to estimating its spectrum. This analogy suggests that we should choose the $\phi_a(\lambda)$ so that the spectrum (7.3) is near constant for λ_j in some finite region. The estimate (7.4) is seen to have the important advantage of allowing the use of existing spectral programs and also of allowing a simultaneous prewhitening of the data.

The proposed analysis may be related to the analysis of a continuous time series in another way. The basic statistic of our analysis is

$$\int_0^T \exp\{-i\lambda t\}h(t/T)\, d\mathbf{X}(t). \tag{7.6}$$

If we approximate (7.6) by a Stieltjes sum, then we obtain

$$\sum_{t=0}^{T-1} \exp\{-i\lambda t\}h(t/T)[\mathbf{X}(t+1) - \mathbf{X}(t)]. \tag{7.7}$$

An examination of expression (7.7) shows that it corresponds to carrying out an empirical spectral analysis on the time series of first differences. This procedure is common in the analysis of economic time series.

Computations involved in forming (7.6) may be prohibitive. Therefore there is much to be said for a procedure involving splitting the data into N segments of length S, forming an estimate

$$f^{(S)}_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_k)_n \tag{7.8}$$

for the nth segment, $n = 1, \cdots, N$, and taking

$$N^{-1} \sum_{n=1}^{N} f^{(S)}_{a_1,\cdots,a_k}(\lambda_1, \cdots, \lambda_k)_n \tag{7.9}$$

as a final estimate. Authors recommending such a procedure include: Bartlett [2], Welch [23], Lewis [19], and Huber *et al* [15]. The asymptotics of such estimates are directly determinable from the results of Theorem 4.3 because, following the remark after Theorem 4.1, Fourier transforms based on disjoint stretches of data are asymptotically independent. A variety of further remarks concerning practical aspects of the calculations in the case of a point process are made in Lewis [19].

We remark that the calculations proposed in this paper reduce, in the case that the interval process $\mathbf{X}(\Delta)$, $\Delta \in \boldsymbol{\Delta}$, is an integral of a continuous time series, to the usual calculations of the frequency analysis of time series.

Extensions of the definitions and theorems of this paper to a case in which t is vector valued, $t \in R^p$, appear fairly immediate if one takes the approach of

Brillinger [7]. A different sort of extension would result from a consideration of processes whose differences of higher order than the first are stationary (see Yaglom [24]).

8. Proofs

Proof of Lemma 2.1. If $M_{a_1,\cdots,a_k}(t_1,\cdots,t_k)$ corresponds to the measure determined by the coordinates $t_1,\cdots,t_k$, let $N_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},t_k)$ correspond to the measure determined by the coordinates $u_1 = t_1 - t_k,\cdots,u_{k-1} = t_{k-1} - t_k, t_k$. The initial measure is invariant under the transformation $t_1,\cdots,t_k \to t_1 + t,\cdots,t_k + t$. The second measure is therefore invariant under the transformation $t_k \to t_k + t$. We see therefore that

$$(8.1)\qquad \begin{aligned} &N_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},t_k) - N_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},0) \\ &\quad = N_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},t_k + t) - N_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1},t). \end{aligned}$$

Suppressing $a_1,\cdots,a_k,u_1,\cdots,u_{k-1}$ this last may be written

$$(8.2)\qquad N(t_k + t) = N(t_k) + N(t) - N(0).$$

Under the given conditions, all solutions of this functional equation have the form

$$(8.3)\qquad N(t_k) = M't_k + N(0),$$

giving the indicated result.

Proof of Theorem 2.1. Assume the results of Theorem 4.1 hold. It will be proved later. Set

$$(8.4)\qquad \mathbf{d}_X^{(T)}(\lambda) = \int_{-T}^{T} \exp\{-i\lambda t\}\, d\mathbf{X}(t),$$

using the notation of Section 4 with $h(t) = 1$ for $|t| \leqq 1$ and $h(t) = 0$ otherwise. One has therefore

$$(8.5)\qquad \mathbf{Z}_X^{(T)}(\lambda) = (2\pi)^{-1}\int_0^{\lambda} d_X^{(T)}(\alpha)\, d\alpha.$$

One now uses expression (4.7) to see that

$$(8.6)\qquad E|Z_a^{(T)}(\lambda) - Z_a^{(S)}(\lambda)|^{2k} \to 0$$

as $S, T \to \infty$ for $k = 1, 2, \cdots; a = 1, \cdots, r$. It follows that there exists $\mathbf{Z}_X(\lambda)$ such that $\mathbf{Z}_X^{(T)}(\lambda) \to \mathbf{Z}_X(\lambda)$ in mean of order ν for any $\nu > 0$.

One next checks that

$$(8.7)\qquad \begin{aligned} &\operatorname{cum}\{Z_{a_1}^{(T)}(\lambda_1),\cdots,Z_{a_k}^{(T)}(\lambda_k)\} \\ &\quad \to \int_0^{\lambda_1}\cdots\int_0^{\lambda_k} \delta\Big(\sum_1^k \alpha_j\Big) f_{a_1,\cdots,a_k}(\alpha_1,\cdots,\alpha_k)\, d\alpha_1\cdots d\alpha_k \end{aligned}$$

as $T \to \infty$, again using expression (4.7). This gives (2.27).

Finally one checks that

$$(8.8)\qquad E\left|X_a(\Delta) - \int_{-\infty}^{\infty}\left[\int_{\Delta}\exp\{i\lambda t\}\,dt\right]dZ_a^{(T)}(\lambda)\right|^2 \to 0$$

as $T \to \infty$. This gives (2.28).

PROOF OF THEOREM 3.1. We first state and prove a lemma.

LEMMA 8.1. *With the conditions and notation of Assumption* 3.1,

$$(8.9)\qquad \lim_{|\Delta_j|\to 0} |\Delta_1|^{-1}\cdots|\Delta_k|^{-1}E\{N_{a_1}(\Delta_1)\cdots N_{a_{m_1}}(\Delta_1)]\cdots$$

$$\cdots[N_{a_{m_{k-1}+1}}(\Delta_k)\cdots N_{a_{m_k}}(\Delta_k)]\}$$

$$= \sum_{\alpha_1,\cdots,\alpha_k=1}^{r}\left[\prod_{j=1}^{m_1}\delta\{a_j-\alpha_1\}\right]$$

$$\cdots\left[\prod_{j=m_{k-1}+1}^{m_k}\delta\{a_j-\alpha_k\}\right]p_{\alpha_1,\cdots,\alpha_k}(t_1,\cdots,t_k)$$

uniformly in $t_1,\cdots,t_k$ *for integers* $1 \leqq m_1 < m_2 < \cdots < m_{k-1} < m_k$.

PROOF. Suppose first that

$$(8.10)\qquad \begin{aligned} a_1,\cdots,a_{m_1} &= \alpha_1,\\ &\vdots\\ a_{m_{k-1}+1},\cdots,a_{m_k} &= \alpha_k.\end{aligned}$$

Now

$$(8.11)\qquad \begin{aligned} &E\{N_{\alpha_1}(\Delta_1)^{m_1}\cdots N_{\alpha_k}(\Delta_k)^{m_k-m_{k-1}}\}\\ &= \sum_{n_j\geqq 1} n_1^{m_1}\cdots n_k^{m_k-m_{k-1}}P[N_{\alpha_1}(\Delta_1)=n_1,\cdots,N_{\alpha_k}(\Delta_k)=n_k]\\ &= P[N_{\alpha_1}(\Delta_1)=1,\cdots,N_{\alpha_k}(\Delta_k)=1] + \sum n_1^{m_1}\cdots n_k^{m_k-m_{k-1}}L(n_1,\cdots,n_k),\end{aligned}$$

with the second summation extending over some $n_j \geqq 2$ and with $|L(n_1,\cdots,n_k)| \leqq K_\delta|\Delta_1|^{n_1}\cdots|\Delta_k|^{n_k}$ from (3.8), and so

$$(8.12)\qquad \lim_{|\Delta j|\to 0}|\Delta_1|^{-1}\cdots|\Delta_k|^{-1}E\{N_{\alpha_1}(\Delta_1)^{m_1}\cdots N_{\alpha_k}(\Delta_k)^{m_k-m_{k-1}}\} = p_{\alpha_1,\cdots,\alpha_k}(t_1,\cdots,t_k),$$

uniformly in $t_1,\cdots,t_k$ from (3.9). Continuing if (8.10) is not satisfied for some $\alpha_1,\cdots,\alpha_k$, then one can see from (3.8) that the limit in (8.9) is 0 uniformly in $t_1,\cdots,t_k$. This completes the proof of the lemma.

Turning to the proof of the theorem; let $\phi_j(t)$ be continuous is some interval of Δ and 0 elsewhere for $j = 1,\cdots,k$. We have

$$(8.13)\qquad \int\phi_j(t)\,dN_{a_j}(t) = \lim_{\varepsilon\to 0}\sum_i\phi_j(i\varepsilon)N_{a_j}(i\varepsilon, i\varepsilon+\varepsilon].$$

By bounded convergence,

$$(8.14)\quad E\left\{\int \phi_1(t)\,dN_{a_1}(t)\cdots\int \phi_k(t)\,dN_{a_k}(t)\right\}$$

$$= \lim_{\varepsilon\to 0}\sum_{i_1}\cdots\sum_{i_k}\phi_1(i_1\varepsilon)\cdots\phi_k(i_k\varepsilon)E\{N_{a_1}(i_1\varepsilon, i_1\varepsilon+\varepsilon]\cdots N_{a_k}(i_k\varepsilon, i_k\varepsilon+\varepsilon]\}$$

$$= \lim_{\varepsilon\to 0}\sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha=1}^{r}[\prod_{j\in\nu_1}\delta\{a_j-\alpha_1\}]\cdots[\prod_{j\in\nu_\ell}\delta\{a_j-\alpha_\ell\}]$$

$$\cdot\sum_{i_1}\cdots\sum_{i_\ell}[\prod_{j\in\nu_1}\phi_j(i_1\varepsilon)]\cdots[\prod_{j\in\nu_\ell}\phi_j(i_\ell\varepsilon)]p_{\alpha_1,\cdots,\alpha}(i_1\varepsilon,\cdots,i_\ell\varepsilon)\varepsilon^\ell,$$

where the summation extends over partitions $(\nu_1,\cdots,\nu_\ell)$ of $(1,\cdots,k)$ if we separate out terms in (8.14) with the same argument and use Lemma 8.1. We now see that expression (8.14) equals

$$(8.15)\quad \sum_{\ell=1}^{k}\sum_{\alpha_1,\cdots,\alpha_\ell=1}^{r}[\prod_{j\in\nu_1}\delta\{a_j-\alpha_1\}]\cdots[\prod_{j\in\nu_\ell}\delta\{a_j-\alpha\}]$$

$$\cdot\int\cdots\int[\prod_{j\in\nu_1}\phi_j(\tau_1)]\cdots[\prod_{j\in\nu_\ell}\phi_j(\tau_\ell)]p_{\alpha_1,\cdots,\alpha_\ell}(\tau_1,\cdots,\tau_\ell)\,d\tau_1\cdots d\tau_\ell.$$

Expression (3.13) now follows from (8.15) taking the $\phi_j(t)$ to be indicator functions.

Proof of Theorem 3.2. One proves (3.28) from (8.15) and then obtains (3.22) by taking the $\phi_j(t)$ to be indicator functions.

Proof of Theorem 3.3. This follows directly from (3.28).

Proof of Theorem 4.1. Let $h_a^{(T)}(t) = h_a(t/T)$. The cumulant at issue is given by

$$(8.16)\quad \int\cdots\int h_{a_1}^{(T)}(t_1)\cdots h_{a_k}^{(T)}(t_k)$$

$$\cdot\exp\left\{-i\sum_1^k\lambda_jt_j\right\}dC'_{a_1,\cdots,a_k}(t_1-t_k,\cdots,t_{k-1}-t_k)\,dt_k$$

$$=\int\cdots\int\left[\int_t h_{a_1}^{(T)}(u_1+t)\cdots h_{a_{k-1}}^{(T)}(u_{k-1}+t)h_{a_k}^{(T)}(t)\right.$$

$$\left.\cdot\exp\left\{-i\sum_1^k\lambda_jt\right\}dt\right]\exp\left\{-i\sum_1^{k-1}\lambda_ju_j\right\}dC'_{a_1,\cdots,a_k}(u_1,\cdots,u_{k-1}).$$

The indicated result now follows as

$$(8.17)\quad \left|\int_t\{h_{a_1}^{(T)}(u_1+t)\cdots h_{a_{k-1}}^{(T)}(u_{k-1}+t)h_{a_k}^{(T)}(t)\right.$$

$$\left.-h_{a_1}^{(T)}(t)\cdots h_{a_k}^{(T)}(t)\exp\{-i\lambda t\}\,dt\right|\leqq C\sum_1^{k-1}|u_j|,$$

for some finite C following Assumption 4.1.

PROOF OF THEOREM 4.2. This follows directly from Theorem 4.1 in the manner of corresponding results in Brillinger [7] and Brillinger [8].

PROOF OF THEOREM 4.3. This follows directly from Theorem 4.1 in the manner of the principal theorems in Brillinger and Rosenblatt [9].

PROOF OF THEOREM 4.4. This follows directly from Theorem 4.1 in the manner of corresponding results in Brillinger [7] and Brillinger [8].

REFERENCES

[1] H. J. BAHBA, "On the stochastic theory of continuous parametric systems and its application to electron cascades," *Proc. Roy. Soc. Ser. A*, Vol. 202 (1950), pp. 301–322.

[2] M. S. BARTLETT, "Smoothing periodograms from time-series with continuous spectra," *Nature*, Vol. 161 (1948), pp. 686–687.

[3] ———, "The spectral analysis or point processes," *J. Roy. Statist. Soc. Ser. B*, Vol. 25 (1963), pp. 264–296.

[4] ———, *Stochastic Processes*, Cambridge, Cambridge University Press, 1966.

[5] S. BOCHNER, *Harmonic Analysis and the Theory of Probability*, Berkeley and Los Angeles, University of California Press, 1960.

[6] D. R. BRILLINGER, "Asymptotic properties of spectral estimates of second order," *Biometrika*, Vol. 56 (1969), pp. 375–390.

[7] ———, "The frequency analysis of relations between stationary spatial series," *Proceedings Twelfth Biennial Seminar*, Montreal, Canadian Mathematical Congress, 1970, pp. 39–81.

[8] ———, *The Frequency Analysis of Vector-Valued Time Series*, New York, Holt, Rinehart, and Winston, to appear 1971.

[9] D. R. BRILLINGER and M. ROSENBLATT, "Asymptotic theory of estimates of kth order spectra," *Spectral Analysis of Time Series* (edited by B. Harris), New York, Wiley, 1967, pp. 153–188.

[10] D. R. BRILLINGER and J. W. TUKEY, *Asymptotic Variances, Moments, Cumulants, and Other Averaged Values*, 1964, unpublished.

[11] D. R. COX and P. A. W. LEWIS, *The Statistical Analysis of Series of Events*, London, Methuen, 1966.

[12] H. CRAMÉR and M. R. LEADBETTER, *Stationary and Related Stochastic Processes*, New York, Wiley, 1967.

[13] D. J. DALEY, "Spectral properties of weakly stationary point processes," *J. Roy. Statist. Soc. Ser. B*, Vol. 33 (1971).

[14] J. L. DOOB, *Stochastic Processes*, New York, Wiley, 1953.

[15] P. J. HUBER, B. KLEINER, TH. GASSER, G. DUMERMUTH, "Statistical methods for investigating phase relations in stationary stochastic processes," International Seminar on Digital Processing of Analog Signals, Zürich, 1970.

[16] K. ITO, "Stationary random distributions," *Mem. Coll. Sci. Univ. Kyoto Ser. A Math.*, Vol. 28 (1954), pp. 209–224.

[17] A. N. KOLMOGOROV, "Curves in Hilbert space invariant with regard to a one parameter group of motions," *Dokl. Akad. Nauk SSSR*, Vol. 26 (1940), pp. 6–9.

[18] P. I. KUZNETSOV and R. L. STRATONOVICH, "A note on the mathematical theory of correlated random points," *Non-Linear Transformations of Stochastic Processes* (edited by P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov), Oxford, Pergamon, 1965, pp. 101–115.

[19] P. A. W. LEWIS, "Remarks on the theory, computation and application of the spectral analysis of series of events," *J. Sound Vib.*, Vol. 12 (1970), pp. 353–375.

[20] A. RAMAKRISHNAN, "Stochastic processes relating to particles distributed in a continuous infinity of states," *Proc. Cambridge Philos. Soc.*, Vol. 46 (1950), pp. 595–602.

[21] S. K. SRINIVASAN, *Stochastic Theory and Cascade Processes*, New York, Elsevier, 1969.
[22] J. W. TUKEY, "An introduction to the calculations of numerical spectrum analysis," *Spectral Analysis of Time Series* (edited by B. Harris), New York, Wiley, 1967, pp. 25–46.
[23] P. D. WELCH, "A direct digital method of power spectrum estimation," *IBM J. Res. Develop.*, Vol. 5 (1961), pp. 141–156.
[24] A. M. YAGLOM, "Correlation theory of processes with stationary n-th order increments," *Amer. Math. Soc. Transl.*, Vol. 8 (1958), pp. 87–142.
[25] ———, *An Introduction to the Theory of Stationary Random Functions*, Englewood Cliffs, Prentice-Hall, 1962.

The Annals of Mathematical Statistics
1972, Vol. 43, No. 2, 534-540

ON THE NUMBER OF SOLUTIONS OF SYSTEMS OF RANDOM EQUATIONS

BY DAVID R. BRILLINGER

University of California, Berkeley

Let $\{f(x, \omega); x \in R^n, \omega \in \Omega\}$ be an n vector-valued stochastic process defined over a probability space $(\Omega, \mathscr{A}, \mu)$. Let $N(f \mid A, y)$ denote the number of elements in the set $A \cap f^{-1}(y)$, that is the number of distinct solutions of the system of equations $f(x, \omega) = y$ for $x, y \in R^n$. We develop expressions for $E\{N(f \mid A, y)\}$ and certain higher-order moments of $N(f \mid A, y)$ under regularity conditions.

1. Introduction. A variety of statistical properties have been developed for the number of solutions of an equation

$$f(x) = y \tag{1.1}$$

in the case that $x, y \in R$ and f is a random function. See, for example, Kac (1943), Rice (1945), Cramér and Leadbetter (1967). Properties have also been developed in the case that $x, y \in C$ and f is a random analytic function, see Paley and Wiener (1934, page 178), Littlewood and Offord (1948), Offord (1965), Offord (1967). In this case (1.1) is equivalent with two real random equations in two real unknowns. Here we determine the expected value and the factorial moments of the number of solutions of n real random equations in n real unknowns under regularity conditions. The results obtained have application to the investigation of the number of extreme points of a random surface defined over R^n, for the extreme points are the solutions of the n equations resulting from setting the first derivatives of the surface to zero. We note that Longuett-Higgins (1957) has investigated the expected number of extreme points for certain random surfaces.

The proofs of the lemmas and theorems of Sections 2 and 3 of the paper have been collected in Section 4.

2. The non-stochastic case. In this section we develop an expression for the number of solutions of a system of n fixed equations in n unknowns. The expression provides a generalization of one due to Kac (1943). In what follows; if $y = (y_1, \cdots, y_n) \in R^n$, the region $|y_1|, \cdots, |y_n| < \varepsilon$ is denoted $|y| < \varepsilon$. If $A \subset R^n$, and f maps R^n into R^n, the restriction of f to A is denoted $f \mid A$. If $f: R^n \to R^n$ is Lipschitz, see Federer (1969), its Jacobian determinant existing almost everywhere is denoted Jf. The number of elements in the set $A \cap f^{-1}(y)$, $y \in R^n$, $A \subset R^n$, is denoted $N(f \mid A, y)$. This is the desired number of distinct solutions of (1.1) in the set A.

Received November 16, 1970; revised September 19, 1971.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_5, © Springer Science+Business Media, LLC 2012

For $\varepsilon > 0$, $v \in R^n$ set

$$N_\varepsilon(f \mid A, y) = (2\varepsilon)^{-n} \int_{|v|<\varepsilon} N(f \mid A, y + v)dv \ . \tag{2.1}$$

Also set

$$\psi_\varepsilon(v) = 1 \quad \text{for } |v| < \varepsilon \,, \qquad = 0 \quad \text{otherwise} \,. \tag{2.2}$$

Then we have,

LEMMA 2.1. *Let A be a measurable subset of R^n and $f: A \to R^n$ be Lipschitz. Then*

$$N_\varepsilon(f \mid A, y) = (2\varepsilon)^{-n} \int_A \psi_\varepsilon[f(x) - y] |Jf(x)| dx \ . \tag{2.3}$$

If in addition

$$\int_A |Jf(x)| dx < \infty \,, \tag{2.4}$$

then

$$N(f \mid A, y) = \lim_{\varepsilon \to 0} (2\varepsilon)^{-n} \int_A \psi_\varepsilon[f(x) - y] |Jf(x)| dx \tag{2.5}$$

for almost all y and indeed

$$\lim_{\varepsilon \to 0} \int |N(f \mid A, y) - N_\varepsilon(f \mid A, y)| dy = 0 \ . \tag{2.6}$$

Finally if $N(f \mid A, u)$ is continuous $u = y$, then (2.5) *folds for that y.*

Expression (2.5) is the promised formula for the number of solutions of interest. The next lemma indicates one set of conditions under which $N(f \mid A, u)$ is continuous at $u = y$. We say that a continuously differentiable $f: R^n \to R^n$ is *normal* above $y \in R^n$ if $Jf(x) \neq 0$ for $x \in f^{-1}(y)$, (see Whitney (1957), page 145).

LEMMA 2.2. *Let A be an open bounded subset of R^n and let $f: A \to R^n$ be normal over y. Then $N(f \mid A, u)$ is continuous at $u = y$.*

This lemma, together with Lemma 2.1, indicates that (2.5) holds for given y if f is normal above y.

3. The stochastic case. We now turn to a determination of the mean number of solutions of a random equation $f(x, \omega) = y$ falling in a set A in the case that $f(x, \omega)$ is a vector-valued stochastic process. We have,

THEOREM 3.1. *Let A be a measurable subset of R^n. Let $\{f(x, \omega); x \in R^n; \omega \in \Omega\}$ be an n vector-valued stochastic process defined over a probability space $(\Omega, \mathscr{A}, \mu)$. Let $f(x, \omega)$ be Lipschitz with probability one for $x \in A$. Let the variates $\alpha = f(x, \omega)$, $\beta = Jf(x, \omega)$ have joint density $p(\alpha; \beta; x)$, $\alpha \in R^n$, $\beta \in R$, $x \in A$, satisfying*

$$\int\int\int_A |\beta| p(\alpha; \beta; x) d\alpha d\beta dx < \infty \ . \tag{3.1}$$

Then

$$E\{N(f \mid A, y)\} = \int\int_A |\beta| p(y; \beta; x) d\beta dx \tag{3.2}$$

for almost all $y \in R^n$.

Expression (3.2) was set down by Rice (1945) in the case $n = 1$. We remark that if $p(\alpha; x)$ denotes the density of $\alpha = f(x, \omega)$, then an alternate form for (3.2), involving a conditional expected value, is

$$E\{N(f \mid A, y)\} = \int_A p(y; x) E\{|Jf(x)| : f(x) = y\} dx \,. \tag{3.3}$$

The solutions of $f(x, \omega) = y$ determine a multidimensional point process in R^n · (These are discussed in Srinivasavan (1969).) If A is taken to be a small parallelipiped of volume $|A|$ and $x \in A$, then from (3.2)

$$E\{N(f \mid A, y)\} = |A| \int |\beta| p(y; \beta; x) d\beta \,, \tag{3.4}$$

showing that the intensity parameter of this point process is $\int |\beta| p(y; \beta; x) d\beta$.

One application of (3.2) is to provide a bound for crossing probabilities of the form Prob$[f(x, \omega) = y$ for some $x \in A]$. Clearly this probability is less than or equal to $E\{N(f \mid A, y)\}$. We may conclude, for example, that the probability is zero if (3.2) holds and $p(y; \beta; x) = 0$ for almost all $\beta \in R^n$, $x \in A$.

Theorem 3.1 provides the expected number of solutions for almost all $y \in R^n$. If some particular value of y is of interest, then the following result may be of use.

COROLLARY 3.1. *Under the conditions of the theorem and if* (i) $N(f \mid A, u)$ *is continuous at* $u = y$ *with probability one*, (ii) $E\{N(f \mid A, u)^{1+\delta}\} < \infty$ *for some* $\delta > 0$ *and for* u *in a neighborhood of* y, (iii) $\int\int_A |\beta| p(u; \beta; x) d\beta dx$ *is continuous at* $u = y$, *then* (3.2) *holds.*

We remark that it follows from Lemma 2.2 that (i) holds if the sample paths $f(x; \omega)$ are normal over y for almost all ω.

We next turn to the investigation of a function related to the higher order moments of the number of solutions. Given measurable subsets $A_1, \cdots, A_k$ of R^n and $f: R^n \to R^n$ consider the number of solutions of the system of equations

$$f(x_1) = y_1, \cdots, f(x_k) = y_k \tag{3.5}$$

for $y_1, \cdots, y_k \in R^n$ with $x_j \in A_j$, $x_i \neq x_j$, $1 \leqq i < j \leqq k$. In the case that the A_j are disjoint, the number of solutions is

$$N(f \mid A_1, y_1) \cdots N(f \mid A_k, y_k) \,. \tag{3.6}$$

In the case that $A_j = A$, $y_j = y$, $N = N(f \mid A, y)$ the number of solutions is

$$N(N - 1) \cdots (N + k + 1) \,. \tag{3.7}$$

Letting $B = \{(x_1, \cdots, x_k) : x_j \in A_j,\ x_i \neq x_j,\ 1 \leqq i < j \leqq k\}$, denoting the map of (3.5) by $\tilde{f}: R^{nk} \to R^{nk}$ and letting $N(f \mid B, y_1, \cdots, y_k)$ denote the number of solutions of (3.5) falling in B we have,

THEOREM 3.2. *Let* $A_1, \cdots, A_k$ *be measurable subsets of* R^n. *Let* $\{f(x, \omega);\ x \in R^n,\ \omega \in \Omega\}$ *be an* n *vector-valued stochastic process defined over the probability space* $(\Omega, \mathscr{A}, \mu)$. *Let* $f(x, \omega)$ *be Lipschitz with probability one for* $x \in A_1 \cdots A_k$.

Let the variates $\alpha_j = f(x_j, \omega)$, $\beta_j = Jf(x_j, \omega)$, $j = 1, \cdots, k$ *have joint density* $p(\alpha_1, \cdots, \alpha_k; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k)$ *for distinct* x_j, $x_j \in A_j$ *with*

$$\int \cdots \int\int_{A_1} \cdots \int_{A_k} |\beta_1| \cdots |\beta_k| p(\alpha_1, \cdots, \alpha_k; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k) \times \prod_1^k (d\alpha_j d\beta_j dx_j) \tag{3.8}$$

finite. Then

$$E\{N(\tilde{f} \mid B, y_1, \cdots, y_k)\} = \int \cdots \int\int_{A_1} \cdots \int_{A_k} |\beta_1| \cdots |\beta_k| \times p(y_1, \cdots, y_k; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k) \times \prod_1^k (d\beta_j dx_j) \tag{3.9}$$

for almost all $y_1, \cdots, y_k \in R^n$.

As one implication of this theorem, we mention that if $A_1, \cdots, A_k$ are small disjoint parallelipipeds of volumes $|A_1|, \cdots, |A_k|$ and $x_j \in A_j$, then

$$\begin{aligned} &E\{N(f \mid A_1, y_1) \cdots N(f \mid A_k, y_k)\} \\ &\quad = |A_1| \cdots |A_k| \int \cdots \int |\beta_1| \cdots |\beta_k| \\ &\qquad \times p(y_1, \cdots, y_k; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k) d\beta_1 \cdots d\beta_k \end{aligned} \tag{3.10}$$

and so

$$\int \cdots \int |\beta_1| \cdots |\beta_k| p(y_1, \cdots, y_k; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k) d\beta_1 \cdots d\beta_k \tag{3.11}$$

may be interpreted as a product density of order k (see Srinivasavan (1969)) of the multidimensional point process resulting from the solutions of (3.5).

If one is interested in the factorial moment of order k of $N = N(f \mid A, y)$ for some prespecified y one has,

COROLLARY 3.2. *Under the conditions of the theorem and if* (i) $N(\tilde{f} \mid B, u_1, \cdots, u_k)$ *is continuous at* $(u_1 \cdots, u_k) = (y, \cdots, y)$ *with probability one,* (ii) $E\{N(\tilde{f} \mid B, u_1, \cdots, u_k)^{1+\delta}\} < \infty$ *for some* $\delta > 0$ *and for* $u_1, \cdots, u_k$ *in a neighborhood of* $(y, \cdots, y)$, (iii) $\int \cdots \int\int_A \cdots \int_A |\beta_1| \cdots |\beta_k| p(u_1, \cdots, u_k; \beta_1, \cdots, \beta_k; x_1 \cdots, x_k) d\beta_1 \cdots d\beta_k dx_1 \cdots dx_k$ *is continuous at* $(u_1, \cdots, u_k) = (y, \cdots, y)$, *then*

$$\begin{aligned} &E\{N(N-1) \cdots (N-k+1)\} \\ &\quad = \int \cdots \int\int_A \cdots \int_A |\beta_1| \cdots |\beta_k| \\ &\qquad \times p(y, \cdots, y; \beta_1, \cdots, \beta_k; x_1, \cdots, x_k) d\beta_1 \cdots d\beta_k dx_1 \cdots dx_k \,. \end{aligned} \tag{3.12}$$

We mention that $\tilde{f}$ will be normal above $(y, \cdots, y)$ when f is normal above y and so following Lemma 2.2 (i) above will hold in the case that $A_1, \cdots, A_k$ are open and bounded and f is normal above y with probability one. Expression (3.12) was set down by Cramér and Leadbetter (1967) in the case of Gaussian $f(x, \omega)$ and $n = 1$.

4. Proofs. We begin with a proof of Lemma 2.1.

PROOF OF LEMMA 2.1. Kirszbaum's Theorem (see Federer (1969), page 201) indicates the existence of a Lipschitz extension of f with domain R^n. Theorem 3.2.5 of Federer (1969) (or Theorem 2, page 374 in Rado and Reichelderfer (1955)) then applies to give

$$\int_A g[f(x)]\,|Jf(x)|dx = \int_{R^n} g(u)N(f\mid A, u)du \tag{4.1}$$

for measurable $g\colon R^n \to R$. Taking $g(u) = (2\varepsilon)^{-n}\psi_\varepsilon(u - y)$ in (4.1) gives (2.3) after a change in variable.

Taking $g(u) = 1$, shows that

$$\int N(f\mid A, u)du = \int_A |Jf(x)|dx \tag{4.2}$$

and so $N(f\mid A, u)$ is integrable in view of (2.4). The conclusions of the lemma now follow from a standard convergence theorem (see, for example, Theorems 1.1.1, 1.3.2 in Bochner (1960).)

Proof of Lemma 2.2. Under the stated conditions the set of solutions can have no limit points for the Jacobian would then vanish at some solution. The solutions are therefore isolated and finite in number. The Inverse Function Theorem then applies to give the existence of a continuously differentiable inverse in the neighborhood of each solution. If y is altered by a sufficiently small amount, it follows that the number of solutions is unchanged and so N is continuous.

Proof of Theorem 3.1. We begin by noting, from (2.1), (4.2), that

$$\int N(f\mid A, u)du\,,\quad \int N_\varepsilon(f\mid A, u)du = \int_A |Jf(x)|dx \tag{4.3}$$

and therefore, in view of (3.1),

$$E\{\int N(f\mid A, u)du\}\,,\quad E\{\int N_\varepsilon(f\mid A, u)du\} < \infty\,. \tag{4.4}$$

In consequence, it follows from bounded convergence, Fubini's Theorem and (2.6) that

$$\begin{aligned}\lim_{\varepsilon\to 0}\int |E\{N(f\mid A, u)\} - E\{N_\varepsilon(f\mid A, u)\}|du \\ = \lim_{\varepsilon\to 0} E\{\int|N(f\mid A, u) - N_\varepsilon(f\mid A, u)|du\} \\ = 0\end{aligned} \tag{4.5}$$

At the same time, we have from (2.3),

$$E\{N_\varepsilon(f\mid A, y)\} = (2\varepsilon)^{-n}\int\int\int_A \psi_\varepsilon(\alpha - y)|\beta|p(\alpha;\beta;x)d\alpha d\beta dx \tag{4.6}$$

and so

$$\lim_{\varepsilon\to 0}\int |E\{N_\varepsilon(f\mid A, y)\} - \int\int_A |\beta|p(y;\beta;x)d\beta dx|dy = 0 \tag{4.7}$$

(by Theorem 1.3.2 of Bochner (1960).) Expressions (4.5) and (4.6) together now give

$$\int |E\{N(f\mid A, y)\} - \int\int_A |\beta|p(y;\beta;x)d\beta dx|dy = 0 \tag{4.8}$$

and thence the conclusion of the theorem.

PROOF OF COROLLARY 3.1. Under the stated conditions

$$\begin{aligned} E\{N(f \mid A, y)\} &= E\{\lim_{u\to y} N(f \mid A, u)\} \\ &= \lim_{u\to y} E\{N(f \mid A, u)\} \\ &= \int\int_A |\beta| p(y; \beta; x) d\beta dx . \end{aligned}$$

PROOF OF THEOREM 3.2. B is a measurable subset of R^{nk}. The Jacobian of the map $\tilde{f}$ is given by $Jf(x_1) \cdots Jf(x_k)$. The conclusion of the theorem now follows directly from Theorem 3.1 taking n to be nk, A to be B and f to be $\tilde{f}$.

PROOF OF COROLLARY 3.2. This result follows directly from Corollary 3.1 in the above manner.

5. Concluding remarks.

We note here that the results obtained are easily modified, in the manner of Leadbetter (1966), to yield the moments of the number of solutions of the equation

$$f(x) = g(x) \tag{5.1}$$

for a fixed measurable n vector-valued function g.

The reader will have noted that the results obtained required expression (4.1) in an essential manner. In fact Federer, Theorem 3.2.5 develops expression (4.1) in the more general setting of maps $f: R^m \to R^n$ with $m \leqq n$ using Hausdorf m-measure. This suggests the possibility of extending the Theorems of this paper to apply to n vector-valued stochastic processes $f(x, \omega)$, $x \in R^m$, $m \leqq n$.

In another direction we mention that if A is a bounded open set, $f: \bar{A} \to R^n$ is continuously differentiable and $\mu(f \mid A, y)$ is the topological index of the mapping f with domain A at the point y (see Rado and Reichelderfer (1955), page 125), then as an analog of (4.1) one has

$$\int_A g[f(x)] Jf(x) dx = \int g(u) \mu(f \mid A, u) du \tag{5.2}$$

(*ibid.* page 374) and so one has, for example, under the conditions of Theorem 3.1

$$E\{\mu(f \mid A, y)\} = \int\int_A \beta p(y; \beta; x) d\beta dx . \tag{5.3}$$

I would like to thank Professor M. W. Hirsch for suggesting Lemma 2.2 to me.

REFERENCES

[1] BOCHNER, S. (1960). *Harmonic Analysis and the Theory of Probability*. Univ. of California Press.

[2] CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York.

[3] FEDERER, H. (1969). *Geometric Measure Theory*. Springer-Verlag, Berlin.

[4] KAC, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* **49** 314–320.

[5] LEADBETTER, M. R. (1966). On crossings of levels and curves by a wide class of stochastic processes. *Ann. Math. Statist.* **37** 260–267.

[6] LITTLEWOOD, J. E. and OFFORD, A. C. (1948). On the distribution of zeros and *a*-values of a random integral function II. *Ann. of Math.* **49** 885–952.

[7] LONGUET-HIGGINS, M. S. (1957). The statistical analysis of a random moving surface. *Philos. Trans. Roy. Soc. London Ser. A* **249** 321–387.

[8] OFFORD, A. C. (1965). The distribution of the values of an entire function whose coefficients are independent random variables I. *Proc. London Math. Soc.* (3) **14A** 199–238.

[9] OFFORD, A. C. (1967). The distribution of zeros of power series whose coefficients are independent random variables. *Indian J. Math.* **9** 175–196.

[10] PALEY, R. E. A. C. and WIENER, N. (1934). *Fourier Transforms in the Complex Domain.* Amer. Math. Soc, Providence.

[11] RADO, T. and REICHELDERFER, P. V. (1955). *Continuous Transformations in Analysis.* Springer-Verlag, Berlin.

[12] RICE, S. O. (1945). Mathematical analysis of random noise. *Bell System Tech. J.* **24** 46–156.

[13] SRINIVASAN, S. K. (1969). *Stochastic Theory and Cascade Processes.* Elsevier, New York.

[14] WHITNEY, H. (1957). *Geometric Integration Theory.* Princeton Univ. Press.

[**Note added in proof.** Beljaev (1967), *Soviet Math. Dokl.* **8** 1107–1109, has announced results that are extensions of those of Longuet-Higgins (1957).]

JOURNAL OF MULTIVARIATE ANALYSIS 12, 64–71 (1982)

Asymptotic Normality of Finite Fourier Transforms of Stationary Generalized Processes*

DAVID R. BRILLINGER

The University of California, Berkeley

Communicated by M. Rosenblatt

This paper indicates a mixing condition under which a net of Fourier transforms, of a stationary generalized process over an abelian locally compact group, has a limiting normal distribution.

1. INTRODUCTION

Finite Fourier transforms of stationary mixing processes have been shown to be asymptotically normal in quite a variety of circumstances. The case of a time series $X(t)$ with t in R is considered in, for example, Leonov and Shiryaev [12], Picinbono [16], Rosenblatt [19], Rozanov [21]. The case of t in Z is considered in Hannan [8, Chap.IV], in Hannan and Thomson [9], in Brillinger [5]. The case of t in R^p is investigated in Brillinger [2] and in Pichard [17]. Morettin [13] considers the case of t in a non-compact locally compact second countable group. In extensions of another sort Brillinger [3] indicates conditions leading to asymptotic normality in the case of stationary random measures on R and Brillinger [4] in the case of stationary random Schwartz distributions with t in R^p. This paper provides a central limit theorem for Fourier transforms of stationary random Schwartz–Bruhat distributions over a locally compact abelian group G. In the general case nets of Fourier transforms are shown to be asymptotically normal. When G is σ-compact the nets become sequences. The Fourier transforms need not be asymptotically normal. Rosenblatt [20] derives a non-Gaussian limit for the transform of a process with long range dependence.

* Prepared with the partial support of National Science Foundation Grant PFR 7901642.

Received April 1981; revised May 24, 1981.
AMS 1970 subject classification: Primary 60B15, 60G20, 60F05; Secondary 43A25, 60M15.
Key words and phrases: asymptotic normality, central limit theorem, Fourier transform, generalized process, group, mixing.

0047-259X/82/010064–08$02.00/0

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_6, © Springer Science+Business Media, LLC 2012

In this work G denotes a locally compact abelian group, Λ denotes its dual. Haar measures on these two groups are denoted by dg and $d\lambda$, respectively. The group operation is denoted by "+". Characters of the groups are denoted by $\langle g, \lambda \rangle$ for g in G and λ in Λ. The Fourier transform of a function ψ in $L_1(G)$ is defined by

$$\Psi(\lambda) = \int \langle g, -\lambda \rangle \, \psi(g) \, dg. \tag{1.1}$$

If Ψ belongs to $L_1(\Lambda)$ one has the inversion formula

$$\psi(g) = \int \langle g, \lambda \rangle \, \Psi(\lambda) \, d\lambda. \tag{1.2}$$

In the case of Euclidean space, the infinitely differentiable functions of rapid decrease provide a setting "par excellence" (Schwartz [22, p. 104]) for harmonic analysis. Similar functions exist in the group case and may be defined as follows (Osborne [15]).

First one defines $A(G)$ the space of functions whose L_∞ norm decreases rapidly off powers of a compact set. Specifically $\psi \in A(G)$ if there exists a compact set $C(\psi) \subset G$ such that for each positive integer n there is a constant M_n such that for each integer $k \geqslant 1$

$$\| \psi|_{G - C(\psi)^k} \|_\infty \leqslant M_n k^{-n}. \tag{1.3}$$

One has, for example, $A(G) \subset L_p(G)$, $p \geqslant 1$, $L_\infty(G) \cdot A(G) \subset A(G)$ and $A(G) * A(G) \subset A(G)$. $A(G)$ is translation invariant.

The Schwartz–Bruhat space, $S(G)$, of rapidly decreasing test functions consists of functions $\psi \in A(G)$ with Fourier transforms $\Psi \in A(\Lambda)$. The space has a topology arising from the inequalities defining $A(G)$ and $A(\Lambda)$. One has $S(G) \subset A(G)$, $A(G) * S(G) \subset S(G)$, $S(G) \cdot S(G) \subset S(G)$. $S(G)$ is translation invariant.

A Schwartz–Bruhat tempered distribution is a continuous linear functional on $S(G)$. Such a distribution $x(\psi)$, $\psi \in S(G)$ has a Fourier transform $X(\Psi)$, $\Psi \in S(\Lambda)$, whose (inverse) Fourier transform is $x(\psi)$ in turn.

A final space that will be made use of is $D(G) = S(G) \cap K(G)$, where $K(G)$ denotes the space of continuous functions of compact support on G. One introduces a finer topology on $D(G)$ than that of $S(G)$. (See the references below.) One has $D(G) * D(G) \subset D(G)$ and $D(G) \cdot D(G) \subset D(G)$, and $D(G)$ is dense in $S(G)$. A Schwartz–Bruhat (ordinary) distribution is a continuous linear functional on $D(G)$. A tempered distribution is an (ordinary) distribution which is continuous with respect to the $S(G)$-topology.

The above concepts are discussed in Bruhat [6], Wawrzynczyk [25], Osborne [15], Argabright and Gil de Lampard [1].

2. Stationary Generalized Processes

This paper is concerned with random Schwartz–Bruhat distributions. These are defined as mean–square continuous (in the $D(G)$-topology) random linear functionals on $D(G)$, whose values are equivalence classes of complex-valued random variables of second-order defined on some probability space $(\Omega, \mathbb{Q}, P)$ and which form a Hilbert space $L_2(\Omega)$. (See Ponomarenko [18].) In the stationary case the covariance function of such a process has a spectral representation

$$\operatorname{cov}\{x(\psi_1), x(\psi_2)\} = \int \Psi_1(\lambda)\, \overline{\Psi_2(\lambda)}\, F_2(d\lambda). \tag{2.1}$$

F_2 has a non-negative measure on Λ and the process itself has a spectral representation

$$x(\psi) = \int \Psi(\lambda)\, Z(d\lambda) = X(\Psi) \tag{2.2}$$

with Z a stochastic measure satisfying $\operatorname{cov}\{Z(A), Z(B)\} = F_2(A \cap B)$. The domain of x may be expanded to $\psi \in S(G)$ and to ψ with $\Psi \in L_2(F_2)$. In the case that the process x is real $Z(d\lambda) = \overline{Z(-d\lambda)}$. (See Ponomarenko [18].)

The above spectral representations were developed by Kampé de Fériet [11] for the case of a stationary ordinary process over a group G and by Niemi [14] for the case of a stationary random measure on a group G and by Cramér [7] originally.

A mixing condition will be required in order to develop the central limit theorems of this paper. It is,

Assumption I. x is a real stationary generalized process. For $k = 1, 2,...$ the cumulant of order $k+1$ of $x(\psi)$ is assumed to exist and be given by

$$\int \cdots \int \Psi(\lambda_1) \cdots \Psi(\lambda_k)\, \overline{\Psi(\lambda_1 + \cdots + \lambda_k)}\, f_{k+1}(\lambda_1,..., \lambda_k)\, d\lambda_1 \cdots d\lambda_k \tag{2.3}$$

with

$$\operatorname{vrai\,sup} |\Psi(\lambda_1) \cdots \Psi(\lambda_k)\, \Psi(\lambda_1 + \cdots + \lambda_k)\, f_{k+1}(\lambda_1,..., \lambda_k)| < \infty \tag{2.4}$$

for all $\Psi \in S(\Lambda)$.

The functions f_{k+1} are called the cumulant spectra of the process x. They are given in terms of the stochastic measure Z by

$$\operatorname{cum}\{Z(d\lambda_1),\dots, Z(d\lambda_{k+1})\} \\ = \delta(\lambda_1 + \cdots + \lambda_{k+1}) f_{k+1}(\lambda_1,\dots,\lambda_k)\, d\lambda_1 \cdots d\lambda_{k+1}, \tag{2.5}$$

where δ is the Dirac delta function.

Making an observation on the process x will be viewed as being able to compute $x(\psi)$ for all ψ in $D(G)$ having support within a given compact set T, (the domain of observation). The sample sum will be viewed as the value $x(\psi)$ with ψ near 1 throughout T. The finite Fourier transform will be taken to be

$$x(\langle g, -\omega\rangle \psi) = \int \Psi(\omega - \lambda)\, Z(d\lambda) = X(\Psi(\omega - \cdot)) \tag{2.6}$$

for such a ψ. The function ψ will need to be in $D(G)$ and will be called a taper.

In the case that G is not compact and that T is large the finite Fourier transform (2.5) might be expected to be approximately normally distributed when the process x is mixing. The notion that T is large will be formalized by basing the Fourier transform on a net of tapers $\{\psi_\alpha\}$ such that $\lim \psi_\alpha(g) = 1$ (pointwise) for all $g \in G$. (In the case that G is σ-compact the net may be taken to be a sequence.) It will be seen that nets of tapers exist such that the finite Fourier transform is asymptotically normal.

3. The Asymptotic Normality

The following regularity conditions will be required in connection with the net of tapers.

Assumption II. $\{\Psi_\alpha\}$ is a net of functions in $S(\Lambda)$ with the property that Ψ_α may be written $\Psi_\alpha = \Phi_\alpha \Psi$ with Φ_α, $\Psi \in S(\Lambda)$, $\Psi(0) = 1$

$$\int |\Phi_\alpha|^2 = 1, \qquad \lim \int_{\Lambda/U} |\Phi_\alpha|^2 = 0 \tag{3.1}$$

for U any neighborhood of 0 in Λ and

$$\lim \int |\Phi_\alpha| = 0. \tag{3.2}$$

In Appendix 1 of the paper such a net will be constructed in the case of non-compact G with the further properties that $\psi_\alpha \in D(G)$ and that $\psi_\alpha(g)$ is proportional to a function with limit $= 1$ for $g \in G$. In the case of σ-compact G, the net may be taken to be a sequence.

The principal result of the paper may now be stated.

THEOREM. *Let the stationary generalized process x satisfy Assumption I and have mean 0. Suppose further that the second-order spectrum* $f_2(\lambda)$ *is continuous at the distinct points* $\lambda = \omega_1, \ldots, \omega_J$ *and does not vanish at those points. Let the net (or sequence)* $\{\Psi_\alpha\}$ *satisfy Assumption II. Then the finite Fourier transforms* $X(\Psi_\alpha(\omega_j - \cdot))$, $j = 1, \ldots, J$ *are asymptotically independent normals with zero means and variances* $f_2(\omega_j)$, $j = 1, \ldots, J$, *respectively.*

Proof. The argument proceeds by showing that the cumulants of order greater than 2 tend to 0 and that those of order 2 tend to the values indicated. Because the normal distribution is determined by its cumulants (those of order greater than 2 vanish), and because the cumulants exist and tend to 0 as indicated, the corresponding net of probability measures is tight and asymptotic normality follows. The argument is carried through for the univariate case. The joint normality argument may be reduced to this case. Using abbreviated notation, one has for cumulants of order $k + 1$

$$|\mathrm{cum}\{X(\Psi_\alpha), \ldots, X(\Psi_\alpha)\}|$$

$$= \left| \int \Psi_\alpha(\lambda_1) \cdots \Psi_\alpha(\lambda_k) \overline{\Psi_\alpha(\lambda_1 + \cdots + \lambda_k)} f_{k+1}(\lambda_1, \ldots, \lambda_k) \, d\lambda_1 \cdots d\lambda_k \right|$$

from (2.3)

$$\leqslant M \int |\Phi_\alpha(\lambda_1) \cdots \Phi_\alpha(\lambda_k) \Phi_\alpha(\lambda_1 + \cdots + \lambda_k)| \, d\lambda_1 \cdots d\lambda_k \text{ from (2.4)}$$

$$\leqslant M \left(\int |\Phi_\alpha(\lambda)|^{(k+1)/k} \, d\lambda \right)^k \text{ from Appendix 2.}$$

Now $(\int |\Phi|^{(k+1)/k})^k \leqslant \int |\Phi|^2 (\int |\Phi|)^{k-1}$ by Holder's Inequality and so the cumulant indicated tends to 0 from (3.2) in the case $k > 1$. (Use is also made here of the fact that the space $S(G)$ is translation invariant.)

In the case $k = 1$

$$\mathrm{var}\, X(\Psi_\alpha(\omega - \cdot)) = \int |\Psi_\alpha(\omega - \lambda)|^2 f_2(\lambda) \, d\lambda$$

$$= \int |\Phi_\alpha(\omega - \lambda)|^2 \, |\Psi(\omega - \lambda)|^2 f_2(\lambda) \, d\lambda$$

tending to $f_2(\omega)$ at points of continuity of f_2 in view of (3.1), the approximate identity nature of the net $|\Phi_\alpha|^2$. In the covariance case one has

$$|\operatorname{cov}\{X(\Psi_\alpha(\omega_1-\cdot)), X(\Psi_\alpha(\omega_2-\cdot))\}|$$
$$= \left| \int \Psi_\alpha(\omega_1-\lambda)\, \overline{\Psi_\alpha(\omega_2-\lambda)}\, f_2(\lambda)\, d\lambda \right|$$
$$\leqslant M \int |\Phi_\alpha(\omega_1-\lambda)\, \Phi_\alpha(\omega_2-\lambda)|\, d\lambda$$

from (2.4). This last integral may be seen to tend to 0 for $\omega_1 \neq \omega_2$ by splitting it up into an integral over a neighborhood of ω_1 not containing ω_2 and a remainder and then using Schwarz's Inequality separately on each integral.

This completes the proof of the Theorem.

4. Extensions

The theorem extends to the case of vector-valued processes by means of the same arguments. This extension includes the complex-valued case.

Appendix 1

This Appendix demonstrates the existence of a net of functions, $\Psi_\alpha \in S(\Lambda)$ satisfying Assumption II.

Simon [24] and Hewitt and Ross [10, p. 298] construct nets, $\Gamma_\alpha \in L_1(\Lambda)$, such that: (i) $\Gamma_\alpha \geqslant 0$, $\int \Gamma_\alpha(\lambda)\, d\lambda = 1$; (ii) $0 \leqslant \gamma_\alpha(g) \leqslant 1$, $\gamma_\alpha \in K(G)$; (iii) given U a neighborhood of 0 in Λ and $\beta > 0$, there is an α_0 such that for $\alpha > \alpha_0$ one has $\int_U \Gamma_\alpha(\lambda)\, d\lambda > 1-\beta$ and $\Gamma_\alpha(\lambda) < \beta$ for $\lambda \in \Lambda/U$; (iv) $\lim \gamma_\alpha(g) = 1$ uniformly on compact subsets of G; and (v) the net becomes a sequence for G σ-compact.

Bruhat [6] demonstrates the existence of functions $\theta \in D(G)$ such that $\theta \geqslant 0$ and $\int \theta(g)\, dg = 1$ and one has $|\Theta| \leqslant 1$.

Now set

$$\phi_\alpha = \gamma_\alpha * \theta \Big/ \left(\int (\gamma_\alpha * \theta)^2 \right)^{1/2}, \qquad \Phi_\alpha = \Gamma_\alpha \Theta \Big/ \left(\int (\Gamma_\alpha \Theta)^2 \right)^{1/2}.$$

One sees that $\gamma_\alpha * \theta(g) \geqslant 0$ and tends to $\int \theta(g)\, dg = 1$. This implies that $\int \Gamma_\alpha^2 \Theta^2$ tends to ∞ for non-compact G. Further, from (iii),

$$\int_{\Lambda/U} \Gamma_\alpha^2 \Theta^2 \leqslant \beta(1-\beta)$$

giving (3.1). Next, $\int |\Gamma_\alpha \Theta| \leqslant \int \Gamma_\alpha \leqslant 1$ and (3.2) follows.

The construction is completed by taking $\Psi = \Theta$.

APPENDIX 2

LEMMA. *Let k be a positive integer and Φ in $L_p(\Lambda)$ with $p = (k+1)/k$. Then*

$$\int \cdots \int |\Phi(\lambda_1) \cdots \Phi(\lambda_k)\, \Phi(\lambda_1 + \cdots + \lambda_k)|\, d\lambda_1 \cdots d\lambda_k \leqslant \left(\int |\Phi(\lambda)|^p \, d\lambda \right)^k.$$

Proof. The result follows from Holder's Inequality, namely,

(a) $\|fg\|_1 \leqslant \|f\|_p \|g\|_q$, $1/p + 1/q = 1$, Young's inequality (Segal and Kunze [23, p. 204], namely,

(b) $\|f * g\|_r \leqslant \|f\|_p \|g\|_q$, $1/r = 1/p + 1/q - 1 \geqslant 0$ and Fubini. Define $H_0(\lambda) = \Phi(\lambda)$,

$$H_{j+1}(\lambda) = \int |\Phi(\lambda + \mu)\, H_j(\mu)|\, d\mu$$

and note that if the integral exists it is given by $\int |\Phi(\lambda)\, H_{k-1}(\lambda)|\, d\lambda$. Now,

$$\int |\Phi(\lambda)\, H_{k-1}(\lambda)|\, d\lambda \leqslant \|\Phi\|_p \|H_{k-1}\|_{k+1}$$

$$\|H_{k-1}\|_{k+1} \leqslant \|\Phi\|_p \|H_{k-2}\|_{(k+1)/2} \leqslant \cdots \leqslant \|\Phi\|_p^{k-1} \|H_0\|_{(k+1)/k}$$
$$= \|\Phi\|_p^k.$$

REFERENCES

[1] ARGABRIGHT, L., AND GIL DE LAMADRID, J. (1974). Fourier analysis of unbounded measures on locally compact abelian groups. *Mem. Amer. Math. Soc.* **145**.

[2] BRILLINGER, D. R. (1970). The frequency analysis of relations between stationary spatial series. In *Proceedings, Twelfth Biennial Seminar Canadian Mathematics Congress* (R. Pyke, Ed.) pp. 39–81. Canad. Math. Congress, Montreal.

[3] BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. In *Proceedings, Sixth Berkeley Symposium Math. Statist. Prob.* (L. M. Le Cam, J. Neyman, and E. L. Scott, Eds.), pp. 483–513. Univ. Calif. Press, Berkeley.

[4] BRILLINGER, D. R. (1974). The Fourier analysis of stationary processes. *Proc. Inst. Electr. Engrs.* **62**, 1628–1643.

[5] BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart & Winston, New York.

[6] BRUHAT, F. (1961). Distributions sur un groupe localement compact et applications à l'étude des représentations des groupes p-adiques. *Bull. Soc. Math. France* **89**, 43–75.

[7] CRAMÉR, H. (1942). On harmonic analysis in certain functional spaces. *Arkiv Math. Astron. Fys.* **28**, 1–7.

[8] HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

[9] HANNAN, E. J., AND THOMSON, P. J. (1971). Spectral inference over narrow bands. *J. Appl. Probab.* **8**, 157–169.

[10] HEWITT, E., AND ROSS, K. A. (1970). *Abstract Harmonic Analysis, II*. Springer-Verlag, Berlin.

[11] KAMPÉ DE FÉRIET, J. (1948). Analyse harmonique des fonctions aléatoires stationaires d'ordre 2 sur un group abélien compact. *C.R. Acad. Sci. Paris* **226**, 868–870.

[12] LEONOV, V. P. AND SHIRYAEV, A. N. (1960). Some problems in the spectral theory of higher-order moments, II. *Theory Probab. Appl.* **4**, 319–329.

[13] MORETTIN, P. A. (1980). Homogeneous random processes on locally compact abelian groups. *Anais Acad. Brasil. Ciênc.* **52**, 1–6.

[14] NIEMI, H. (1978). Stationary vector measures and positive definite translation invariant bimeasures. *Ann. Acad. Sci. Fenn. Ser. A* **4**, 209–226.

[15] OSBORNE, M. S. (1975). On the Schwartz–Bruhat space and the Paley–Wiener theorem for locally compact abelian groups. *J. Funct. Anal.* **19**, 40–49.

[16] PICINBONO, B. (1960). Tendance vers le caractère gaussien par filtrage sélectif. *C.R. Acad. Sci. Paris* **250**, 1174–1176.

[17] PICHARD, J.-F. (1974). Normalisation par filtrage selectif de fonctions aléatoires de plusieurs variables réeles de classe $\Delta^{(\infty)}$. *C.R. Acad. Sci. Paris* **278**, 443–446.

[18] PONOMARENKO, A. I. (1974). Harmonic analysis of generalized wide-sense homogeneous random fields on a locally compact commutative group. *Theoret. Prob. Math. Statist.* **4**, 119–137.

[19] ROSENBLATT, M. (1961). Some comments on narrow band-pass filters. *Quart. Appl. Math.* 18, 387–394.

[20] ROSENBLATT, M. (1981). Limit theorems for Fourier transforms of functionals of Gaussian sequences. *Z. Wahrsch. Verw. Gebiete* **55**, 123–132.

[21] ROZANOV, YU. A. (1961). An application of the central limit theorem. In *Proceedings, Fourth Berkeley Symposium Math. Statist. Prob.* (J. Neyman, Ed.) Vol. 1, pp. 445–454.

[22] SCHWARTZ, L. (1959). *Théorie des distributions, II*. Hermann, Paris.

[23] SEGAL, I. E., AND KUNZE, R. A. (1978). *Integrals and Operators*. McGraw-Hill, New York.

[24] SIMON, A. B. (1966). Cesàro summability of groups: characterization and inversion of Fourier transforms. In *Function Algebras*, pp. 208–215. Scott, Foresman, New York.

[25] WAWRZYNCZYK, A. (1968). On tempered distributions and Bochner–Schwartz theorem on arbitrary locally compact abelian groups. *Colloq. Math.* **19**, 305–318.

Journal of Theoretical Probability, Vol. 10, No. 2, 1997

A Particle Migrating Randomly on a Sphere

David R. Brillinger[1]

Consider a particle moving on the surface of the unit sphere in R^3 and heading towards a specific destination with a constant average speed, but subject to random deviations. The motion is modeled as a diffusion with drift restricted to the surface of the sphere. Expressions are set down for various characteristics of the process including expected travel time to a cap, the limiting distribution, the likelihood ratio and some estimates for parameters appearing in the model.

KEY WORDS: Drift; great circle path; likelihood ratio; pole-seeking; skew product; spherical Brownian motion; stochastic differential equation; travel time.

1. INTRODUCTION

There are marine mammals, such as elephant seals, that travel great distances and are tracked. It is of interest to biologists to describe the routes. One can wonder for example if the animals follow great circle paths. The animals will be foraging along the way, i.e., pulled away from the direct route from origin to destination, and this may be modeled as stochastic fluctuations. The great circle route is the geodesic, providing the shortest trip. A ship needs to be changing course continually to stay on it. It is intriguing that some animals apparently do not need to change course, they can keep going straight ahead.

An issue that arises in modeling the physical world is whether to work employing the Itô or the Stratonovich calculus. Reasons have been presented in various places to the effect that, when developing physical applications, it is simpler to start with the Stratonovich form and then switch to the Itô for developing properties of the process. See the discussions in: Bartholdi *et al.*[19] Karlin and Taylor.[9]

To start, some of the previous work on the planar case with drift and the spherical case without drift will be presented.

[1] Statistics Department, University of California, Berkeley, California 94720-3860.

0894-9840/97/0400-0429$12.50/0 © 1997 Plenum Publishing Corporation

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_7, © Springer Science+Business Media, LLC 2012

2. THE PLANAR CASE

Kendall[10] considers the case of a Brownian motion on the plane with an "attractive" polar drift. He works with polar coordinates, (r, ϕ), centered at the target center. The particle, in his case a bird, starts at location $(D, 0)$. In a time interval of length dt it moves a distance $\delta\, dt$ towards the target, then is subject to random Gaussian disturbance, of amount $\sigma\, dU_t$ towards the target and amount $\sigma\, dV_t$ at right angles to the path. Here U_t and V_t are independent standard Brownian motions and σ^2 their common variance. In Itô form the motion may be described by

$$dr_t = \left(\frac{\sigma^2}{2r_t} - \delta\right) dt + \sigma\, dU_t$$
$$d\phi_t = \frac{\sigma}{r_t}\, dV_t \tag{2.1}$$

It will be noted later that these equations correspond approximately to motion on a sphere of large radius. The infinitesimal generator of the r_t process is

$$\frac{\sigma^2}{2}\frac{d^2}{dr^2} + \left(\frac{\sigma^2}{2r} - \delta\right)\frac{d}{dr}$$

Using the criteria developed in Karlin and Taylor[9] or Bhattacharya and Waymire,[3] for this process the point 0 is unreachable, but an entrance point.

Next suppose that there is a circle of radius a about the target, then among the results of the Kendall paper is that the time to get from $(D, 0)$ to the circle has expected value

$$\left(D - a + \frac{\sigma^2}{2\delta} \log D/a\right) \Big/ \delta \tag{2.2}$$

This result may be obtained directly from the formulas recorded in Appendix I, with

$$s(x) = \int_a^x \frac{1}{z} \exp\left\{z \frac{2\delta}{\sigma^2}\right\} dz \tag{2.3}$$

and

$$m'(x) = x \frac{2}{\sigma^2} \exp\left\{-x \frac{2\delta}{\sigma^2}\right\}$$

$x > 0$. Here (2.3) is the exponential integral $Ei(\cdot)$. The naive expression for the expected travel time is $(D-a)/\delta$ corresponding to $\sigma = 0$ in (2.2). Kendall also derives an expression for the variance of the travel time.

The invariant distribution of the process is proportional to $m'(x)$ above, i.e., a gamma. The likelihood ratio relative to the case $\delta = 0$ is

$$\int_0^T \frac{1}{\sigma^2}(-\delta)\, dr_s - \frac{1}{2}\int_0^T \frac{1}{\sigma^2}\left(\delta^2 - \frac{2\delta\sigma^2}{r_s}\right) ds$$

following the Cameron–Martin–Girsanov formula recorded in Appendix I. This is maximized by the choice

$$\hat{\delta} = \frac{1}{T}\left(r_0 - r_T + \sigma^2 \int_0^T \frac{1}{r_s}\, ds\right)$$

When $\delta = 0$ the equations considered are the polar coordinate form of Brownian motion in the plane and for r_t one has the case $n = 2$ of the Bessel process discussed for example in Karlin and Taylor.[9] The transition density function for that process is given there and is

$$p(t; q; r) = g(r) \int_0^\infty e^{-\lambda^2 t/2} G(\lambda q)\, G(\lambda r)\, g(\lambda)\, d\lambda$$

$t > 0$ with

$$g(r) = r, \qquad G(r) = \Gamma(\tfrac{3}{2})\, J_0(r)$$

3. FORMS OF SPHERICAL BROWNIAN MOTION

Perrin[15] working from a model of a randomly rotating sphere, determines the density of the angle, ω_t, subtended at the center of the sphere between the initial position of a point on the surface and its position t time units later. He finds that density, relative to the measure $\sin \omega\, d\omega\, d\phi$ on the sphere is

$$2\pi f(\omega, t)$$

with f satisfying

$$\frac{\sigma^2}{2}\frac{\partial^2 f}{\partial \omega^2} + \frac{\sigma^2}{2}\frac{1}{\tan \omega}\frac{\partial f}{\partial \omega} = \frac{\partial f}{\partial t} \tag{3.1}$$

Perrin shows the solution of (3.1), with appropriate initial conditions, is

$$f(\omega, t) = \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1) P_n(\cos \omega) e^{-n(n+1) t\sigma^2/2} \tag{3.2}$$

for $t > 0$ where P_n is the Legendre polynomial of order n. Perrin provides some graphs of this function. He also shows that

$$E \sin^2 \omega_t = \tfrac{2}{3}(1 - e^{-t3\sigma^2})$$

and further remarks that one can get other moments of Legendre polynomials in $\cos \omega_t$ by integrating them against expression (3.2).

Yosida[23] determines spherical Brownian motion as the unique temporally and spatially homogeneous diffusion process on S^2. Suppose θ and ϕ denote longitude and colatitude respectively, $0 \leqslant \theta \leqslant \pi$ and $0 \leqslant \phi < 2\pi$. Yosida finds the transition density from the position (θ, ϕ) at time 0 to the position (θ', ϕ') at time t to be

$$p(t; \theta, \phi; \theta', \phi') = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} e^{-k(k+1)t} Y_k^m(\theta, \phi)\, Y_k^m(\theta', \phi') \tag{3.3}$$

Here $Y_n^m(\theta, \phi)$ is the spherical harmonic

$$Y_n^m(\theta, \phi) = e^{im\phi} P_n^m(\cos \theta)$$

for $n = 0, 1, 2,..., |m| \leqslant n$, and P_n^m is the associated Legendre function and $P_n^0 = P_n$, see Terras.[20] The representations (3.2) and (3.3) are seen to correspond when one uses the addition formula for spherical harmonics. Yosida[23] gives the infinitesimal generator of the process as $\sigma^2 \Delta^*/2$ where Δ^* is the spherical Laplacian

$$\Delta^* = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \tag{3.4}$$

$0 < \theta < \pi$. The Y_k^m are the eigenfunctions of the operator Δ^*, showing one source of the formula (3.3).

The Itô equations for the process are

$$d\theta_t = \sigma\, dU_t + \frac{\sigma^2}{2 \tan \theta_t} dt \tag{3.5}$$

$$d\phi_t = \frac{\sigma}{\sin \theta_t} dV_t \tag{3.6}$$

with U_t and V_t independent Brownians.

The motion of θ is what Itô and McKean[8] call the Legendre process on $[0, \pi]$. It has generator

$$\frac{\sigma^2}{2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)$$

Writing

$$l(t) = \int_0^t \frac{1}{\sin^2\theta(s)}\,ds$$

and with ψ circular Brownian (that is σB_t $(mod\, 2\pi)$, where B_t is a Brownian), they show that spherical Brownian may be represented as

$$[\theta, \phi] = [\theta, \psi(l)] \tag{3.7}$$

Stroock[18] works with $\mathbf{Y}$ in R^3. He obtains the spherical Brownian as the solution of the Itô equation

$$d\mathbf{Y} = (\mathbf{I} - \mathbf{Y}\mathbf{Y}'/(\mathbf{Y}'\mathbf{Y}))\,d\mathbf{B} - (\mathbf{Y}/(\mathbf{Y}'\mathbf{Y}))\,dt \tag{3.8}$$

supposing $\mathbf{Y} \neq 0$ and that $\mathbf{B}$ is Brownian $R \to R^3$. In this case the Stratonovich form is

$$(S)\; d\mathbf{Y} = (\mathbf{I} - \mathbf{Y}\mathbf{Y}'/(\mathbf{Y}'\mathbf{Y}))\,d\mathbf{B} \tag{3.9}$$

(Here and in the following (S) indicates that the Stratonovich form of equation is being employed.) This process stays on the surface and the final term of (3.8) may be thought of as pulling the process back onto the sphere.

With the change to polar coordinates defined by $\mathbf{Y} = (R\sin\theta\sin\phi,$ $R\sin\theta\cos\phi,\ R\cos\theta)$ expression (3.9) becomes

$$(S)\; dR_t = 0$$

$$(S)\; d\theta_t = \frac{\sigma}{R_t}\,[\cos\theta_t \sin\phi_t\, dB_t^1 + \cos\theta_t \cos\phi_t\, dB_t^2 - \sin\theta_t\, dB_t^3]$$

$$(S)\; d\phi_t = \frac{\sigma}{R_t\sin\theta_t}\,[\cos\phi_t\, dB_t^1 - \sin\phi_t\, dB_t^2]$$

If one sets $\theta_t = r_t/R$ and takes R large, these become

$$(S)\; dr_t \approx \sigma[\sin\phi_t\, dB_t^1 + \cos\phi_t\, dB_t^2]$$

$$(S)\; d\phi_t \approx \frac{\sigma}{r_t}\,[\cos\phi_t\, dB_t^1 - \sin\phi_t\, dB_t^2]$$

If one converts to the Itô form and $\delta = 0$, these become (2.1). In another approach, Oksendal,[14] [pp. 142–143] obtains spherical Brownian as $\mathbf{B}/|\mathbf{B}|$ with a particular time change.

Suppose next that one is focusing on the distance, θ_t, to the North Pole. With the change of variables, $X_t = \cos\theta_t$, (3.5) becomes

$$dX_t = -\sigma^2 X_t\, dt - \sigma\sqrt{1 - X_t^2}\, dU_t \tag{3.10}$$

on $[-1, 1]$. This process is considered in Karlin and Taylor,[9] and Matthews.[12] Its infinitesimal generator is

$$A = \frac{\sigma^2}{2}(1 - x^2)\frac{d^2}{dx^2} - \sigma^2 x\frac{d}{dx}$$

The eigenfunctions are the Legendre polynomials, $P_n(x)$ with eigenvalue $\lambda_n = n(n+1)$, see Karlin and Taylor.[9] One sees another connection with (3.2). The so-called scale and speed functions, defined in Appendix I, are

$$s(x) = \frac{1}{2}\log\frac{1+x}{1-x}$$

and

$$m(x) = \frac{2}{\sigma^2}(x - x_0)$$

respectively. The invariant density, proportional to $m'(x)$, is the uniform. The points ± 1 constitute an entrance boundary and are unreachable. Using the expression (A.3) of the Appendix and carrying out the required integrations, the expected time to travel from x to the point d is

$$\frac{2}{\sigma^2}\log\frac{1-x}{1-d}$$

$-1 < x < d < 1$.

Roberts and Ursell[16] investigate random walks on the sphere with all directions of movement assumed equally probable. They obtain the formula (3.2) as the limit when the step size gets small and suggest an approximation to distribution of ω_t. Hartman and Watson[7] develop various properties of the approximate distribution. [See also Bingham,[4] Watson.[22]]

4. THE GREAT CIRCLE CASE

Suppose that a particle on the sphere is migrating directly towards the North Pole at speed δ and subject to Brownian disturbances. The North

Pole is taken for convenience. The Itô differential equations for the process are

$$d\theta_t = \sigma\, dU_t + \left(\frac{\sigma^2}{2\tan\theta_t} - \delta\right) dt$$
$$d\phi_t = \frac{\sigma}{\sin\theta_t}\, dV_t \qquad (4.1)$$

so long as $\theta_t \neq 0$ and with ϕ_t defined $mod\, 2\pi$. It will be supposed that the particle does not start at $\theta = 0$ or π. (These points are inaccessible.) The equations extend (3.4).

The infinitesimal generator of the process is

$$\left(\frac{\sigma^2}{2\tan\theta} - \delta\right)\frac{\partial}{\partial\theta} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial\theta^2} + \frac{1}{2}\frac{\sigma^2}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \qquad (4.2)$$

Clearly the process is bounded, simplifying derivations.

Figure 1 shows a simulation corresponding to the application motivating this research. It refers to elephant seals migrating from the California coast into the NW Pacific. One notices the particle meandering once it reaches the neighborhood of its destination, as was to be anticipated. Meandering around the destination may be thought of as the animal foraging there.

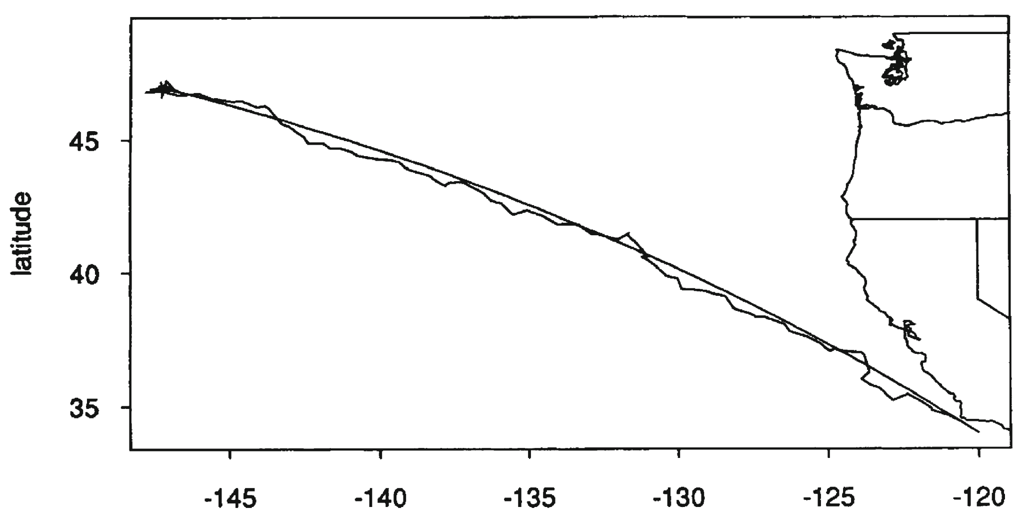


Fig. 1. Simulation of diffusion with drift on sphere.

With the change of variables $X_t = \cos\theta_t$, and using Itô's lemma, Eq. (4.1) becomes

$$dX_t = (-\sigma^2 X_t + \delta\sqrt{1-X_t^2})\,dt - \sigma\sqrt{1-X_t^2}\,dU_t$$

whose infinitesimal generator is

$$B = \frac{\sigma^2}{2}(1-x^2)\frac{d^2}{dx^2} - (\sigma^2 x - \delta\sqrt{1-x^2})\frac{d}{dx}$$

on $[-1, 1]$. This reduces to (3.10) when $\delta = 0$.

In the case of a sphere of large radius R, if one writes $\theta_t = r_t/R$ one sees that the Eq. (4.1) become

$$\frac{1}{R}dr_t \approx \sigma\,dU_t + \left(\frac{\sigma^2}{2}\frac{R}{r_t} - \delta\right)dt$$

$$d\phi_t \approx \sigma\frac{R}{r_t}dV_t$$

and replacing $R\sigma$ by σ and $R\delta$ by δ leads back to (2.1). The endpoints of the interval here are inaccessible, but following the general discussion of the topic in Karlin and Taylor[9] can be treated as points of entry. The process is recurrent.

Various characteristics may be derived from the expressions in the Appendix. Following that material one has

$$I(x, z) = \frac{2\delta}{\sigma^2}[\cos^{-1}x - \cos^{-1}z] - [\log(1-x^2) - \log(1-z^2)]$$

The scale function is given by

$$s(x) = \int_{x_0}^{x} \exp\left\{\frac{2\delta}{\sigma^2}\cos^{-1}z\right\}\frac{1}{1-z^2}\,dz$$

and the speed function by

$$m(x) = \frac{1}{\sigma^2}\int_{x_0}^{x} \exp\left\{-\frac{2\delta}{\sigma^2}\cos^{-1}z\right\}dz$$

The invariant density of the process is proportional to

$$m'(x) = \exp\left\{-\frac{2\delta}{\sigma^2}\cos^{-1} x\right\}$$

$-1 < x < 1$. For δ/σ^2 small, the density is approximately uniform.

Consider next the expected travel time for the process. Suppose the process starts at x and heads to d, $1 > d > x > -1$. Following the expression (A.4) in the Appendix this may be evaluated to

$$\int_x^d \frac{2}{\sigma^2}\left[\int_{-1}^{y} \exp\left\{-\frac{2\delta}{\sigma^2}\cos^{-1} z\right\} dz\right] \exp\left\{\frac{2\delta}{\sigma^2}\cos^{-1} y\right\} \frac{1}{1-y^2}\, dy$$

In the case that $\delta = 0$, it is

$$\frac{2}{\sigma^2}\log\frac{1-x}{1-d}$$

as given in Section 3.

The skew product representation (3.7) given earlier holds in the present case as well. The proof of Itô and McKean,[8] [p. 200], applies equally.

5. DETERMINATION OF THE PARAMETERS

Following the expression (A.5) in Appendix I, the likelihood ratio of the process, relative to that of the case $\delta = 0$, is

$$\frac{1}{\sigma^2}\left[(-\delta)\int_0^T d\theta_s - \frac{1}{2}\int_0^T \left(-\frac{2\delta\sigma^2}{\tan\theta_s} + \delta^2\right) ds\right]$$

This leads to the maximum likelihood estimate

$$\hat{\delta} = \frac{1}{T}\left[(\theta_0 - \theta_T) + \sigma^2 \int_0^T \frac{1}{\tan\theta_s}\, ds\right] \Big/ T$$

Because the particle reaches the region of its destination eventually, this estimate becomes unreasonable as $T \to \infty$.

One can obtain an exact estimate of σ^2 on the basis of the usual sort of result for quadratic variation

$$\sum_i [\tilde{\phi}_{t_{i+1}} - \tilde{\phi}_{t_i}]^2 \xrightarrow{P} \sigma^2 \int_0^T \frac{1}{\sin^2\theta_s}\, ds \qquad (5.1)$$

derived in the Appendix. Here $\{t_i\}$ is a partition of the interval that gets finer under the limiting process, the result is conditional on the (continuous) realization of θ_s, $0 \leqslant s \leqslant T$, and it is assumed that there exists $\varepsilon > 0$ such that $|\sin \theta_s| \geqslant \varepsilon$. The curve $\tilde{\phi}_t$ refers to a continuous curve obtained from the curve ϕ_t by either patching together continuous segments or by reflecting ϕ_t whenever it reaches the barriers $\phi = 0, \pi$. (It is assumed that $0 < \phi_0 < 2\pi$.) These two constructions are illustrated in Fig. 2. The top

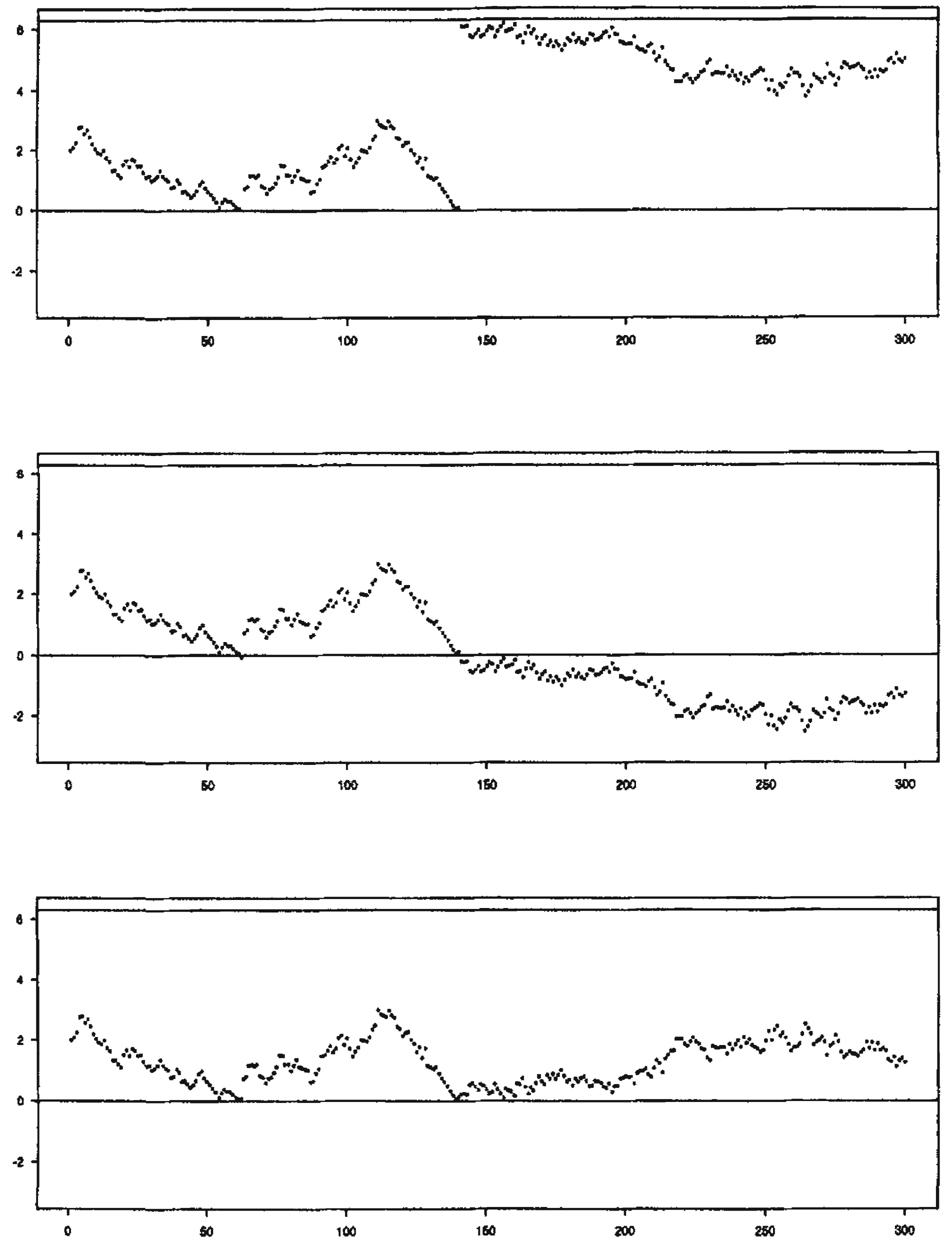


Fig. 2.

graph represents a realization of ϕ_t. The middle results from joining the two segments continuously and the bottom from reflecting.

When working in practice the data will be available at discrete time points and the above likelihood ratio is not obtainable. However if one has an expression for the transition density, then it can be employed to obtain the likelihood function and estimates of the parameters obtained. An approximate approach here is to do what a ship's navigator has done traditionally. Specifically at the start of a day based on a ship's position the navigator determines the great circle course and that is followed for a day. The next day the navigator determines the ship's new position, again the great circle course based on that position is determined and followed for a day. Unless the ship is heading due north or south, during the course of the days it will be pulled off the great circle route, but with the course revisions the destination is approached. This method leads to approximating the desired transition density by a succession of motions in the plane.

Other approaches to approximating the transition density include: numerical solution of the partial differential equations, some form of quadrature (e.g., that of Dawson[5]) and simulation.

6. NAVIGATOR'S COORDINATES

To obtain the planar approximation just referred to and to prepare Fig. 1 provided earlier, traditional coordinates are helpful. The relations for these are as follows. Let ϕ and $\bar{\theta}$ refer to longitude and latitude on the sphere, in radians, $0 \leqslant \phi < 2\pi$ and $-\pi/2 \leqslant \bar{\theta} \leqslant \pi/2$. Suppose that a ship is at location ϕ_t, $\bar{\theta}_t$ at time t heading towards position Φ, $\bar{\Theta}$. The great circle distance, ρ_t, and course, η_t, satisfy

$$\cos \rho_t = \cos\left(\frac{\pi}{2} - \bar{\theta}_t\right) \cos\left(\frac{\pi}{2} - \bar{\Theta}\right)$$

$$+ \sin\left(\frac{\pi}{2} - \bar{\theta}_t\right) \sin\left(\frac{\pi}{2} - \bar{\Theta}\right) \cos(\Phi - \phi_t)$$

$$\cos \eta_t = \left[\cos\left(\frac{\pi}{2} - \bar{\Theta}\right) - \cos\left(\frac{\pi}{2} - \bar{\theta}_t\right) \cos \rho_t\right] \Big/ \left[\sin\left(\frac{\pi}{2} - \bar{\theta}_t\right) \sin \rho_t\right]$$

with appropriate choice of quadrant in the latter case. These formulas come from spherical trigonometry and are developed, for example, in Various.[21]

7. DISCUSSION

Corresponding Itô, Stratonovich and differential equation developments have been presented. Each has something to offer and will be employed in the practical study of elephant seal paths in progress.

Another paper working with a diffusion on the sphere, with drift, is Le Gall and Yor.[11] They add general drift terms to Brownian on the sphere and study the equations

$$d\theta_t = dU_t + \frac{1}{2 \tan \theta_t} dt + c_1(\theta_t, \phi_t) \, dt$$

$$d\phi_t = \frac{1}{\sin \theta_t} (dV_t + c_2(\theta_t, \phi_t) \, dt)$$

They obtain asymptotic properties of the windings of the process.

Rogers and Williams[17] develop Brownian motion on submanifolds of R^d via a Stratonovich equation. This could be extended to include motion preferring one direction.

APPENDIX

Appendix I—Some Formulas

General results for diffusion processes are developed in Gihman and Skorokhod,[6] Karlin and Taylor,[9] and Bhatacharya and Waymire[3] for example. The notation of the latter work is used here.

Consider a diffusion process X_t on the line satisfying

$$dX_t = \mu(X_t) \, dt + \sigma(X_t) \, dB_t \tag{A.1}$$

where B_t is Brownian motion. Define

$$I(x, z) = \int_x^z \frac{2\mu(y)}{\sigma^2(y)} \, dy$$

Then the scale function is given by

$$s(x) = \int_{x_0}^x \exp\{-I(x_0, z)\} \, dz \tag{A.2}$$

and the speed function by

$$m(x) = \int_{x_0}^{x} \frac{2}{\sigma^2(z)} \exp\{I(x_0 z)\}\, dz \qquad \text{(A.3)}$$

Following Bhatacharya and Waymire[3] the expected travel time may be written

$$\int_{x}^{d} (m(y) - m(-1))\, ds(y) \qquad \text{(A.4)}$$

if -1 is the lower bound of the state space.

The Cameron-Martin-Girsanov formula for the log likelihood ratio of the process (A.1), relative to the case of $\mu(x) = \mu_0(x)$, leads to the expression

$$\int_0^T \frac{[\mu(X_t) - \mu_0(X_t)]}{\sigma^2(X_t)}\, dX_t - \frac{1}{2}\int_0^T \frac{[\mu^2(X_t) - \mu_0^2(X_t)]}{\sigma^2(X_t)}\, dt \qquad \text{(A.5)}$$

See also Gihman and Skorokhod[6] [p. 90].

Appendix II—Determining σ

One reference is Basawa and Rao.[2] Consider the quantity

$$\sum_i [\tilde{\phi}_{t_{i+1}} - \tilde{\phi}_{t_i}]^2 = \sigma^2 \sum_i \left[\int_{t_i}^{t_{i+1}} \frac{1}{\sin\theta_s}\, d\tilde{V}_s \right]^2 \qquad \text{(A.6)}$$

Its behavior will be considered conditional on the continuous realization $\{\theta_t,\ 0 \leqslant t \leqslant T\}$ satisfying $|\sin\theta_t| \geqslant \varepsilon$ for some $\varepsilon > 0$.

The expected value of (A.6) is

$$\sigma^2 \int_0^T \frac{1}{\sin^2\theta_s}\, ds$$

which is $\leqslant \sigma^2 T/\varepsilon^2$. The variance of (A.6) is bounded by $2\sigma^4 T \max\{t_{i+1} - t_i\}/\varepsilon^4$ which tends to 0 as the partition gets finer. This gives the result. The conclusion is basically a result for the quadratic variation of a martingale.

ACKNOWLEDGMENTS

It was a privilege to get to work with Murray Rosenblatt early in my career. He set standards of the highest order. I thank him for all the help he has given.

I have had many enlightening conversations with Brent Stewart concerning the elephant seal data collection program and the behavior of the seals. I thank Ed Ionides, Jim Pitman, and Dan Stoock for useful remarks and I thank Geoff Watson for pointing out the Roberts and Ursell[16] reference.

This paper was prepared with the partial support of the National Science Foundation Grant DMS-9625774 and of the Office of Naval Research Grant N00014-94-1-0042.

REFERENCES

1. Bartholdi, E., Wokaun, A., and Ernst, R. R. (1976). Theory of stochastic NMR spectroscopy, application of the Itô and Stratonovich calculus. *Chem. Phys.* **18**, 57–84.
2. Basawa, I. V., and Rao, B. L. S. P. (1980). *Statistical Inference for Stochastic Processes.* Academic, New York.
3. Bhattacharya, R. N., and Waymire, E. (1990). *Stochastic Processes with Applications.* John Wiley, New York.
4. Bingham, N. H. (1972). Random walk on spheres. *Z. Wahrsch. verw. Geb.* **22**, 169–172.
5. Dawson, D. A. (1981). Galerkin approximation of nonlinear Markov processes. *Statistics and Related Topics.* North-Holland, Amsterdam.
6. Gihman, I. I., and Skorohod, A. V. (1972). *Stochastic Differential Equations.* Springer, New York.
7. Hartman, P., and Watson, G. S. (1974). "Normal" distribution functions on spheres and the modified Bessel functions. *Ann. Prob.* **2**, 593–607.
8. Itô, K., and McKean, H. P. (1965). *Diffusion Processes and Their Sample Paths.* Springer, Berlin.
9. Karlin, S., and Taylor, H. M. (1981). *A Second Course in Stochastic Processes.* Academic, New York.
10. Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *J. Roy. Statist. Soc.* **36B**, 365–417.
11. Le Gall, J. F., and Yor, M. (1986). Etude asymptotique de certains mouvements browniens complexes avec drift. *Prob. Th. Rel. Fields* **71**, 183–229.
12. Matthews, P. (1988). Covering problems for Brownian motion on spheres. *Ann. Prob.* **16**, 189–199.
13. Mortensen, R. E. (1969). Mathematical problems of modelling stochastic nonlinear dynamic systems. *J. Stat. Physics* **1**, 271–296.
14. Oksendal, B. (1995). *Stochastic Differential Equations,* Fourth Edition. Springer, New York.
15. Perrin, M. F. (1928). Mouvement brownien de rotation. *Ann. l'École Norm. Sup.* **45**, 1–51.
16. Roberts, P. H., and Ursell, H. D. (1960). Random walk on a sphere and on a Riemanian manifold. *Philos. Trans. Roy. Soc. London* **252A**, 317–356.
17. Rogers, L. C. G., and Williams, D. (1995). *Diffusions, Markov Processes and Martingales,* Vol. 3. Wiley, New York.
18. Stroock, D. W. (1971). On the growth of stochastic integrals. *Z. Wahrsch. verw. Geb.* **18**, 340–344.
19. Sussmann, H. J. (1978). On the gap between deterministic and stochastic ordinary differential equations. *Ann. Prob.* **6**, 19–41.

20. Terras, A. (1985). *Harmonic Analysis on Symmetric Spaces and Applications*. Springer, New York.
21. Various (1938). *Admiralty Navigation Manual*, Vol. III. H. M. Stationery Office, London.
22. Watson, G. S. (1983). *Statistics on Spheres*. Wiley, New York.
23. Yosida, K. (1949). Brownian motion on the surface of the 3-sphere. *Ann. Math. Statist.* **20**, 292–296.

The Annals of Statistics
1988, Vol. 16, No. 1, 1–54

THE 1983 WALD MEMORIAL LECTURES

SOME STATISTICAL METHODS FOR RANDOM PROCESS DATA FROM SEISMOLOGY AND NEUROPHYSIOLOGY[1]

BY DAVID R. BRILLINGER

University of California, Berkeley

To Jeff Austin Brillinger, B.A.

Examples are presented of statistical techniques for the analysis of random process data and of their uses in the substantive fields of seismology and neurophysiology. The problems addressed include frequency estimation for decaying cosinusoids, signal estimation, association measurement, causal connection assessment, estimation of speed and direction and structural modeling. The techniques employed include complex demodulation, nonlinear regression, probit analysis, deconvolution, maximum likelihood, singular value decomposition, Fourier analysis and averaging.

Table of contents

Received December 1986; revised July 1987.
[1]Prepared with the partial support of National Science Foundation Grants CEE-79-01642, MCS-83-16634 and while the author was a Guggenheim Fellow.
AMS 1980 *subject classifications*. Primary 62M99, 62P99.
Key words and phrases. Array data, average evoked response, binary data, Fourier inference, complex demodulation, neurophysiology, point processes, probit analysis, spatial–temporal processes, seismology, system identification, time series.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_8, © Springer Science+Business Media, LLC 2012

I. Introduction

The purpose of statistics, . . . , is to describe certain real phenomena.
A. Wald (1952)

The concern of these lectures is raw data distributed in time and/or space. The basic data are curves and surfaces. If n denotes the sample size and p denotes the dimension, then the concern is with the case of n much less than p. In the situations addressed, the phenomena have developed or are developing in time or space. They are complex, so that subject matter plays essential roles in the analyses made and in the interpretations and conclusions drawn. There need to be combinations of both physical and statistical reasoning. Indeed, a principal goal of the lectures is to bring out the key role that subject matter plays in the analysis of random process data. A further intention is to show that the fields of seismology and neurophysiology are rich in problems for statisticians, particularly those individuals with some interest in applied mathematics. The work presented involves a mixture of data analysis and structural modeling. The problems discussed are specific, but the techniques employed are broadly applicable. The data concerned is of high quality, so that detailed analyses are possible. The material presented consists of personal (collaborative) work and a few success stories of other particular methods that serve to tie the development together. An attempt is made to present problems from a unified point of view. Emphasis is on techniques, rather than novel substantive results.

The study of random process data provides a major interface of statistics with science and technology. Indeed, there has been an explosion in the collection of spatial–temporal measurements (corresponding in part to much of modern technology having become digital). Some particular issues and procedures become emphasized as a result of the interaction of statistics with technology. These include system identification, systems analysis, inverse problems, Fourier inference, bias versus variability (resolution versus precision), averaging functions, dynamics and micro-versus-macro study. These strains run through the examples presented. There is also a desire to display the broad range of data types with whose analysis statisticians must now be concerned.

Some provisos are necessary. There is no claim made that the analyses are definitive. What is presented is an overview, rather than specific details. Furthermore, there is little presentation of formalism. The reader is referred to the papers referenced for greater detail.

There are two lectures. The first concentrates on some statistical methods in seismology, the second on some corresponding methods in neurophysiology. It is interesting to see the same methods playing central roles in the analysis of data from two quite disparate fields. Indeed, one of the principal goals of the lectures was to bring out the universality of statistical techniques—by examples from these two fields.

II. Seismology

Jeffreys... attention to scientific method and statistical detail has been one of the main forces through which Seismology has attained its present level of precision.

Bullen and Bolt (1985)

1. The field and its goals. The term seismology refers to the scientific investigation of earthquakes and related phenomena. It has been defined as the "science based on data called seismograms, which are records of mechanical vibrations of the Earth" [Aki and Richards (1980)]. This latter definition allows the admission that seismologists also study vibrations caused by the sea, by volcanoes or by man. One further definition that has been given is: the science of strain–wave propagation in the Earth.

Whatever the definition, the broad goals of seismology are to learn the Earth's and a planet's interior composition and to predict the time, size, location and strength of ground motion in future earthquakes. Workers in the field seek to provide valid explanations of earthquake-related phenomena and to understand these phenomena so that life may be made safer.

Specific problems addressed include the detection, location and quantification of earthquakes, the distinguishing of earthquakes from nuclear explosions and the determination of wave velocity in the Earth's interior as a function of depth.

The accumulation of knowledge in seismology has displayed a steady back-and-forth between new insight concerning the waves and new insight concerning the media through which the waves propagate. Among major "discoveries" one can list are the inner core, the liquid central core, the Mohorovic discontinuity, the movement of tectonic plates causing earthquakes themselves and the locating of numerous gas and oil fields.

The field is largely observational, with the basic instruments the seismogram and clock. There are important experiments too, where tailored impulses are input to the Earth and the resulting vibrations studied. The field experienced the "digital revolution" in the 1950s and now poses problems exceeding the capabilities of even today's supercomputers.

Statistical methods have played an important role in seismology for many years—in large part due to the efforts of Harold Jeffreys [see Jeffreys (1977), for example]. Vere-Jones and Smith (1981) provide a review of many contemporary instances. Statistics enters for a variety of reasons. The data sets are massive. There is substantial inherent variability and measurement error. Models need to be refined, fitted and revised. Inverse problems need to be addressed. Experiments need to be designed. Sometimes the researcher must fall back on simulations. The basic quantity of concern is often a (risk) probability. In particular, it may be pointed out, that in the construction of the Jeffreys and Bullen (1940) travel time tables, one has an early, perhaps greatest success, of the use of robust/resistant methods. [B. A. Bolt's (1976) presidential address "Abnormal seismology" is well worth reading in this connection.]

Seismologists deal with data of a variety of types. The important forms are digital waveforms from spatial arrays of seismometers of various dimensions (where the instruments have been arranged in such a fashion that an earthquake signal may be seen as a moving, changing entity) and catalogs (containing lists of an event's times, locations, sizes and other characteristics) for geographic regions of interest.

Seismology is not without its controversies. There are fundamental ones, such as whether or not plate tectonics is a validated theory. There are practical ones, such as does the size of the motion of an earthquake increase steadily as one approaches the fault or does it level off? As is so often the case, the existing data and analysis methods prove inadequate to resolve these disputes conclusively.

A general reference that provides much of the pertinent seismological background is Bullen and Bolt (1985). We turn to a presentation of some specific problems and techniques.

2. Free oscillations of the Earth. This subject is one of the principal developments in seismology over the last 25 years. Whenever there is a great earthquake, the Earth vibrates for days afterwards. The seismogram then consists, approximately, of a sum of an infinite number of exponentially decaying cosinusoids plus noise; see expression (2). The frequencies of the cosinusoids and the corresponding rates of decay relate to the Earth's composition. Measured values may be used to make inferences about that composition. The techniques of complex demodulation, nonlinear regression and regularization may be employed in this connection. Some details on these techniques will follow.

As is the case with many natural systems, the vibratory motion of the Earth may be described by a system of equations of the form

$$\frac{d\mathbf{Y}(t)}{dt} = \mathbf{A}\mathbf{Y}(t) + \mathbf{X}(t), \tag{1}$$

with $\mathbf{X}(\cdot)$ a (vector-valued) input. In the case that the input is $\mathbf{b}\delta(t)$, with $\delta(\cdot)$ the Dirac delta function (corresponding to the earthquake shock) and initial conditions $\mathbf{Y}(0-) = 0$, the general solution of (1) may be written as

$$\begin{aligned}\mathbf{Y}(t) &= \exp\{\mathbf{A}t\}\mathbf{b} \\ &= \sum_j \zeta_j \exp\{\mu_j t\}\mathbf{u}_j, \qquad t > 0,\end{aligned}$$

where $\mu_j, \mathbf{u}_j$ are the (assumed distinct) latents of the matrix $\mathbf{A}$. The spectrum occurring is discrete because of the finiteness of the Earth as a body. Focusing on one of the coordinates of $\mathbf{Y}(t)$ and assuming the presence of noise, one has

$$Y(t) = \sum_k \alpha_k \exp\{-\beta_k t\}\cos(\gamma_k t + \delta_k) + \varepsilon(t), \tag{2}$$

with $-\beta_k$ and γ_k the real and imaginary parts of the μ_j and $\varepsilon(\cdot)$ the noise. This model may be checked by complex demodulation of the series $Y(t)$ in the neighborhood of frequencies γ_k, as estimated from the periodogram. Provided the bandwidth of the demodulation is not too great, a single cosinusoid should be

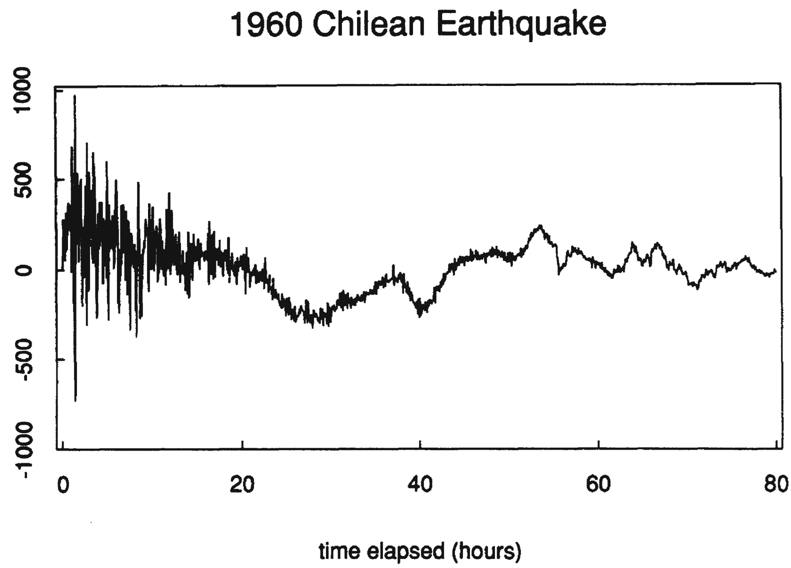


FIG. 1. *Record of the Chilean great earthquake of May* 22, 1960, *as recorded by the tiltmeter in the Grotta Gigante at Trieste. The tides have been partially removed.*

included, the log amplitude should fall off linearly with time and the phase angle should be approximately constant. Details are given later, specifically at (5).

Figure 1 is a plot of the seismogram recorded at Trieste of the 1960 Chilean great earthquake after partially removing the tides. Details re the data and the tidal removal procedure may be found in Bolt and Marussi (1962). Figure 2, a plot of the lower-frequency portion of the periodogram of this data, suggests the

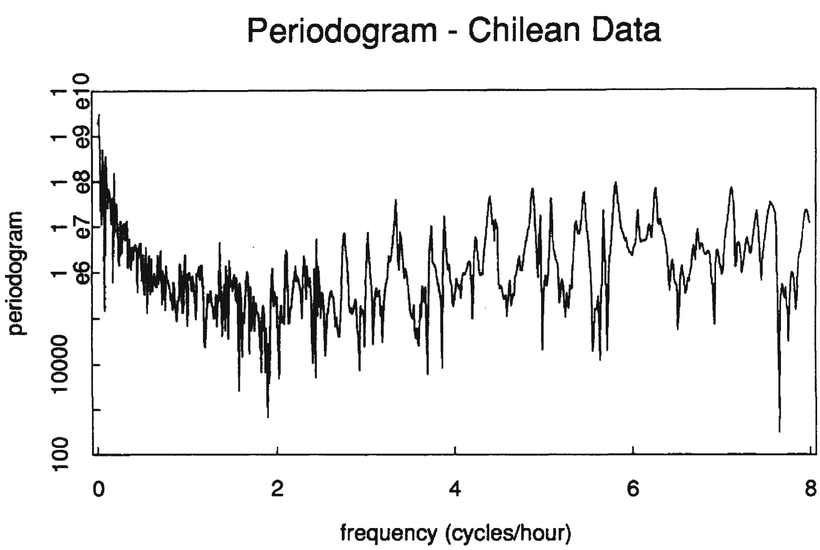


FIG. 2. *The periodogram of the data of Figure* 1 *based on* 2548 *data values. Only ordinates corresponding to frequencies less than* 8 *cycles/h have been graphed. The y-axis is logarithmic.*

presence of a variety of periodic components. The periodogram of a stretch of time-series values $Y(t)$, $t = 0, \ldots, T-1$, is defined as follows. Set

$$d_Y^T(\lambda) = \sum_{t=0}^{T-1} Y(t)\exp\{-i\lambda t\}, \qquad -\infty < \lambda < \infty. \tag{3}$$

Then the periodogram at frequency λ is defined as

$$I_{YY}^T(\lambda) = (2\pi T)^{-1}|d_Y^T(\lambda)|^2. \tag{4}$$

For data from the model (2), $I_{YY}^T(\lambda)$ may be expected to show peaks for λ near the γ_k.

The basic ideas of complex demodulation are frequency isolation by narrow-band filtering to focus on a single term in expression (2), followed by frequency translation to slow the oscillations down. The specific steps are: (i) $Y(t) \to Y(t)\exp\{i\lambda t\}$ (modulation), followed by (ii) local smoothing in t of $Y(t)\exp\{i\lambda t\}$ to obtain $Y(t, \lambda)$, the complex demodulate at frequency λ. In the case that $Y(t) = \alpha\exp\{-\beta t\}\cos(\gamma t + \delta)$, one has

$$\begin{aligned} Y(t,\lambda) &\approx \tfrac{1}{2}\alpha e^{i\delta} e^{-\beta t} e^{i(\lambda-\gamma)t}, \quad \text{for } \lambda \text{ near } \gamma \\ &\approx 0, \quad \text{otherwise.} \end{aligned} \tag{5}$$

Hence $\log|Y(t, \lambda)| \approx \log(\alpha/2) - \beta t$ and $\arg\{Y(t, \lambda)\} \approx \delta + (\lambda - \gamma)t$. Plots of these quantities versus t provide checks on model adequacy and provide preliminary estimates of parameters. Figures 3 and 4 present such plots for the Chilean data at two frequencies, 3.885 and 5.6775 cycles/h. These frequencies were determined by noticing the locations of peaks in the periodogram, setting λ equal to them, demodulating and then in some cases employing a nearby λ to get a more nearly horizontal phase plot. The fluctuations in the amplitude plot can be due to noise, to leakage from other frequency components or to split peaks among other things. The rate of decay β is found to generally vary with frequency in the present seismological situation. Results for the Chilean data for a variety of frequencies may be found in Bolt and Brillinger (1979).

The parameters could be estimated from the complex demodulate pictures, for example, by fitting regression lines. It is generally more effective to proceed via nonlinear regression. This has the further advantage of providing estimated standard errors. Suppose one has a model

$$Y(t) = S(t; \theta) + \varepsilon(t),$$

with $S(\cdot)$ known up to the finite-dimensional parameter θ and $\varepsilon(\cdot)$ a noise series. In the present case, $S(t) = \alpha\exp\{-\beta t\}\cos(\gamma t + \delta)$ and $\theta = \{\alpha, \beta, \gamma, \delta\}$. For the next step, it is convenient to take $\lambda_j = 2\pi j/T$ and to write $Y_j = d_Y^T(\lambda_j)$, $E_j = d_\varepsilon^T(\lambda_j)$ and $S_j(\theta) = d_S^T(\lambda_j)$. One will estimate θ by minimizing

$$\sum_{j \text{ in } J} |Y_j - S_j(\theta)|^2, \tag{6}$$

for J a range of subscripts with λ_j near γ. The logic of this is as follows. There are a variety of central limit theorems for empirical Fourier transforms [see, for example, Brillinger (1983)]. Suppose that the noise series $\varepsilon(\cdot)$ is stationary and

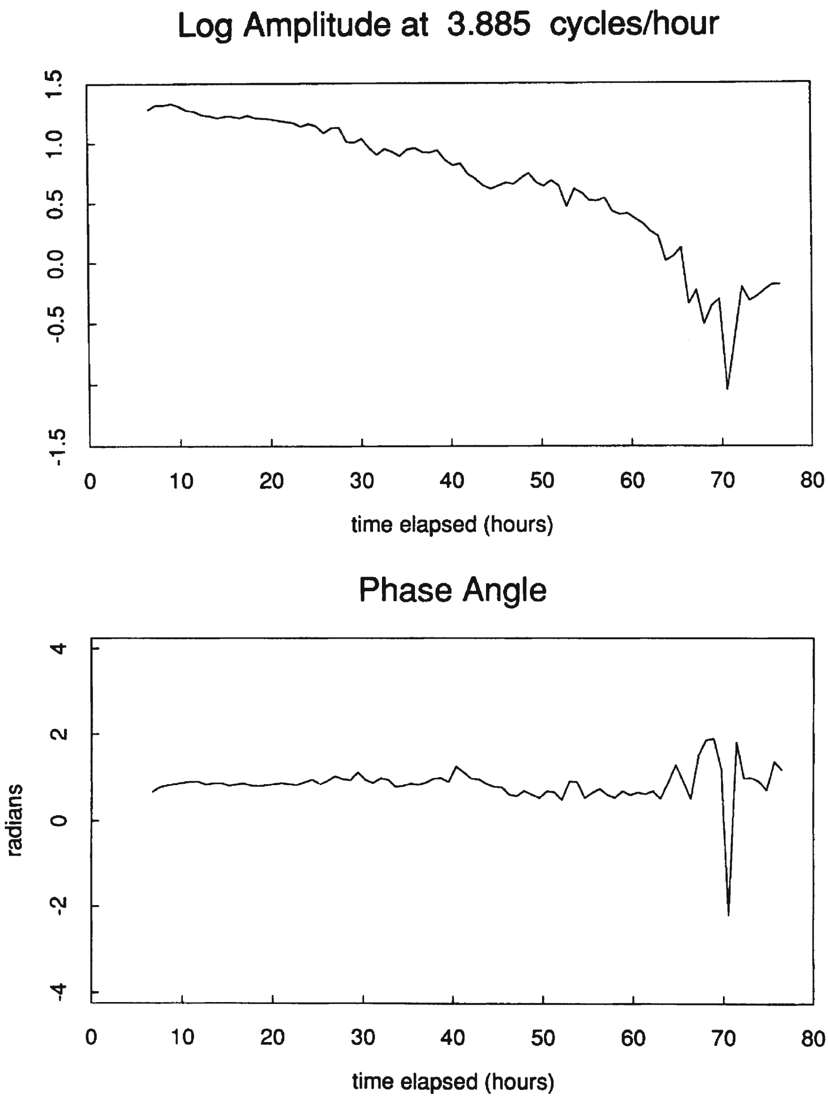


FIG. 3. *The result of complex demodulating the data of Figure* 1 *at a frequency of* 3.885 *cycles/h. The upper graph gives the logarithm of the running amplitude. The lower graph gives the running phase. The bandwidth of the filter employed is* 0.594 *cycles/h.*

mixing with power spectrum $f_{\varepsilon\varepsilon}(\lambda)$. Then for large T, E_j is approximately complex normal with mean 0 and variance $2\pi T f_{\varepsilon\varepsilon}(\lambda_j)$. Further the variates E_j, E_k are approximately independent. It follows that the determination of an estimate of θ to minimize expression (6) is approximately the maximum likelihood procedure. The statistical properties of such estimates were indicated in Bolt and Brillinger (1979) and developed in detail in Hasan (1982). For example, one finds the asymptotic variance of $\hat{\gamma}$ to be proportional to

$$\frac{4\pi f_{\varepsilon\varepsilon}(\gamma)}{T^3\alpha^2},$$

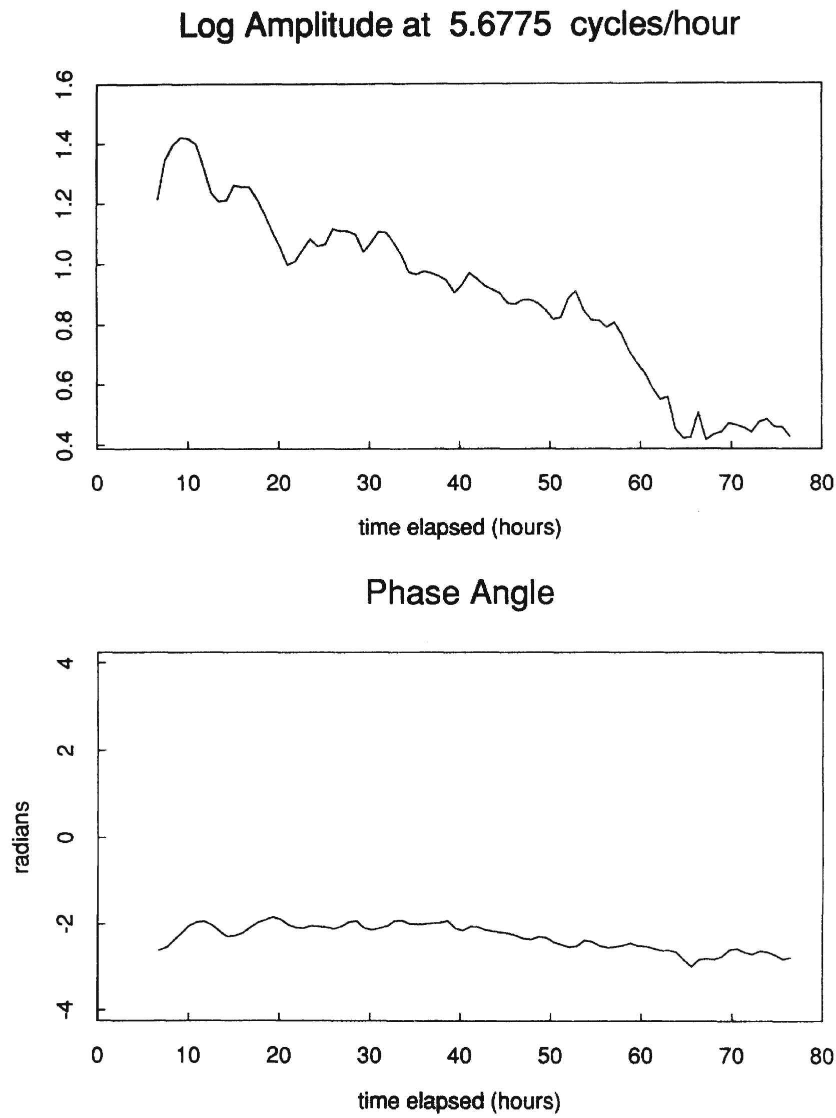


FIG. 4. *Complex demodulation as for Figure* 3, *but at the frequency of* 5.6775 *cycles/h.*

having considered a limiting process with $\beta = \phi/T$ as $T \to \infty$. The inverse cubic dependence on sample size is on first glance surprising. It comes from the narrowness of the peaks when they are present.

Complex demodulation is an exploratory technique. Hence one has to be conscious of the possibility of employing it at frequencies of "false" peaks. In practice, it is found that the nearness of the phase plot to a straight line is a highly sensitive indicator of the presence of a periodic component.

Earlier in the paper, it was noted that progress in seismology shows a to-and-fro between new knowledge of waves and new knowledge of the structure of the Earth. This occurs in the case of free oscillations. Suppose one has an initial model for the Earth in terms of some physical parameters, e.g., expres-

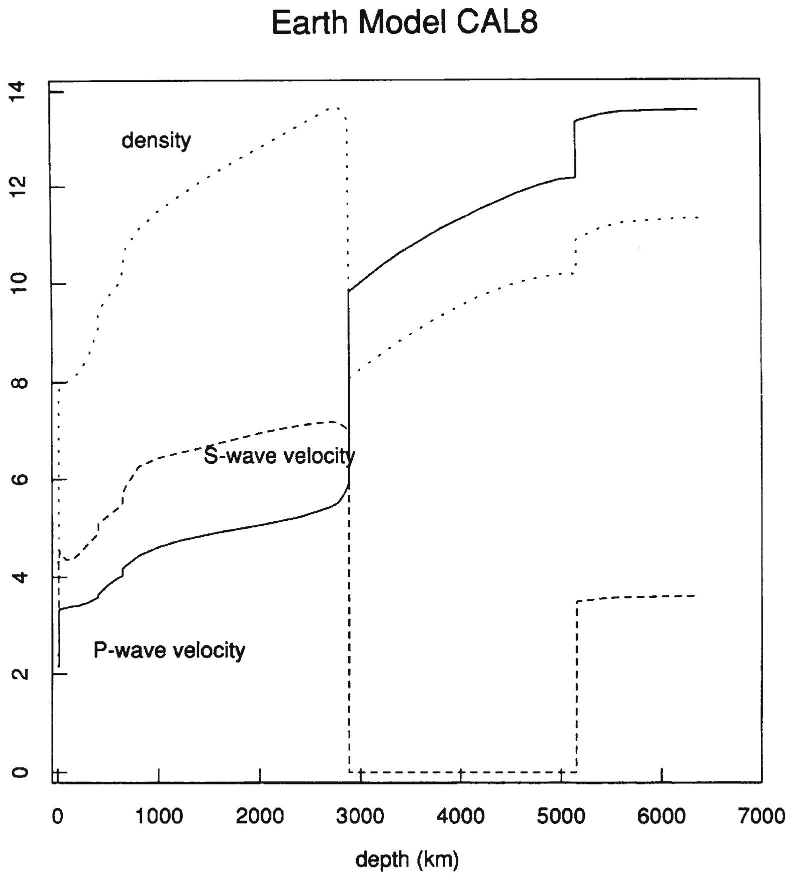


FIG. 5. *The* CAL 8 *Earth model. The curves give the assumed density (grams per cubic centimeter), P-wave velocity (kilometers per second) and S-wave velocity (kilometers per second) as a function of depth assuming a spherical Earth. Given such a model, one can compute implied periods of free oscillation. Of interest is the inverse problem, given periods what is a corresponding Earth model?*

sions for density, shear wave velocity and compression wave velocity as functions of depth, say $\rho(r)$, $c_S(r)$ and $c_P(r)$, respectively, r denoting depth. Figure 5, based on the data in Tables 3 and 4 of Bolt (1982), shows what is meant by an Earth model. Given such a model, one can compute the implied frequencies of free oscillation γ_k. How to do this is described in Chapter 6 of Lapwood and Usami (1981), for example. The relationship involved is nonlinear, but perturbations may be expressed linearly via kernels. Specifically, suppose one perturbs the parameters by amounts $\Delta\rho$, Δc_S and Δc_P, respectively, then the perturbation of the frequency of the kth free oscillation is given by

$$\Delta\gamma_k \approx \int_0^R A_k(r)\,\Delta\rho(r)\,dr + \int_0^R B_k(r)\,\Delta c_S(r)\,dr + \int_0^R C_k(r)\,\Delta c_P(r)\,dr,$$

for kernels A_k, B_k and C_k. This expression is said to lay out the "direct problem": Given $\Delta\rho$, Δc_S and Δc_P find $\Delta\gamma_k$. Now suppose a great earthquake occurs. Then new estimates of the frequencies γ_k are available. One has the "inverse problem": Given the observed $\Delta\gamma_k$, find $\Delta\rho$, Δc_S and Δc_P. Because the new frequencies are just estimates, one seeks a model only approximately achieving them. It seems worth setting out the type of problem involved here in a specific notation. Let Y and Θ denote normed spaces. Let $\mathbf{X}$ denote a map from Θ to Y, $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta}$. The values $\mathbf{Y}$ and $\mathbf{X}$ are given, a value for $\boldsymbol{\theta}$ is desired. Let α denote a scalar. Some, basically similar, methods for selecting a $\boldsymbol{\theta}$ currently being employed include: (a) regularization, choose $\boldsymbol{\theta}$ to minimize $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\|^2 + \alpha\|\boldsymbol{\theta}\|^2$; (b) sieve, choose $\boldsymbol{\theta}$ subject to $\|\boldsymbol{\theta}\| \leq \alpha$ to minimize $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\|$; (c) residual, choose $\boldsymbol{\theta}$ subject to $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\| \leq \alpha$ to minimize $\|\boldsymbol{\theta}\|$. A characteristic of the solutions obtained is that one has to be content with the estimation of some form of average of the unknown $\boldsymbol{\theta}$. Chapter 12 of Aki and Richards (1980) contains a discussion of inverse problems in geophysics. A characteristic that distinguishes the present Earth model problem, from the usual inverse problems, is that there are discontinuities present in the model—corresponding to the Earth's layers. The above perturbation approach of a nonlinear problem to a linear one has been employed by geophysicists for many years; see Jeffreys and Bullen (1940), for example.

Several other references to the study of free oscillations may be noted. Hansen (1982) extends the procedure of Bolt and Brillinger (1979) to handle the case of several eigenfrequencies present in the nonlinear regression fit. Dahlen (1982) sets down the asymptotic results for the case of tapered data, that is, when convergence factors have been introduced into the Fourier transform computations. Zadro and Caputo (1968) look for nonlinearities via bispectral analysis.

3. Estimation of fault-plane parameters. That there exists a see-saw between the study of the Earth's structure and the study of earthquake sources was pointed out earlier. In this section it will be indicated how a (nonlinear) probit analysis may be employed to estimate basic characteristics of the source of an earthquake.

An important quantity read off the seismic trace of an earthquake at a particular observatory is the sign of the increment at the arrival of the first energy from the event. This sign corresponds to whether the initial motion is a compression or a dilation. In many cases, following the observation of an earthquake at a number of stations, if the observed signs of first motion are plotted on a map centered at the epicenter of the event a (radiation) pattern results. Figure 6, taken from Brillinger, Udias and Bolt (1980), provides such a plot for one of the aftershocks (event 4) of the Good Friday 1964 Alaskan event. (Unfortunately, due to the locations of the particular stations recording the event, this figure does not provide a particularly good example of the ideal radiation pattern, but the data were of special interest. Were the stations well scattered, in an ideal circumstance one would see mainly solid circles in two opposite quadrants and mainly open circles in the other two quadrants. In this case only two of the four quadrants have been covered. The implication will be that one of the planes will be poorly determined.) Following Byerly (1926), plots such as this have been employed to learn about the source. Before describing

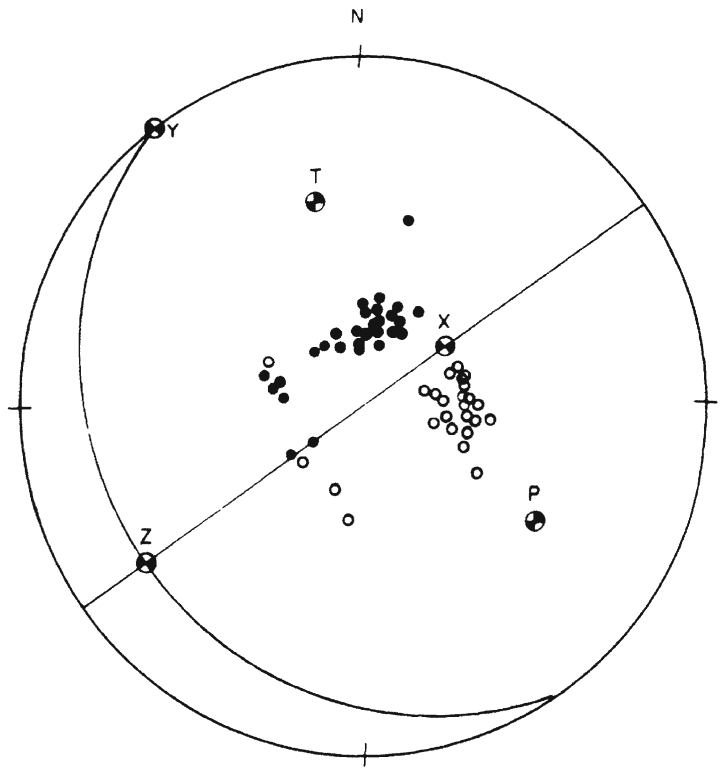


FIG. 6. *The P-wave first-motion data for the earthquake of the Alaskan sequence that took place March* 30, 1964 *at* 0200. *The solid circles refer to compressions, i.e., first motion upward, the open circles to dilations, i.e., first motion downward.* [*Reproduced with permission from Brillinger, Udias and Bolt* (1980).]

what may be learned, some details of the earthquake process will be set down. The usual assumption (the elastic rebound theory) is that earthquakes are due to faulting. A crack initiates at a point and (in the case of pure slip) spreads out to form a fault plane. As the crack passes a given point, slip takes place (on the fault plane) resulting in a stress drop and the radiation of seismic waves. The radiated (P-) waves may be shown to have a quadrantal pattern with one of the axes parallel and the other perpendicular to the fault plane of the event. It follows, and this is what Byerly (1926) contributed, that the data may be used to estimate the fault-plane orientation. Having an estimate of the fault plane and the direction of motion on that plane is important to geology and geophysics. Researchers seek to tie together surface and subsurface features, to consider regional stress directions and to use the results to confirm and extend the theory of plate tectonics. The results can be crucial to seismic risk computations.

Byerly proceeded graphically and this has continued to generally be the working approach. However, the results so obtained are subjective, have no attached measure of uncertainty and may not be easily combined with estimates derived from other events at the same site.

The problem may be approached in formal statistical fashion as follows. The data available consist of hypocenter of earthquake, locations of observatories, directions of observed first motions (compressions or dilations) at the observatories and a store of knowledge concerning the Earth's structure [velocity models as given in Jeffreys and Bullen (1940), for example]. It may further be argued that seismographic noise is approximately Gaussian [see Haubrich (1965)]. Let a fault plane be described by three angles $(\theta_T, \phi_T, \theta_P)$. Let $A_{ij}(\theta_T, \phi_T, \theta_P)$ denote the theoretical expression for the wave amplitude on the focal sphere for event i at station j. This expression may be found in Brillinger, Udias and Bolt (1980). (The focal sphere is a "little" sphere of unit radius around the hypocenter. In carrying out the amplitude computation, one has to trace the ray from the hypocenter to the observatory through the focal sphere.) Let Y_{ij} denote the realized amplitude of the seismogram at the onset of the event. Then one can write $Y_{ij} = \alpha_{ij}A_{ij} + \varepsilon_{ij}$, with α_{ij} a scale factor and ε_{ij} normal mean 0 and variance σ_{ij}^2 variate. Here α_{ij} reflects the attenuation the signal experiences in traveling from the source to the observing station, whereas ε_{ij} represents noise caused by disturbances unrelated to the earthquake of concern. Let $y_{ij} = 1$ if $Y_{ij} > 0$ and $= 0$ otherwise. It follows that

$$\text{Prob}\{y_{ij} = 1\} = \text{Prob}\{Y_{ij} > 0\} = \Phi(\rho_{ij}A_{ij}),$$

writing $\rho_{ij} = \alpha_{ij}/\sigma_{ij}$, for this signal-to-noise ratio. The model may be further expanded by including a term γ_{ij} to allow for reader and recorder errors, now writing

$$\text{Prob}\{y_{ij} = 1\} = \gamma_{ij} + (1 - 2\gamma_{ij})\Phi(\rho_{ij}A_{ij}). \tag{7}$$

Precise data correspond to γ and σ small (hence ρ large) and imprecise to γ near 0.5 or ρ near 0.

The model is seen to take the form of a nonlinear probit (with a term γ corresponding to "natural mortality"). An example of a corresponding likelihood is provided by

$$\prod_{ij}\Phi(\rho_i A_{ij})^{y_{ij}}\left(1 - \Phi(\rho_i A_{ij})\right)^{1-y_{ij}}, \tag{8}$$

assuming ρ to depend on event alone and $\gamma = 0$. One can now proceed to estimate the unknown parameters $\theta_T, \phi_T, \theta_P, \rho_i$ by maximum likelihood.

Figure 6 includes the fitted planes for the case of event 4 of the Alaska sequence. These particular estimates were computed restricting the likelihood (8) to the observations of event $i = 4$ and including a γ term as in (7).

It is critical to assess the fit of any model. In Brillinger, Udias and Bolt (1980), this was done by comparing the theoretical and estimated probability functions. Figure 7 is based on a pooled analysis of some 16 of the Alaskan events (labeled by i previously) that seemed to go together. It has been assumed that the ρ_i are all equal in the fit studied. The figure provides the empirical probability that the observed first motion agrees with the theoretical as a function of amplitude. The fitted values $z = \hat{\rho}\hat{A}_{ij}$ have been grouped into cells of width 0.1 in the analysis.

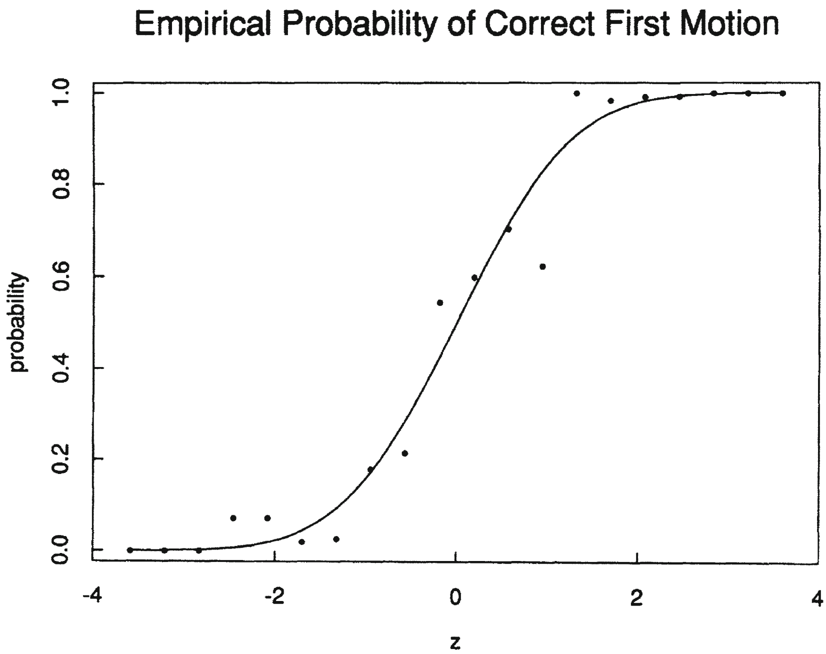


FIG. 7. *A plot of the statistic* (9) *of Section* 3 *and* $\Phi(z)$ *for the data of* 16 *events of the Alaskan sequence of* 1964. *The plot is meant to assess the validity of the model* (7). *Here z refers to the values* $\hat{\rho}\hat{A}_{ij}$.

What is plotted are $\Phi(z)$ and

$$(9)\quad \begin{aligned} &\#\{(i,j) | \operatorname{sgn} Y_{ij} \\ &\quad = \operatorname{sgn} \hat{A}_{ij},\ z - h < \hat{\rho}\hat{A}_{ij} < z + h\} / \#\{(i,j) | z - h < \hat{\rho}\hat{A}_{ij} < z + h\}, \end{aligned}$$

for $h = 0.5$. Here $\hat{A}$ refers to $A(\hat{\theta}_T, \hat{\phi}_T, \hat{\theta}_P)$ and # refers to the count of the number of elements in the set. The fit seems adequate.

The results of further computations of this type may be found in Brillinger, Udias and Bolt (1980) and Buforn (1982). The maximization program VA09A of the Harwell subroutine library, see Hopper (1980), proved effective in determining the maximum likelihood values. The estimates were, however, nonunique and poorly determined in some cases of small data sets.

An important by-product of such analyses is to form clusters of like fault-plane solutions for events in the same region, in order to get at motions occurring on the same fault plane; see Udias, Munoz and Buforn (1985), for example. The maximum likelihood standard errors are useful in this connection. The practical implication of the work just reported is that first motions for large collections of events may be handled routinely and that geophysical conjectures may be checked formally. The final fault-plane solution may be plotted in traditional fashion allowing examination of the data for difficulties. What remains is for more realistic seismic source models than the one treated in the papers listed to be fitted statistically. An elementary reference to the subject matter of concern here is Boore (1977).

4. Quantification of earthquakes. One of the important and difficult questions of seismology is how to measure the "size" of an earthquake. Size is an essential feature that a seismologist makes use of in attempts to deal with earthquake hazards and to understand the basic phenomena of concern. Specifically, the seismologist is not only interested in estimating the direction of movement at the source, he is further interested in the overall deformation that took place and the amount of energy that was released. Among the physical quantities of interest for a given earthquake are the seismic moment (a measure of the seismic energy released from the entire fault) and the stress drop (difference between the initial and final stress.)

For a variety of seismic source models, seismologists have related the seismic moment and stress drop to characteristics of the amplitude spectrum $|S(\lambda)|$, the modulus of the Fourier transform of the signal. Suppose that the seismogram is written as

$$Y(t) = s(t; \theta) + \varepsilon(t),$$

where $s(\cdot)$ is the signal, θ is an unknown parameter and $\varepsilon(\cdot)$ is a noise disturbance. If $S(\lambda; \theta)$ denotes the Fourier transform of $s(t; \theta)$, then what is given, from the source model, is the functional form of $|S(\lambda, \theta)|$. A reason for working in the Fourier domain here is that distracting phase information is eliminated. Common forms (for displacement measurements) include

$$|S(\lambda; \theta)| = \alpha / \sqrt{1 + (\lambda/\lambda_0)^\beta} \text{ and } \alpha / \left\{1 + (\lambda/\lambda_0)^2\right\},$$

with $\theta = \{\alpha, \beta, \lambda_0\}$. The seminal paper on the determination of such functional forms and on the relationship of their parameters to the "size" of the earthquake is Brune (1970/1971). Estimates of the seismic moment and stress drop may be determined once estimates of α and λ_0 are available. That the parameters relate to size and duration will be seen for a particular functional form in the discussion that follows. The empirical practice has been to estimate the unknowns graphically from a plot of the modulus of the amplitude of the empirical Fourier transform $|d_Y^T(\lambda)|$. The following formal procedure was suggested in Brillinger and Ihaka (1982).

The asymptotic distribution of $|d_Y^T(\lambda)|$ may be evaluated in the case of stationary ε using a central limit theorem of the type mentioned in Section 2. The asymptotic distribution is found to depend on $|S(\lambda; \theta)|$ and $f_{\varepsilon\varepsilon}(\lambda)$ alone. Hence one needs an expression only for the modulus of S, and as stated above, this is what the seismologist generally provides. Next, with the model $Y(t) = s(t; \theta) + \varepsilon(t)$ and small noise,

$$|d_Y^T(\lambda)| = |S(\lambda; \theta)| + \left(d_\varepsilon^T(\lambda) + d_\varepsilon^T(-\lambda)\right)/2 + \cdots,$$

showing variation around $|S|$ not depending on $|S|$. However, when deviations of $|d_Y^T|$ from a final fitted form are plotted versus the fitted values, dependence of the error on $|S|$ is apparent. An example is provided in Figure 8. This is the

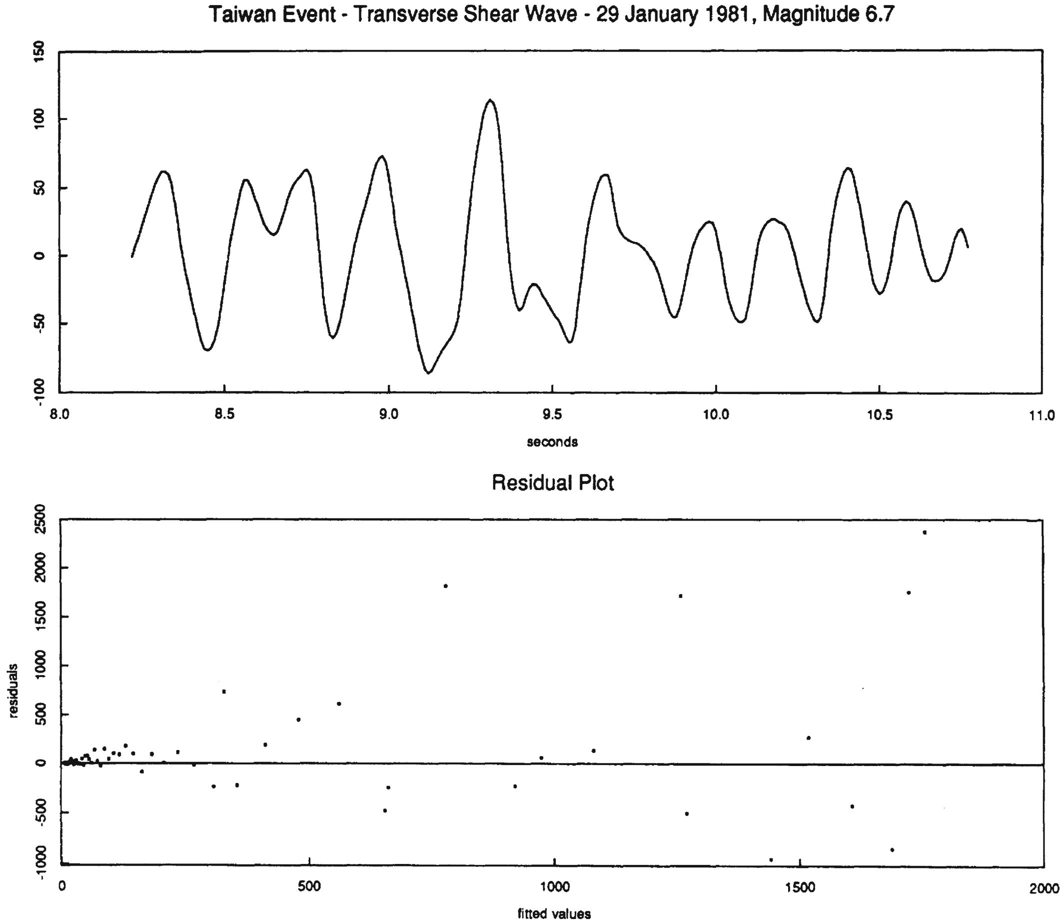


FIG. 8. *The upper graph gives the transverse S-wave component of the vibrations of the January* 29, 1981, *magnitude* 6.7, *Taiwan earthquake as recorded by the central accelerometers of the Smart* 1 *array. The array is approximately* 30 *km northwest of the epicenter of the event. The lower graph plots the differences between the amplitudes of the Fourier transform of the data and corresponding (final) fitted values. The data stretch consisted of* 256 *points.*

result of computations for an earthquake of magnitude 6.7 that occurred in Taiwan on January 29, 1981. The data were recorded by one of the instruments of the Smart 1 array; see Bolt, Tsai, Yeh and Hsu (1982). The upper graph of the figure provides the transverse S-wave portion of the recorded accelerations. The lower graph provides the deviations plot just referred to. This plot suggests that the noise is in part "signal generated" in this case. There are various physical phenomena that can lead to signal-generated noise. These include multipath transmission, reflection and scattering. The following is an example of a model that includes signal-generated noise:

$$Y(t) = s(t) + \sum_k \left(\gamma_k s(t - \tau_k) + \delta_k s^H(t - \tau_k)\right) + \varepsilon(t), \qquad (10)$$

with the τ_k time delays, with s^H the Hilbert transform of s and with γ_k, δ_k

reflecting the vagaries of the transmission process. [The inclusion of the Hilbert transform allows the presence of phase shifts. The Hilbert transform is discussed, for example, in Brillinger (1975a), page 32]. With the γ_k, δ_k, τ_k random and after evaluating the large sample variance, one is led to approximate the distribution of $Y_j = d_Y^T(\lambda_j)$ by a complex normal with mean $S(\lambda_j; \theta)$ and variance $\Gamma_j = 2\pi T(\rho^2|S(\lambda_j; \theta)|^2 + \sigma^2)$, where now ε has been assumed to be white noise (of variance σ^2), and also it is assumed that $E\gamma_k$, $E\,\delta_k = 0$ and that the process τ_k is Poisson. The ratio ρ^2/σ^2 measures the relative importance of signal-generated noise. This variance is seen to depend on the "signal" through $|S|$ and leads to

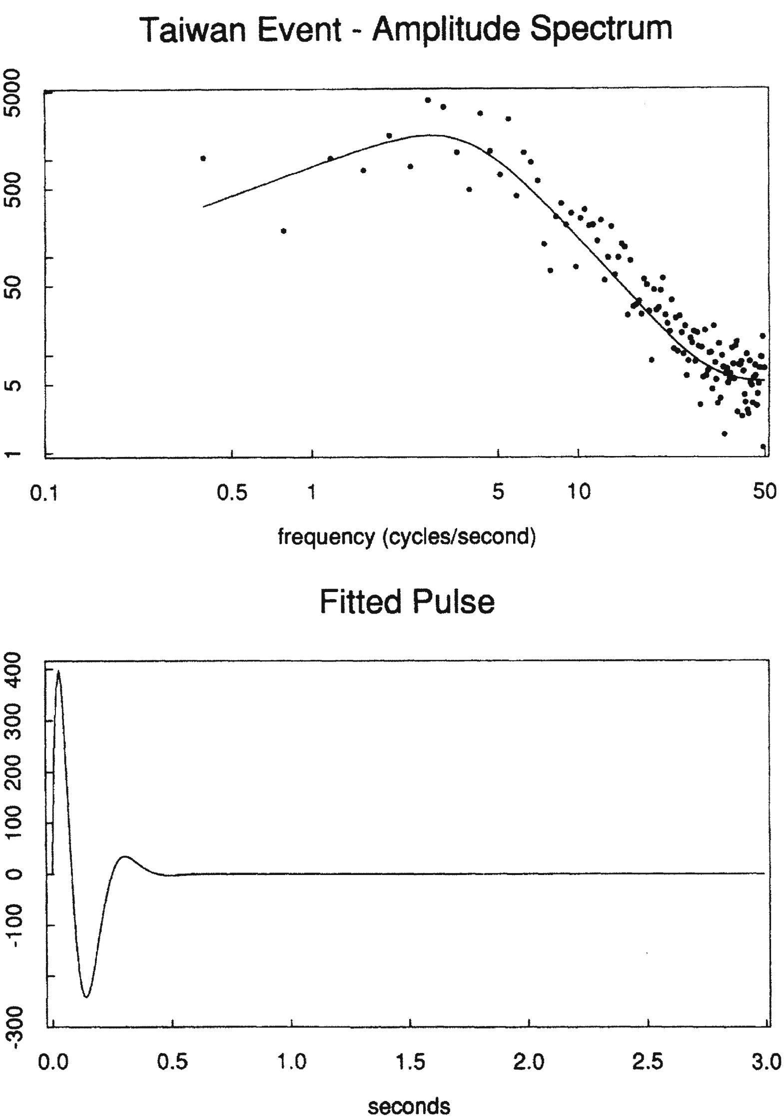


FIG. 9. *The upper graph provides the amplitudes of the Fourier transforms of the Taiwan data of Figure 8 and the corresponding fitted expected values as computed for the model of Section 4. Both scales of the plot are logarithmic. The lower graph provides the fitted pulse $\hat{s}(t)$.*

wedging of the type present in Figure 8. One can proceed to estimate θ by deriving the marginal likelihood based on the $|Y_j|$. This likelihood may be evaluated and found to be

$$\prod_j \left(\exp\left(-\frac{|Y_j|^2 + |S_j|^2}{\Gamma_j} \right) I_0\left(\frac{2|Y_j|\,|S_j|}{\Gamma_j} \right) \frac{1}{\Gamma_j} \right),$$

where I_0 denotes a modified Bessel function. The upper graph of Figure 9 shows a fit of the model $|S(\lambda)| = \alpha|\lambda|/(1 + (\lambda/\lambda_0)^4)$ to the data of Figure 8. This functional form was settled on after the degree of fit of two more elementary forms was examined.

Details may be found in Ihaka (1985). We remark that this model fit corresponds to a time domain pulse $s(t) = \alpha\lambda_0^2 p(\lambda_0 t)$, where

$$p(t) = \left[\sin\frac{t}{\sqrt{2}} - t\sin\left(\frac{t}{\sqrt{2}} + \frac{\pi}{4} \right) \right] e^{-t/\sqrt{2}},$$

for $t > 0$ and $p(t) = 0$ otherwise. The expression $s(t) = \alpha\lambda_0^2 p(\lambda_0 t)$ indicates how λ_0 corresponds (inversely) to the duration of the event and how α corresponds to size. The lower graph of Figure 9 provides a plot of the fitted pulse. Once estimates of α, λ_0 are at hand, these may be converted to estimates of the seismic moment and stress drop via theoretical relationships developed by geophysicists.

The maximum likelihood fit of the model was carried out by a computer program written by Ihaka. This program also generates standard error estimates and standardized residuals. These later may be used to assess the goodness of fit of the model. Figure 10 provides a plot of the standardized residuals against the

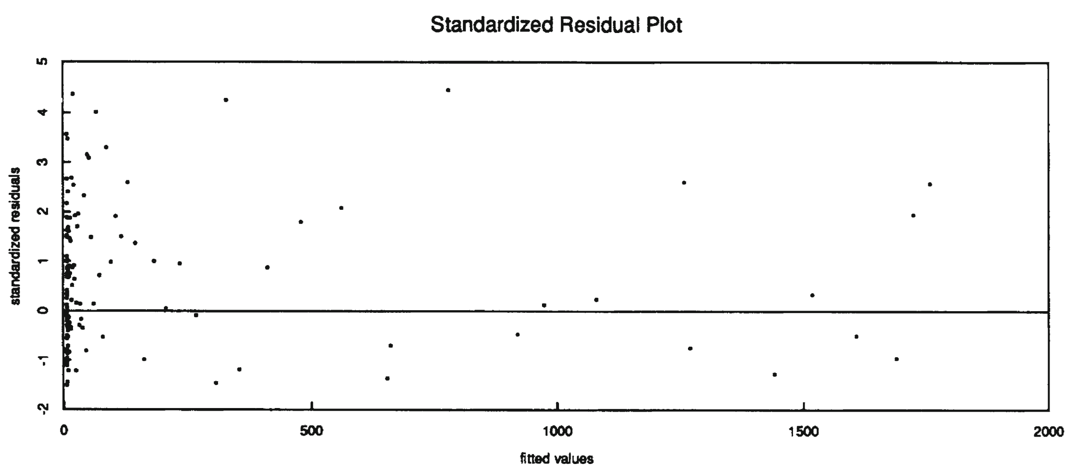


FIG. 10. *A standardized residual plot, based on the model* (10), *corresponding to the lower graph of Figure* 8. *The differences between the amplitudes of the Fourier transform values and their fitted expected values have been divided by their fitted standard deviations to obtain standardized residuals.*

fitted values of the same format as the residual plot of Figure 8. The wedging corresponding to signal-generated noise in the later plot is no longer present; however, there is a definite suggestion that the fit might be improved in the region where the signal has low amplitude. Luckily, this is the region of least importance. It awaits future analysis. It might be handled by allowing the series $\varepsilon(t)$ to have a nonconstant spectrum.

5. Array data. Today it would be a strange thing indeed for an earthquake to be recorded on just one seismometer. In fact, from the very earliest days, readings of the same event at geographically scattered observatories have been made use of. Since the 1960s, seismometers have been deliberately arranged in geometric designs over distances of the order of miles to hundreds of miles in order to allow extraction of traditional information and sometimes ellicitation of new information.

An important use has been the estimation of the direction from which a seismic signal is arriving and the velocity with which it is moving. One manner in which this is done is by the computation of estimates of frequency–wavenumber spectra. The procedure may be described as follows. Suppose one has array data; $Y(x_j, y_j, t)$, $j = 0, \ldots, J$ and $t = 0, \ldots, T-1$. Here (x_j, y_j) denotes the coordinates of the location of the jth sensor. The frequency–wavenumber periodogram of this data is given by

$$\text{(11)} \qquad \left| \sum_j \sum_t Y(x_j, y_j, t) \exp\{ -i(\mu x_j + \nu y_j + \lambda t)\} \right|^2, \qquad -\infty < \mu, \nu, \lambda < \infty.$$

A motivation for this definition is the following. Suppose one has a plane wave $Y(x, y, t) = \rho \cos(\alpha x + \beta y + \gamma t + \delta)$ of temporal frequency γ and wavenumber $\kappa = (\alpha, \beta)$. Then the periodogram will have a peak near (α, β, γ). (Incidentally, this wave is moving with apparent velocity $\gamma/\sqrt{\alpha^2 + \beta^2}$ from azimuth given by $\tan \phi = \beta/\alpha$.) An example of array data is given by Figure 11. What is plotted are the locations of nine of the seismometers of the Smart 1 array located in Taiwan. Also plotted are the portions of the traces used in the computations. These traces correspond to the vertical P-wave part, of the January 29, 1981 earthquake. (The initial near-flat part is the noise, saved in a buffer, just before the onset of the wave.) The estimated epicenter of this earthquake was 30 km southeast of the array. Figure 12 gives a central portion of the frequency–wavenumber periodogram, for this data, as computed via formula (11), at frequency λ corresponding to 1.944 cycles/s. (The temporal frequency 1.944 was picked on the basis of a times-series analysis of the individual seismograms.) There is seen to be a large peak in the southeast quadrant, at an azimuth that turns out to correspond to that of the epicenter of the event. The radial distance corresponds to the velocity of P-waves.

Seismologists working with this type of data have often preferred to employ, what they call, the "high-resolution" or "Capon" statistic [see Capon (1969)] instead of the periodogram (11). The high-resolution statistic typically shows more dramatic peaks than the periodogram. Before defining it, we introduce

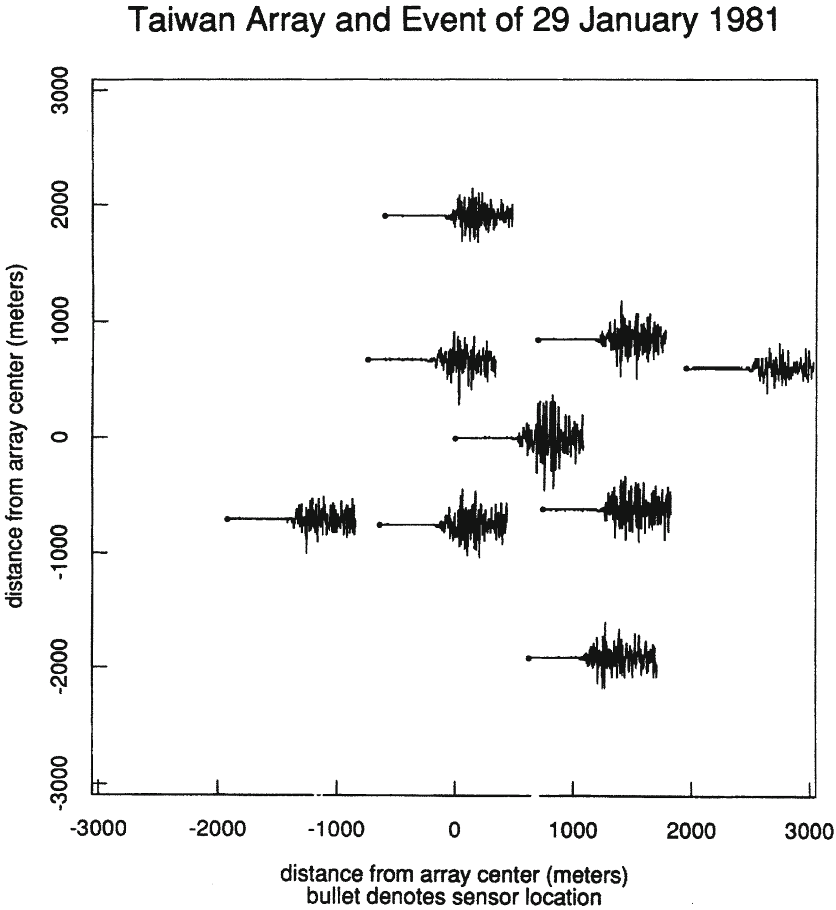


FIG. 11. *The vertical P-wave portion of the January 29 Taiwan earthquake as recorded at nine of the sensors of the Smart 1 array. The "bullets" are plotted at the physical locations of the sensors. Noise immediately preceding the arrival of energy from the event had been saved in a buffer.*

some notation. Let $\mathbf{Y}(t)$ denote the j-vector $[Y(x_j, y_j, t)]$. Set

$$\mathbf{Y}_k = T^{-1} \sum_{t=0}^{T-1} \mathbf{Y}(t) \exp\left\{-i\frac{2\pi kt}{T}\right\},$$

for $k = 0, 2, \ldots$. Further let $\mathbf{B} = [\exp\{-i(\mu x_j + \nu y_j)\}]$. If $\lambda = 2\pi l/T$, l an integer, then the periodogram (11) is proportional to $|\overline{\mathbf{B}}^{\tau}\mathbf{Y}_l|^2$. Next define

$$\mathbf{M} = \sum \mathbf{Y}_k \overline{\mathbf{Y}}_k^{\tau},$$

with the sum over k with $2\pi k/T$ near λ. Now the high-resolution statistic at frequency λ may be defined as $1/\overline{\mathbf{B}}^{\tau}\mathbf{M}^{-1}\mathbf{B}$. If $Y(x, y, t) = \rho\cos(\alpha x + \beta y + \gamma t + \delta)$ + noise, this statistic may be expected to show a peak for (μ, ν) near (α, β) and λ near γ. This statistic has been introduced, in part, in order to be able to present the next example.

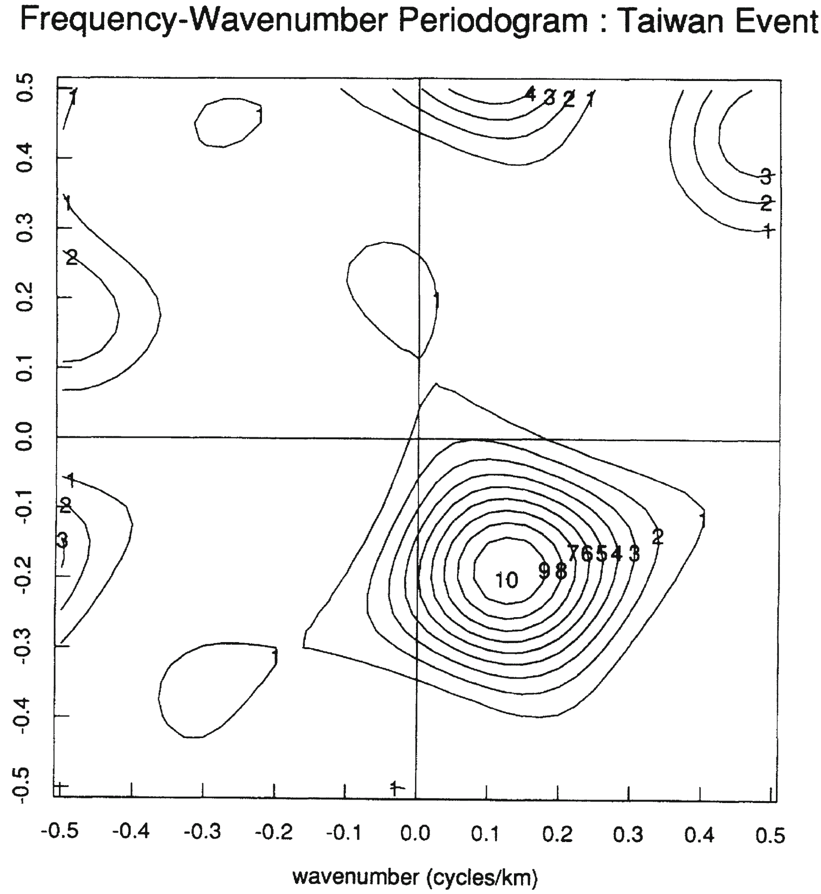


FIG. 12. *The frequency–wavenumber periodogram of the data of Figure* 11. *The time series stretches contain* 720 *points. The temporal frequency employed is* 1.94 *cycles/s.*

Figure 13 is reproduced from Scheimer and Landers (1974). It shows the high-resolution statistic computed for two portions of data recorded by the Large Aperture Seismic Array (LASA) in Montana following a strip-mining blast. These computations confirmed the validity of the high-resolution approach. The statistic for one portion shows a single large peak in the direction of the blast. The statistic for the following portion shows energy arriving from various directions. This analysis provided empirical proof of the existence of scattering of seismic waves. That this phenomenon existed had been theorized for years. A frequency–wavenumber data analysis has provided the confirmation.

Spectral analyses are (too) often thought of as being appropriate only for stationary data. As the preceding example shows, the technique may be highly useful in nonstationary cases as well. As a second example we mention the results of Bolt, Tsai, Yeh and Hsu (1982). If, in fact, an earthquake is caused by faulting, then the direction of the source of seismic energy will be changing as the fault is ripping, that is, as the fault tip is advancing. In the paper cited, Bolt, Tsai, Yeh and Hsu present high-resolution spectra for succeeding time stretches

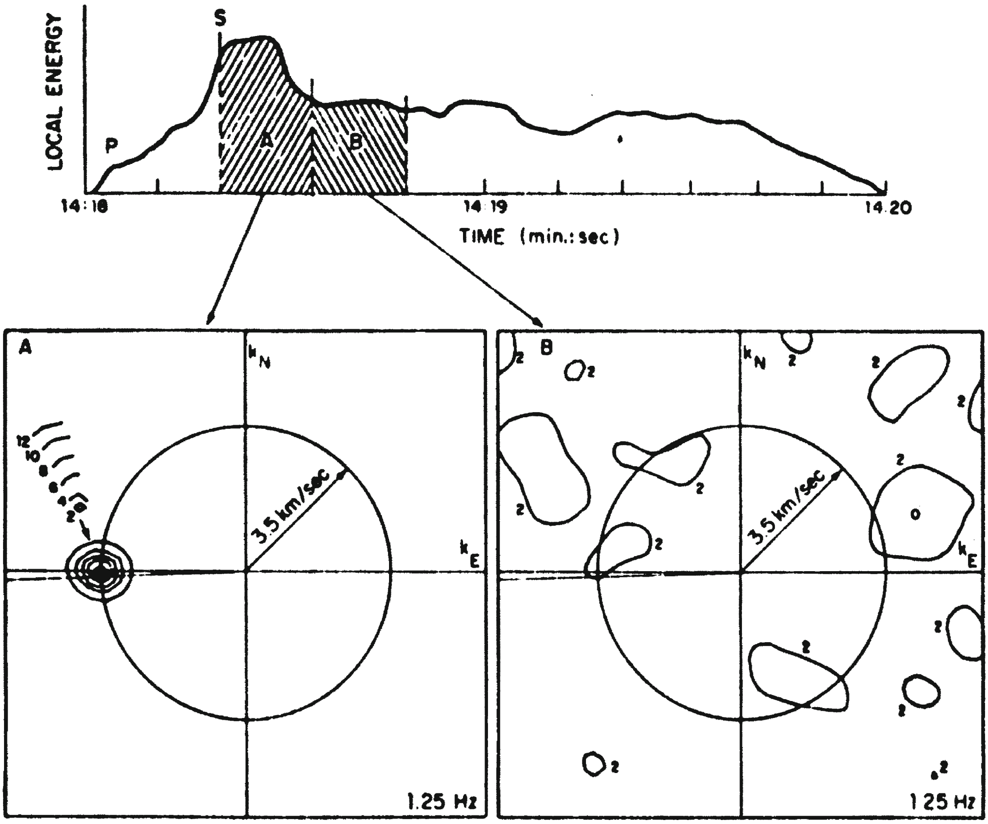


FIG. 13. *The high-resolution (or Capon) spectrum computed for S-wave data recorded, following a strip-mining blast, at the Large Aperture Seismic Array located in Montana. The temporal frequency was* 1.25 *cycles/s (corresponding to the principal frequency of the S-wave).*

of the January 29, 1981 Smart 1 event. There is an apparent shift in direction with time. Their work may have been the first experimental measurement of a seismic dislocation moving along a rupturing fault.

In each of the preceding two examples, frequency–wavenumber analysis has allowed researchers to confirm the presence of suspected scientific phenomena.

6. Exploration seismology (reflection seismology). The problem of learning the Earth's crustal structure can be approached as one of system identification. The approach to be described takes advantage of the fact that the Earth happens to be made up of layered strata. Signals, such as powerful impacts or explosions, can be deliberately input to the Earth and the consequent vibrations recorded by an array of seismometers or geophones. Such experiments may be carried out in a search for gas and oil, or in a scientific study of the general geological makeup of a region of interest. The results of these experiments may be viewed as one of the grand success stories for statistical techniques generally, and of least squares particularly. An unusual aspect of the inferences made is that in many cases one gets to examine their validity, by the later drilling of a well. No worked example is presented in this section of the paper, in large part because data sets are hard to come by. The material is presented, however, because it provides a case where a rather complete solution (design through confirmation) can be presented and because the basic experimental technique is also employed in the neurophysiological case, where so much less is known.

In its simplest form, the energy of an initiated seismic disturbance propagates through the Earth with a spreading wavefront. When it meets an interface between geological strata, part of the energy may be reflected back and part continue forward due to the difference in acoustic impedance at the interface. The sensors record the returning reflected energy echoes. Knowledge of subsurface velocities allows estimation of the depths and angles of inclination of the various reflectors, whereas knowledge of the locations of reflectors allows estimation of velocities. (One notes again a see-saw in the collection of knowledge.) In practice, the initiating impacts will be repeated a number of times at the same location and at points of a grid. The power of averaging is again used.

If the input signal is taken to be $X(t)$ and if $Y(t)$ denotes the corresponding output, then the two may be modeled as related, assuming linearity and time invariance, by

$$Y(t) = \int a(t-s)X(s)\,ds. \tag{12}$$

The function $a(\cdot)$ is called the impulse response, since if the Dirac delta function $\delta(t)$ is taken as input, then the resulting output is $Y(t) = a(t)$. The function $a(\cdot)$ evidences the reflectors and velocities in the earth beneath the source and receiver. The model and its interpretation may be motivated as follows. Suppose a pulse is applied at time τ. Suppose in consequence a wave is generated, travels at velocity ν_1 to a reflector at distance d_1 and a proportion α_1 is reflected back. With $X(t) = \delta(t-\tau)$, then $Y(t) = \alpha_1\delta(t - \tau - 2d_1/\nu_1)$. (This is actually the naive model for radar or sonar.) Suppose further that the transmitted portion continues downward at velocity ν_2 to a reflector at distance d_2 and a portion of its energy is reflected back, some of which is transmitted by the first reflector to reach the receiver. Now the response has the form $Y(t) = \alpha_1\delta(t - \tau - 2d_1/\nu_1) + \alpha_2\delta(t - \tau - 2d_1/\nu_1 - 2d_2/\nu_2)$. This last is seen to correspond to the system of expression (12) with impulse response $a(t) = \alpha_1\delta(t - 2d_1/\nu_1) + \alpha_2\delta(t - 2d_1/\nu_1 - 2d_2/\nu_2)$. One can clearly extend this model to situations with many layers, many velocities and many corresponding transmission and reflection coefficients. Peaks in the function $a(t)$ may be seen as corresponding to reflectors. (It must be noted that unfortunately such an elementary interpretation is likely to be complicated in practice by interfering phenomena such as ghost reflections. Some techniques have been developed to handle these.) The basics of exploration seismology are discussed in Wood and Treitel (1975), Waters (1978) and Robinson (1983).

The problem has now been formulated as one of system identification; given stretches of corresponding input X and output Y, determine an estimate of the impulse response $a(\cdot)$. In the case that a pulse close to a Dirac δ may be generated and that the function $a(\cdot)$ drops off to 0 reasonably quickly, a convenient procedure results from taking

$$X(t) = \sum_{m=1}^{M} \delta(t - m\,\Delta),$$

as input, and the "average evoked response"

$$\hat{a}(s) = \frac{1}{M} \sum_{m=1}^{M} Y(s + m\,\Delta),$$

as an estimate of $a(s)$. This input corresponds to applying pulses periodically. The estimate corresponds to stacking and averaging.

Suppose one sets $m_{YX}(t) = Y^{*}X(t)$ for some convolution operation "$*$." Then from (12) one has

$$m_{YX}(t) = \int a(t - s) m_{XX}(s)\, ds$$

and one has a deconvolution (or inverse) problem to solve. Suppose one decides to seek an $X(\cdot)$ such that

$$\int a(t - s) m_{XX}(s)\, ds \approx a(t),$$

to allow elementary processing. In terms of Fourier transforms, the left-hand side here may be expressed as $\int \exp\{i\lambda t\} A(\lambda) M_{XX}(\lambda)\, d\lambda$, with M_{XX} a Fourier transform of m_{XX}. Then what is wanted is an X such that $M_{XX}(\lambda) \approx 1$ on the support of $A(\cdot)$. If $A(\lambda)$ is known to be near 0 for $0 < \lambda < \lambda_0$ and for $\lambda > \lambda_1$, then a possible function is the "chirp" signal

$$X(t) = \cos\left(\left[\lambda_0 + (\lambda_1 - \lambda_0)\frac{t}{\tau}\right] t\right), \quad \text{for } 0 \le t \le \tau.$$

In the seismic case, the values of λ_0, λ_1 have been determined in various experiments. The chirp probe originated in radar work during World War II [see Cook and Berenfield (1967)]. It may be seen to attach near equal power to the frequencies between λ_0 and λ_1. In the seismic case special devices have been developed to input the chirp signal to the earth. The signal is input repeatedly and the results averaged. The response is then convolved with the chirp function, that is, m_{YX} is formed to estimate $a(\cdot)$. Structure can appear dramatically during the cross-correlation processing described here.

In practice, subtle further processing is employed to handle wavefront curvature, ghost reflections and other natural phenomena that may be present.

7. Other topics. There are other problems arising in seismology to which statistical methodology can be applied fruitfully. These include analysis of the coda (i.e., of the irregular trailing part of the disturbance), analysis of scattering, risk analysis, nonlinear phenomena, point process studies, polarization, cepstral analysis, discrimination of earthquakes from explosions [see, e.g., Tjøstheim

(1981)], seismicity study, travel time table construction, attenuation laws, earthquake location and azimuthal dependence of characteristics. Vere-Jones and Smith (1981) discuss several of these problems. In some cases work has begun.

8. Discussion. Seismologists have long been serious users of statistical methods. One finds Harold Jeffreys making the following statement in the entry, "Seismology, statistical methods," in the *International Dictionary of Geophysics*: "The uncertainty is as important a part of the result as the estimate itself. ...An estimate without a standard error is practically meaningless." Hudson (1981) remarks: "The success of the Jeffreys–Bullen travel time tables was due in large part to Jeffreys' consistent use of sound statistical methods." When I asked my colleague B. A. Bolt what he saw as the role of statistics in seismology, he replied: "Seismology is largely an inferential science. ...The role of statistics in seismology is to provide a rigorous procedure for turning observations on seismic waves, etc., into probabilistic statements about properties of the (real) Earth."

One may note that work in seismology is characterized by massive data sets, inherent variability and measurement error, defining/fitting/refining models, design of experiments, simulation, probabilistic description, needs for robust/resistant procedures, predictive situations, inverse problems and combination of observations. Statistics has much to offer in all these connections.

9. Update. Since the lectures were presented in 1983, work has progressed on various of the topics covered. Abrahamson (1985) has employed Smart 1 data to better see the movement of the fault rupture tip. Chiu (1986) studies the problem of estimating the parameters of a moving energy source. Lindberg (1986) develops "optimal" tapers to employ in the estimation of the frequencies of free oscillations. The approach of Kitagawa and Gersch (1985) to nonstationary data seems likely to prove of broad practical applicability. The book by Udias, Munoz and Buforn (1985) goes into substantial detail over the formal estimation of fault-plane parameters. Copas (1983) sets down an expression for the variance of a statistic like that of (9). Brillinger (1985) develops a maximum likelihood statistic for detection and estimation of a plane wave given array data. Donoho, Chambers and Larner (1986) develop a robust/resistant procedure for better aligning the seismic traces of a section. Mendel (1983, 1986) presents maximum likelihood state space-based methods for handling the data of reflection seismology. Shumway and Der (1985) indicate how the EM method may be employed to deconvolve pulses hidden in seismic traces. The nongaussianity of seismograms obtained in reflection seismology is being taken specific advantage of; see Giannakis and Mendel (1986). The techniques of Donoho (1981) and Lii and Rosenblatt (1982) seem bound to prove useful in the seismological case. The thesis, Ihaka (1985), has been completed. Ogata [e.g., Ogata (1983) and Ogata and Katsura (1986)] has carried out a variety of likelihood-based analyses of earthquake times as a point process. Many statisticians have begun working on statistical aspects of inverse problems. We specifically mention O'Sullivan (1986).

One can speculate on where the field of statistical seismology will go in the coming years. It seems clear that there will be much concern with non-Gaussian noise and signals, that vector-valued spatial–temporal data and analysis will become the norm, that large-scale conceptual models will be set down and that there will be a variety of techniques developed for borrowing strength in situations with scanty data, e.g., risk estimation.

III. Neurophysiology

> *..., modern biometry is the interdisciplinary endeavor to build structural stochastic models of biological phenomena.*
>
> J. Neyman (1974)

10. The field and its goals. Neurophysiology is the branch of science concerned with how the elements of the nervous system function and work together. The functioning is seen to involve chemical mechanisms, electrical mechanisms and physical arrangement. The studies extend from the movements of individual ions, through to the mass behavior of the components of the brain.

The goals of neurophysiologists range to the heroic: how to explain things like memory, emotion, learning, sleep, expectation, behavior. At a less ambitious level, neurophysiologists are concerned with how a single nerve cell responds to stimuli, transmits information and changes with alterations of the environment.

The neuron is both the functional and structural unit of the nervous system. The brain is a multiprocessor of dramatic complexity. The elements of the nervous system may be said to differ from those in the seismic case, in that they apparently have purposes.

The field is largely experimental with researchers collecting varied and extensive data sets. The data include photographs made via electron microscopes, fluctuating voltages and current levels within single nerve cells and finally electroencephalograms (the brain's electrical potential at points near the skull.) The studies are sometimes simply observational, but often complex experimental designs are employed.

Important techniques that are made use of include staining to identify individual neurons, insertion of microelectrodes to make measurements within individual cells and the averaging of whole suites of responses to a stimulus of interest in order to reduce what can be the dominant effects of noise. Many experiments are computer controlled and computer processed.

Discoveries made by neuroscientists include the following. Nerve cells communicate with each other in both a chemical and electrical fashion, the voltage pulse that travels along a neuron's output fiber is of near constant shape and there are a broad variety of nonlinear phenomena that occur. A number of verifiable physical laws and effective deterministic models (such as the Hodgkins–Huxley equations) have been set down. Much insight has been gained, especially at the level of small groups of neurons. At the level of the brain itself, knowledge is mainly phenomenological. Here the brain is viewed as a black box and studied by system identification techniques. Whatever the approach, discoveries have been made leading to lifesaving and life-improving clinical diagnoses.

Statistical methods entered with the quantification of the field. No single individual scientist seems to have had a dominating effect, rather there have been many contributing workers—researchers concerned with electroencephalograms (EEGs) and researchers concerned with small collections of neurons. Statistical methods entered both because of high noise levels and because a variety of phenomena seemed to be inherently stochastic. Evidence for this last is presented in Burns (1968) and Holden (1976). Pertinent books on neurophysiology include Freeman (1975), Aidley (1978) and Segundo (1984). General reviews of statistical models and methods in neurophysiology are given in Moore, Perkel and Segundo (1966) for the cases of single neurons and of small groups of neurons and by Glaser and Ruchkin (1976) for EEGs. Statistical methods for classification and pattern recognition, for handling artifacts and for data summarization are in common use.

Neurobiology is one of the most active branches of science. The physiological phenomena with which it is concerned are fundamental and in most cases barely understood.

11. Neuronal signaling. One of the important means by which nerve cells communicate is via spike trains. The inlays at the tops of the three graphs of Figure 14 give examples of spike times representative of three different sorts of neuronal behavior; pacemaker (near-periodic), bursting (activity occurs in bursts) and bursting with acceleration (of firing within bursts).

Suppose that a neuron fires at times τ_n, $n = 0, \pm 1, \pm 2, \ldots$. A convenient formal representation of its temporal behavior is provided by writing

$$Y(t) = \sum_n \delta(t - \tau_n),$$

with $\delta(\cdot)$ the Dirac delta function. This representation leads to results analogous to ordinary time-series results in many cases. In the case that the τ_n are random, one has a stochastic point process $\{\tau_n\}$. A principal descriptor of a point process is provided by its rate function. This is given by

$$\lim_h \text{Prob}\{\text{point in } (t, t + h]\}/h,$$

as h tends to 0. In the stationary case, where the stochastic properties of the process do not depend on the time origin, the rate function is constant and so only crudely useful then.

The autointensity function is an important parameter in the stationary case. It is defined as

$$\lim_h \text{Prob}\{\text{point in } (t, t + h] | \text{point at } 0\}/h,$$

as h tends to 0. It is a point process analog of the autocovariance function of time-series analysis in a general sense. This parameter may be used, for example, to describe the behavior of spontaneously firing neurons. Figure 14 presents examples for three cases. In the first case, the neuron is firing approximately periodically. The (estimate of) the autointensity is seen to oscillate (with period

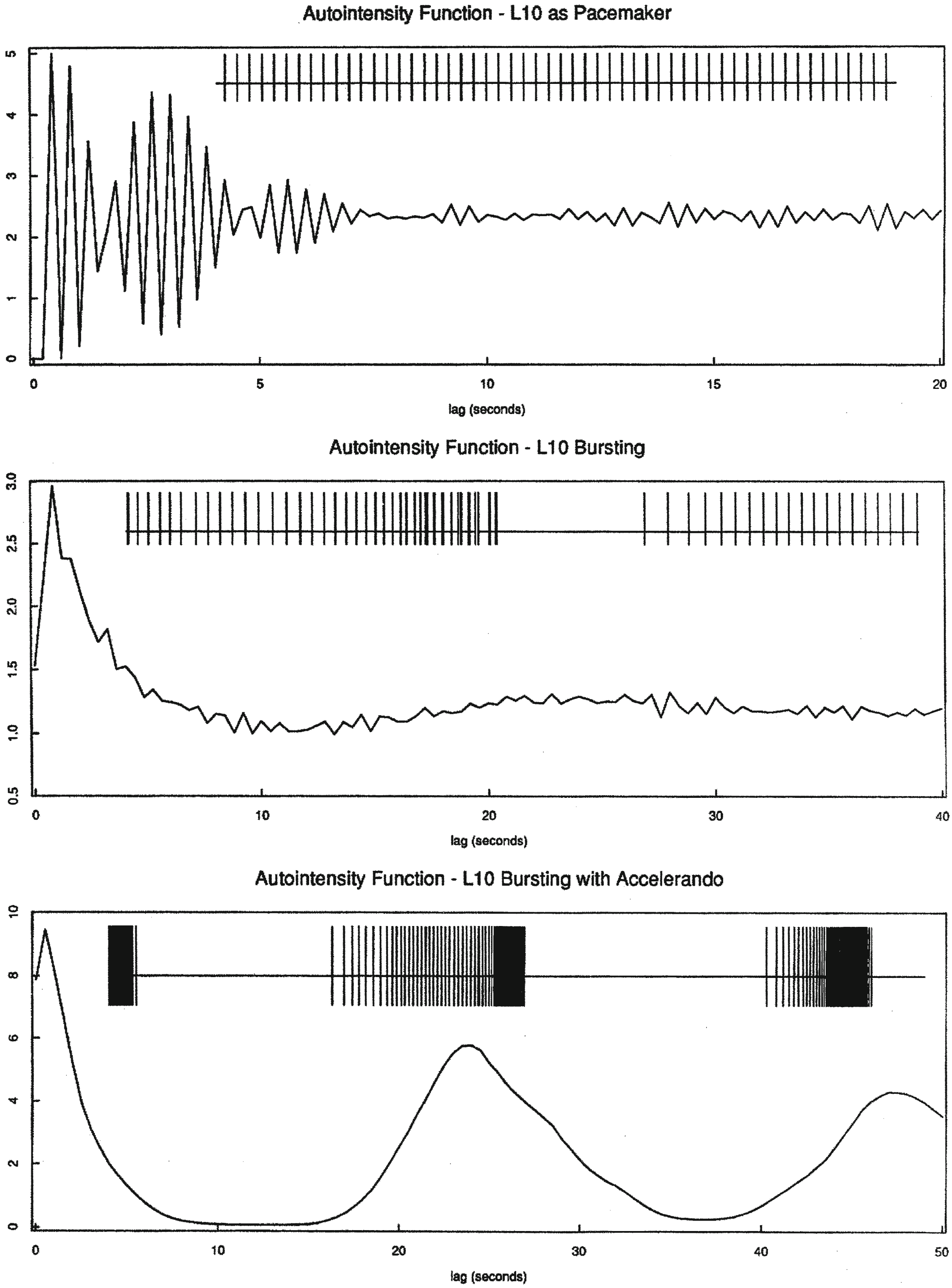


FIG. 14. *Point process data (spike train) from the nerve cell* L10 *of Aplysia californica. The cell is behaving in three different fashions. The inlays at the tops of the three graphs give brief stretches of the data (but not on the same time scales as the autointensities). The functions plotted are estimates of the autointensity functions based on* 1538, 1019, 1631 *spikes, respectively.*

equal to the interval between the points). In the second case, the neuron is evidencing activity in bursts. The probability that the neuron fires again soon after it has fired is high. In the third case, the neuron is also firing in bursts; however, now there is structure within the bursts, the rate of firing is seen to increase therein. The bursts here are at regular intervals.

The autointensity functions have been estimated, for this figure, by the statistic

$$\#\{|\tau_n - \tau_m - t| < h/2\}/Nh,$$

with N the total number of points, with h a small binwidth and with t lag. (Here # refers to the count of the number of points in the set.) The data analyzed are for the cell L10 of *Aplysia californica*, the sea hare. They were collected and previously analyzed by Bryant, Marcos and Segundo (1973). The experimental procedures and details of the data preparation may be found in that reference.

A question that arises in the study of small networks of neurons is which neurons are interacting with which? In other words, which spike trains are associated with which others? A useful parameter to employ in the study of such questions is provided by the cross-intensity function. Supposing one has spike trains named M and N, then the cross-intensity function of N given M at lag t is defined as

$$\lim_{h} \text{Prob}\{N \text{ point in } (t, t+h] | M \text{ point at } 0\}/h,$$

as h tends to 0. If the M spike train consists of points σ_m and the N train of points τ_n, then this cross-intensity may be estimated by

$$\#\{|\tau_n - \sigma_m - t| < h/2\}/Mh,$$

with M denoting the number of M points in the data set, with h a small binwidth and with t lag. Figure 15 presents three examples of estimated cross-intensity functions. The first graph refers to data from cells L3 and L10 of *Aplysia californica*. The behavior exhibited here is that of negative association; L10's firing is inhibiting the firing of L3 (for approximately 0.5 s). If one asks whether the values at negative lags differ from the level of no-association by more than sampling fluctuations, one finds they do not. This result is consistent with the cell L10 driving the cell L3. The middle graph corresponds to positive association. It is for a cell in the right visceropleural connective (RVP) and cell R15. The first cell tends to excite the second for about 0.25 s. The final graph represents a more complicated (polyphasic) situation. These data sets were also analyzed in Bryant, Marcos and Segundo (1973), where further details may be found. The approximate sampling distributions of such statistics were developed in Brillinger (1975b). It was found, for example, that it could be more convenient to graph the square root of the estimate in some circumstances.

The cross-intensity function, being a point process analog of covariance, may be expected to be an inadequate measure of relationship (as usual, correlation does not imply causation). In the case of elementary statistical data, it is usual

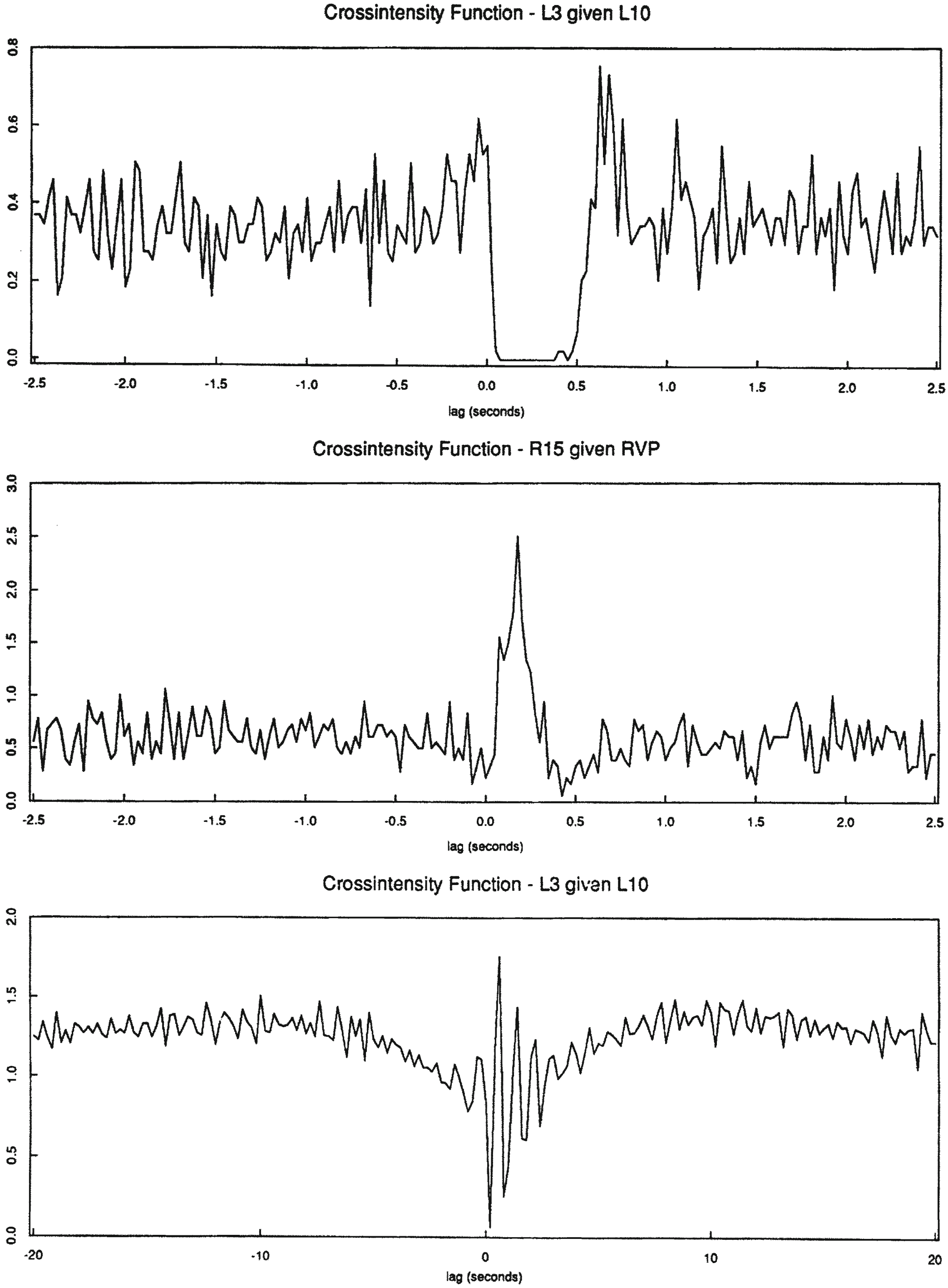


FIG. 15. *Estimates of the cross-intensity functions for three pairs of Aplysia neurons. The estimates are based on* (1746,302), (1101,288), (1019,993) *spikes in the pairs of trains, respectively.*

to turn to regression as a better technique. In the point process case it is possible to carry out regression-type analyses. For example, one may fit the following form of model:

$$\lim_h \operatorname{Prob}\{N \text{ spike in } (t, t+h) | M \text{ spike train}\}/h = \mu + \sum_m a(t - \sigma_m),$$

as h tends to 0. The function $a(t)$ appearing in this model is referred to as the impulse response. This model may be fit as follows. Set

$$d_M^T(\lambda) = \sum_{m=1}^{M} \exp\{-i\lambda\sigma_m\},$$

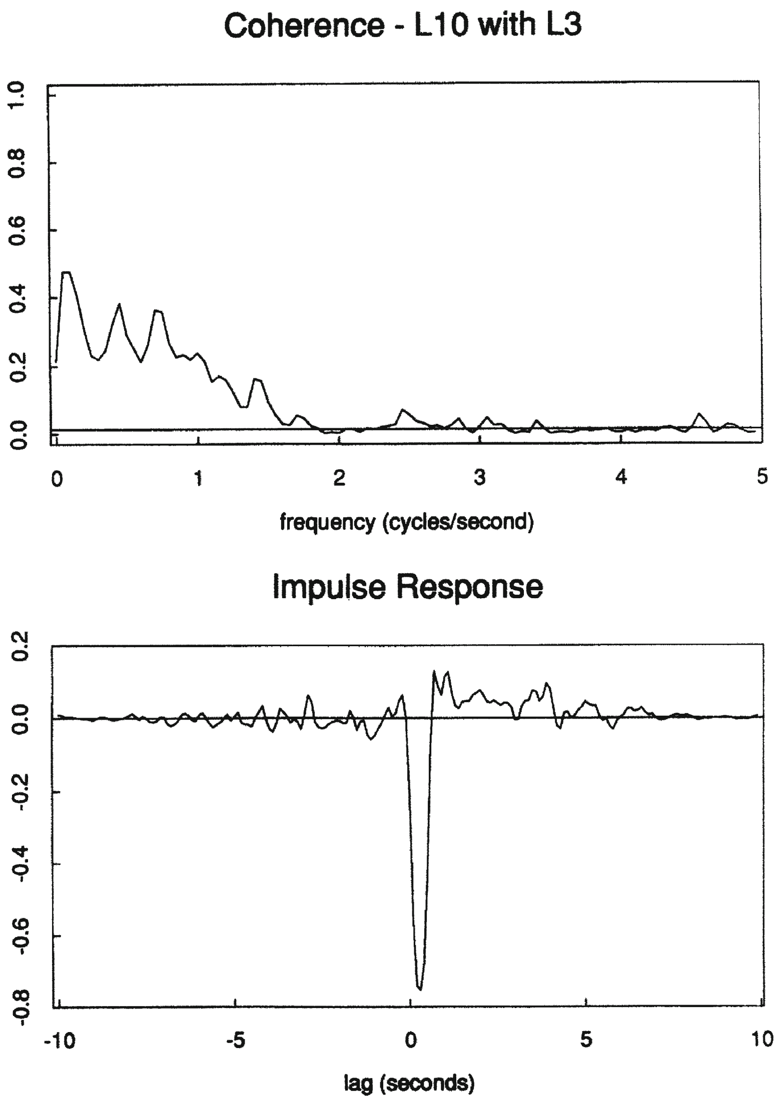


FIG. 16. *The estimated coherence and impulse response for the data of the upper graph of Figure 15. The horizontal line gives an estimate of the level exceeded by chance only 5% of the time when the spike trains are independent.*

with a similar definition for $d_N^T(\lambda)$. These are point process analogs of the empirical Fourier transform (3) of time-series data. The cross-periodogram of the given data at frequency λ is defined as

$$I_{NM}^T(\lambda) = (2\pi T)^{-1} d_N^T(\lambda) \overline{d_M^T(\lambda)}.$$

If the cross-periodogram is smoothed to obtain $f_{NM}^T(\lambda)$, then $f_{NM}^T(\lambda)$ is an estimate of the cross-spectrum in the case that $\{M, N\}$ is a bivariate stationary point process. Now $A(\lambda)$, the Fourier transform of the impulse response $a(t)$, may be estimated by $f_{NM}^T(\lambda) f_{MM}^T(\lambda)^{-1}$. The impulse response itself may be

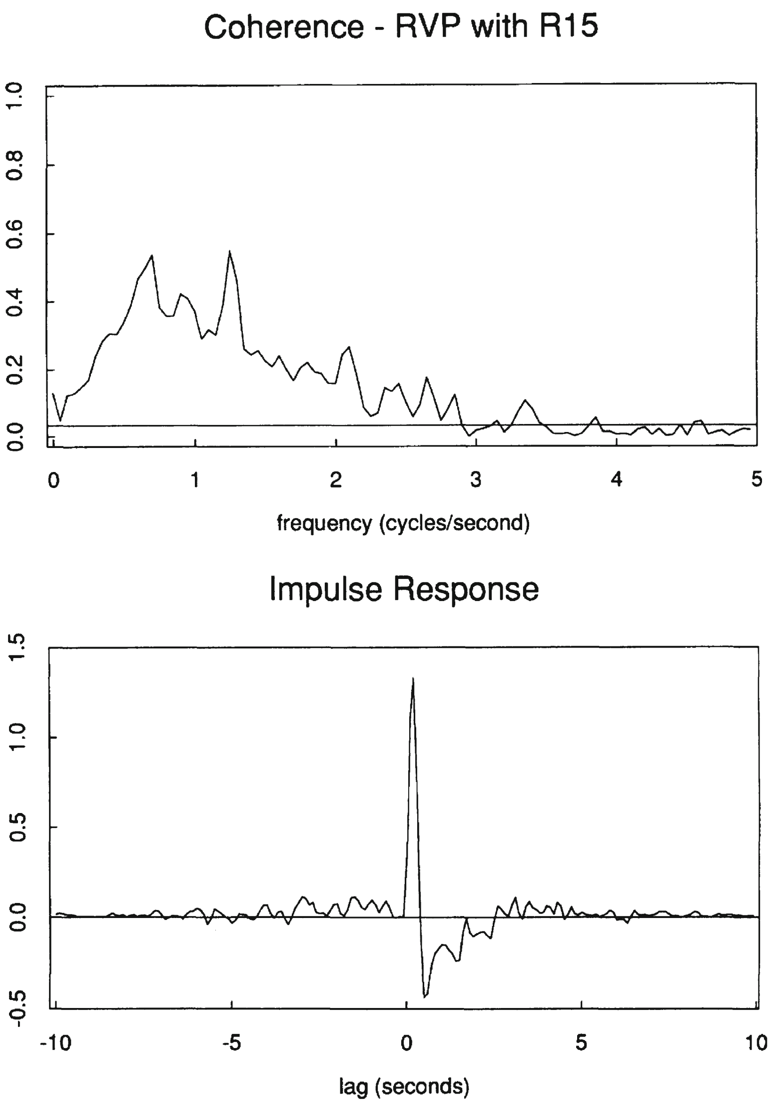


FIG. 17. *The estimated coherence and impulse response for the data of the middle graph of Figure 14. The horizontal line in the upper graph gives the approximate upper 95% null point of the distribution of the sample coherence.*

estimated by back Fourier transforming A^T. The strength of the relationship proposed in the model may be measured, at frequency λ, by the sample coherency function $R^T_{NM}(\lambda) = f^T_{NM}(\lambda)/\sqrt{f^T_{MM}(\lambda)f^T_{NN}(\lambda)}$. Its modulus squared is called the sample coherence. The coherence lies between 0 and 1, being nearer to 1 the stronger the relationship. More details of these computations may be found in Brillinger (1975b) and Brillinger, Bryant and Segundo (1976). [We here follow the use of the terms "coherency" and "coherence" in Wiener (1930).]

Figures 16 and 17 provide the results of such an analysis for the first two data sets of Figure 15. In each case the first graph is of the sample coherence. The coherences are at some distance from the value 1.0, but above the 95% null significance level (given by the horizontal lines in the figures). The relationship is inherently nonlinear, so it could have been anticipated that the coherence estimate would not be close to 1.0. Further discussion of these and similar analyses may be found in Brillinger, Bryant and Segundo (1976).

12. Assessing connectivities. Questions that can arise with small networks of neurons include; is one neuron driving the rest and if one apparently is, which one is it? The next data analysis to be presented addresses this question for three *Aplysia* cells L2, L3 and L10. From other experiments the neurophysiologists knew that cell L10 was driving cells L2 and L3. It was not known if there were any direct connections between L2 and L3. The first three graphs of Figure 18 present estimates of the three coherences, L10 with L2, L2 with L3 and L10 with L3. As might have been anticipated, these suggest relationship exists in each case.

It is possible to address the question of the direct connection of cells L2 and L3, in the presence of L10, by partial coherence analysis. Suppose that $\{A, B, C\}$ is a trivariate stationary point process. Let $R_{AB}(\lambda)$ denote the coherency function of processes A and B, with similar definitions of R_{AC} and R_{BC}. Then the partial coherency of the processes B and C, having removed the (linear time invariant) effects of process A, is given by

$$R_{BC|A} = \frac{R_{BC} - R_{BA}R_{AC}}{\sqrt{(1 - |R_{BA}|^2)(1 - |R_{CA}|^2)}}, \tag{13}$$

suppressing the dependence on λ. This definition may be motivated several ways. For example, it is the coherency between the processes resulting when their best linear predictors based on A are removed. Or, it is given by

$$\lim_{T\to\infty}\left|\operatorname{corr}\left\{d_B^T - \frac{f_{BA}}{f_{AA}}\,d_A^T,\, d_C^T - \frac{f_{CA}}{f_{AA}}\,d_A^T\right\}\right|^2.$$

Here corr denote the (complex) correlation coefficient and f_{BA}/f_{AA}, f_{CA}/f_{AA} are approximate regression coefficients. An estimate may be determined by substituting estimates for the quantities appearing on the right-hand side of expression (13). If there is no connection between the processes B and C beyond their

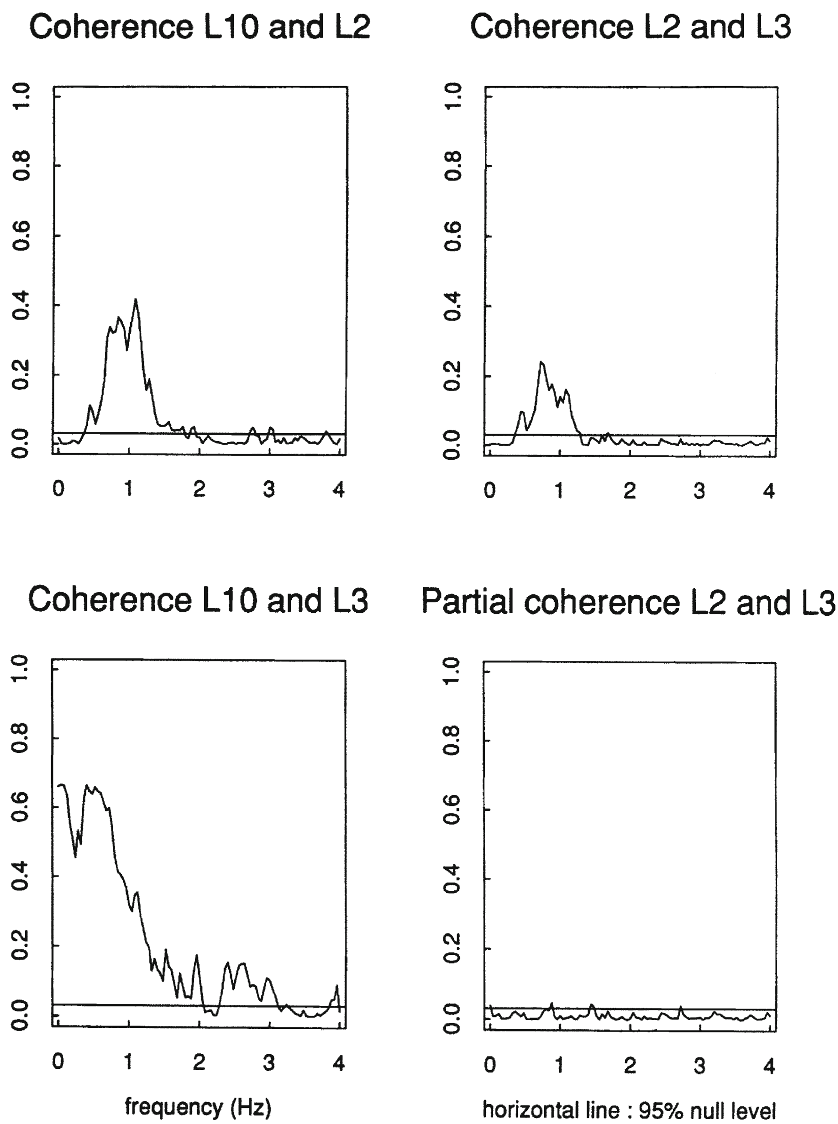


FIG. 18. *Data for a network of three Aplysia neurons. The partial coherence estimate is based on expression* (13) *of Section* 12. *In each case the horizontal solid line gives the approximate upper* 95% *null level.*

joint dependence on A, then the sample partial coherence $R^T_{BC|A}$ may be expected to be near zero.

The final graph of Figure 18 provides the results of the computation for the cells L2, L3 and L10. There is no suggestion of a direct connection being present. On reflection, it is quite astonishing the degree to which the linear models and quadratic statistics have apparently captured the dependency in a highly nonlinear situation. Further discussion and other examples of partial coherence computations may be found in Brillinger, Bryant and Segundo (1976).

13. A structural stochastic model. The analyses of neuronal firing, so far presented, are of the correlation and regression type. Parameters with direct biological interpretations have not been introduced. In Brillinger and Segundo (1979), a conceptual model is constructed and fitted by the method of maximum likelihood. The model involves the following elements.

Input to a nerve cell leads to electrical-current genesis. This current flows to a trigger zone, being filtered in the course of its passage. When the voltage level at the current zone exceeds a threshold value, the nerve cell fires. The neuron remembers back only to the time of previous firing. This process may be specified analytically as follows. Let $U(t)$ denote the voltage (membrane potential) at the trigger zone at time t. Let $B(t)$ denote the time elapsed since the neuron last fired. Let $X(t)$ denote the (measured) input to the cell. Then, assuming linearity and time invariance, one can write

$$U(t) = \int_0^{B(t)} a(s)X(t-s)\,ds,$$

for some summation function (impulse response) $a(\cdot)$. The neuron fires when the process $U(t)$ crosses a threshold level $\theta(t)$. Depending on the level at which the threshold is set and the internal mechanics of the nerve cell, the input will either accelerate (excite) or slow (inhibit) the firing. In Brillinger and Segundo (1979), this mechanism was completed and discretized as follows. Input to the cell was written X_t, $t = 0, \ldots, T-1$. Corresponding output was Y_t, $t = 0, \ldots, T-1$, with $Y_t = 1$ if there was a firing in the (small) interval immediate to t and with $Y_t = 0$ otherwise. With B_t denoting the time elapsed at t since the preceding time that $Y = 1$, they set

$$U_t = \sum_{s=0}^{B_t - 1} a_s X_{t-s}.$$

The presence of B_t in the model had the effect of introducing a form of feedback. Finally, they assumed that the threshold function had the form $\theta_t = \theta + \varepsilon_t$ with the ε's independent normals having mean 0, variance 1 and cumulative distribution function $\Phi(\cdot)$.

The likelihood function of the given data and the model then took the form

$$\prod_{t=0}^{T-1} \Phi(U_t - \theta)^{Y_t}(1 - \Phi(U_t - \theta))^{1-Y_t}.$$

Parameter estimates were determined by maximizing this likelihood with respect to θ and the a_s. Approximate standard errors were determined by procedures traditional to maximum likelihood.

Figure 19 presents the results of one such analysis. In this case fluctuating current $X(t)$ was injected directly into the cell R2 of *Aplysia*. The current level was taken to have marginal distribution that was approximately uniform (but that is not crucial to the technique). The sampling rate was 50 samples/s. The upper graph of the figure gives a stretch of the noise signal injected and the corresponding times at which the neuron fired. Details of the experiments are given in Bryant and Segundo (1976). It is very difficult, if not impossible, to see a

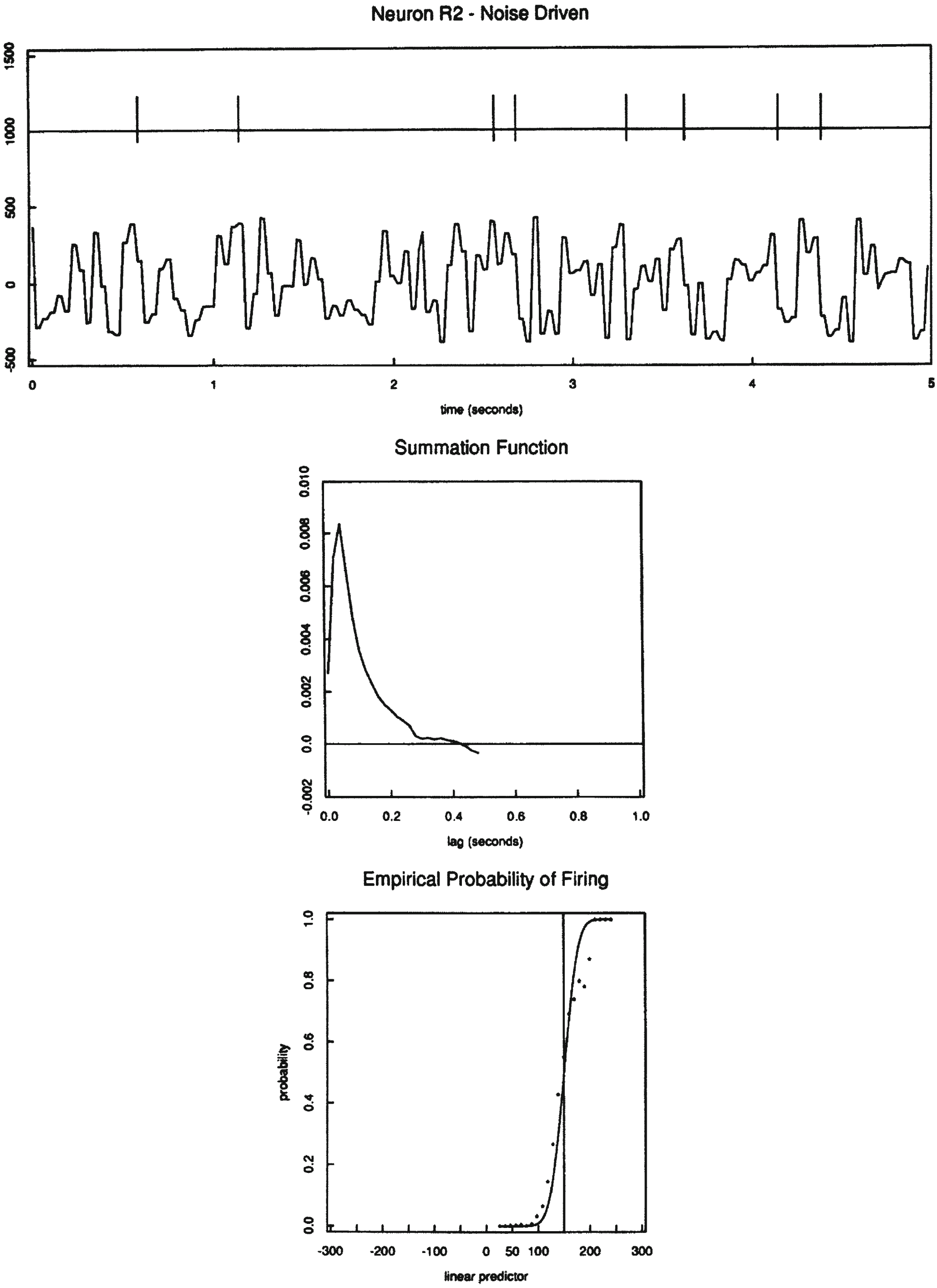


FIG. 19. *The results of fitting the neuron model of Section* 13 *to data obtained in an experiment with the cell* R2 *of Aplysia. The upper graph is a segment of the data. Noise (lower trace) is injected into the cell. The upper trace gives corresponding observed firing times. The middle graph gives the maximum likelihood estimate of the summation function* $a(\cdot)$, *estimated at* 25 *lags. The lower graph provides the statistic* (14) *of Section* 13 *and the curve* $\Phi(U - \hat{\theta})$ *with* $\hat{\theta}$ *the estimated mean threshold. The vertical line is at* $U = \hat{\theta}$.

connection between these two stretches of data. The middle graph gives the estimated summation function $\hat{a}_s$. The lower graph is one means of assessing the fit of the model. It is analogous to expression (9) of Section 3, and given by

$$(14)\quad \#\{Y_t = 1 \text{ with } U - h < \hat{U}_t < U + h\}/\#\{t \text{ with } U - h < \hat{U}_t < U + h\},$$

for small h, plotted versus U. Here

$$\hat{U}_t = \sum_{s=0}^{B_t - 1} \hat{a}_s X_{t-s}$$

is the fitted linear predictor. The smooth curve is the corresponding $\Phi(U - \hat{\theta})$. The fit may be described as adequate. The computations were carried out by a variant of the program developed for handling the seismic first-motion data of Section 3. Further examples and discussion may be found in Brillinger and Segundo (1979). Other types of input are employed and alternate estimating procedures compared there.

The large-sample properties of such estimates may be studied as in Sagalovsky (1982). A great advantage of the model-building approach, of this section, is that the parameters introduced and estimated have biological interpretations. A further advantage of the maximum likelihood approach, over that of partial coherency, is that the spike trains involved can be highly nonstationary.

14. Analysis of evoked responses. A traditional means of studying the nervous system involves applying sensory stimuli to a subject and examining the ongoing electroencephalogram for an evoked response. The stimulus may be auditory, visual (e.g., light flash, checkerboard pattern), olfactory, somatosensory (e.g., an electrical shock), gustatory or a task. Generally, the stimulus is applied for a time interval that is brief in comparison to the duration of the response. Evoked-response experiments play an essential role in quantitative biology. Because the experimenter is able to choose which stimuli to apply and when to apply them, conclusions can pass beyond associations noted, to formal inferences concerning causal mechanisms. These experiments are formally the same as the seismological reflection experiments described in Section 6.

Some dramatic success stories of the technique may be mentioned. One is presented in Bergamini, Bergamasco, Fra, Gandiglio and Mutani (1967). Siamese twins were joined in such a way that it was not possible to determine by traditional means if the peripheral nervous pathways were interconnected. Before operating, it was crucial to determine the interconnections of the twins. Ongoing EEGs were recorded for each. A series of trials were carried out in which each of the twins' legs was stimulated in turn by electrical shocks. What was found was that when a leg of one twin was stimulated, response was noted only in her EEG. On the basis of this information, the twins were separated—successfully. A second notable example of the use of the evoked response technique is provided by hearing exams for newborn infants (including infants asleep.) EEGs are recorded. These are examined for responses after loud clicks are made near the infants' ears. Rapin and Graziani (1967) present an example for an infant with hearing difficulties, both wearing and not wearing a hearing aid. The hearing aid is found to have an objectively measurable effect.

Figure 20 presents an example of evoked-response data recorded at a 4×4 array of sensors implanted in a rabbit. In this case the stimulus was an odor and the sensors were implanted in order to study the rabbit's olfactory system. These responses were recorded concurrently. A second example is given in Figure 21. It gives the 20 successive responses evoked by applying a current pulse to the lateral olfactory tract of a rabbit and recording from a sensor implanted in the depth of the pre-piriform cortex. The signal is fairly pronounced in Figure 20. In Figure 21 the strength of the stimulus was weak and the signal is not apparent. Both of these data sets were collected in the laboratory of W. J. Freeman, University of California, Berkeley. Some details of his experiments may be found in Freeman and Schneider (1982).

Crucial to many evoked response experiments is the fact that it is generally insufficient to apply a stimulus once. Rather it must be applied repeatedly (perhaps thousands of times) and the responses averaged. (This is also true in the case of reflection seismology as mentioned earlier.) In the twins and infant examples discussed previously, M equaled 250 and 100, respectively. Formally, if $Y(t)$ denotes the measured EEG and the stimulus is applied at times σ_m, $m = 1, \dots, M$, then it is usual to take as the basic statistic, the average evoked response

$$\bar{Y}(s) = \frac{1}{M} \sum_{m=1}^{M} Y(s + \sigma_m).$$

The left-hand column in Figure 22 presents the results of averaging the data of Figure 21 with $M = 3, 5, 10, 20$ and 38. With increasing averaging a signal is slowly appearing from the noise. Some alternate evidence for the presence of a signal is provided by the results of the right-hand column. These are averages of 38 responses, where the stimulus has been applied at a succession of increasing strengths.

A variety of questions, which have statistical formulations, arise in the course of work with evoked responses. (1) Does an applied stimulus elicit a response? (2) Do two different stimuli elicit the same response? (3) Is the same response elicited at two different sensor locations? (4) Is the response stationary? (5) If the order of stimuli application is altered, are the corresponding responses altered? (6) Are the effects of different stimuli additive? (7) How does the response depend on the stimulus intensity? (8) How do the responses depend on exogenous variables? To go with answers to these questions, researchers seek quick efficient data collection, precise estimates and indications of variability. Difficulties that commonly arise include small response, large noise, variability in response, artifacts present and superposed effects. Next in this section, two formal set-ups will be presented that may be employed to address the situation.

Suppose, to begin, that there is a single stimulus and that it is applied at times σ_m. Let $a(\cdot)$ denote the response in a single-shock experiment. If the system is time invariant and the effects of the various shocks additive

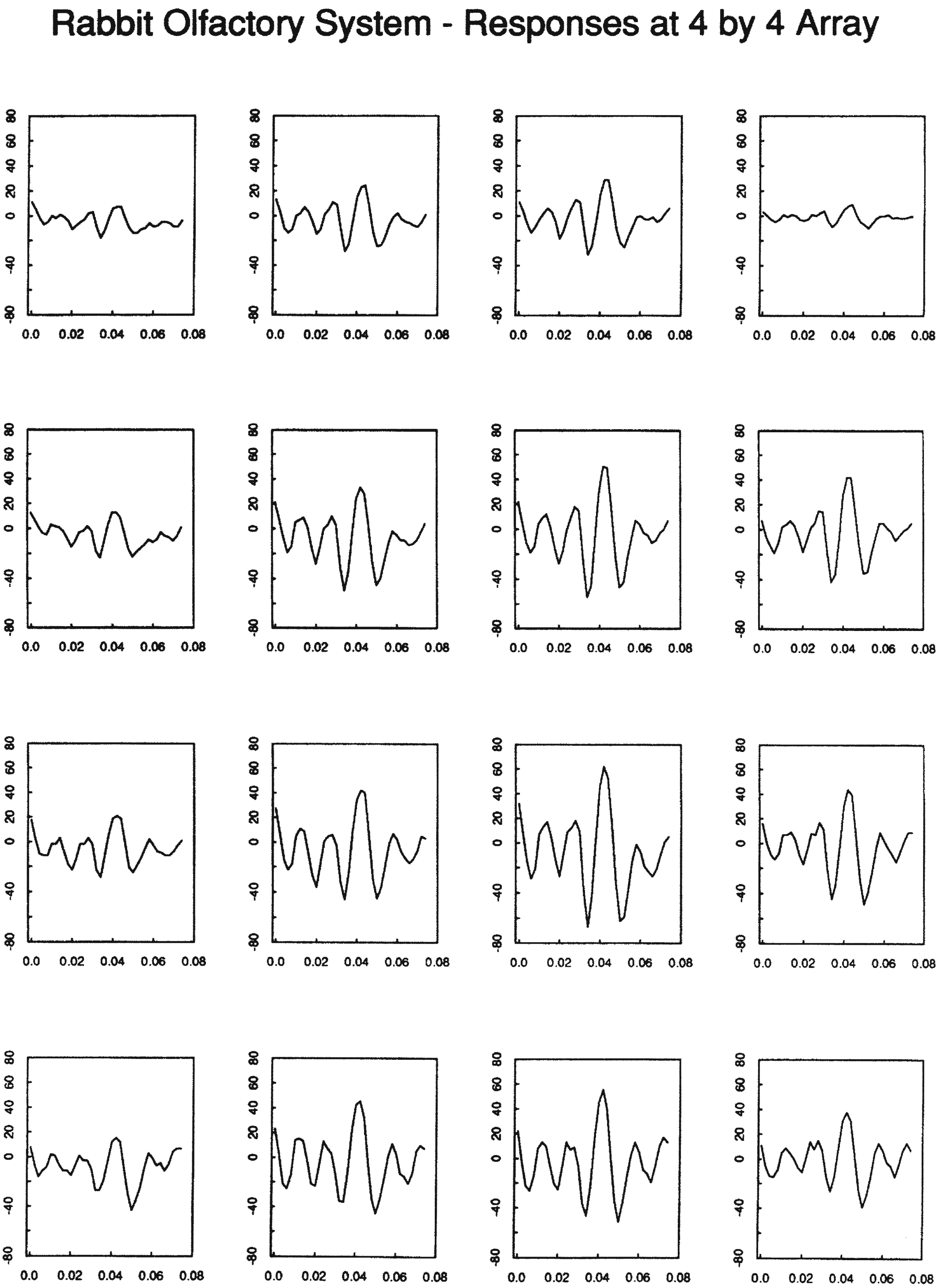


FIG. 20. *The bursts of electroencephalographic activity recorded at the* 16 *sensors of a* 4×4 *array implanted in a rabbit, following the stimulation of the rabbit by an odor. The units of the x-axis are in seconds.*

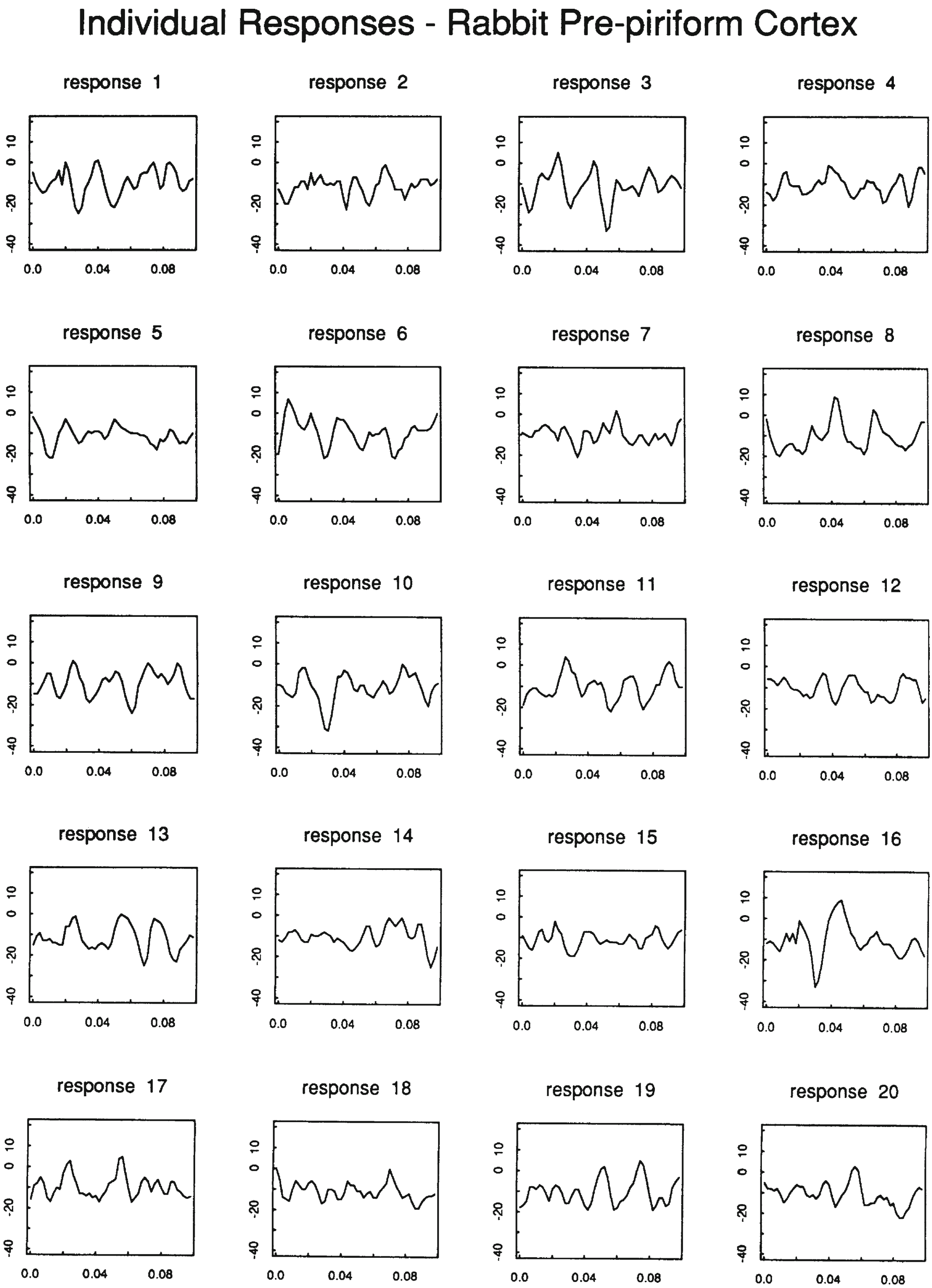


FIG. 21. *Twenty successive responses evoked in the pre-piriform cortex by (electrically) stimulating a rabbit. The x-axis units are in seconds.*

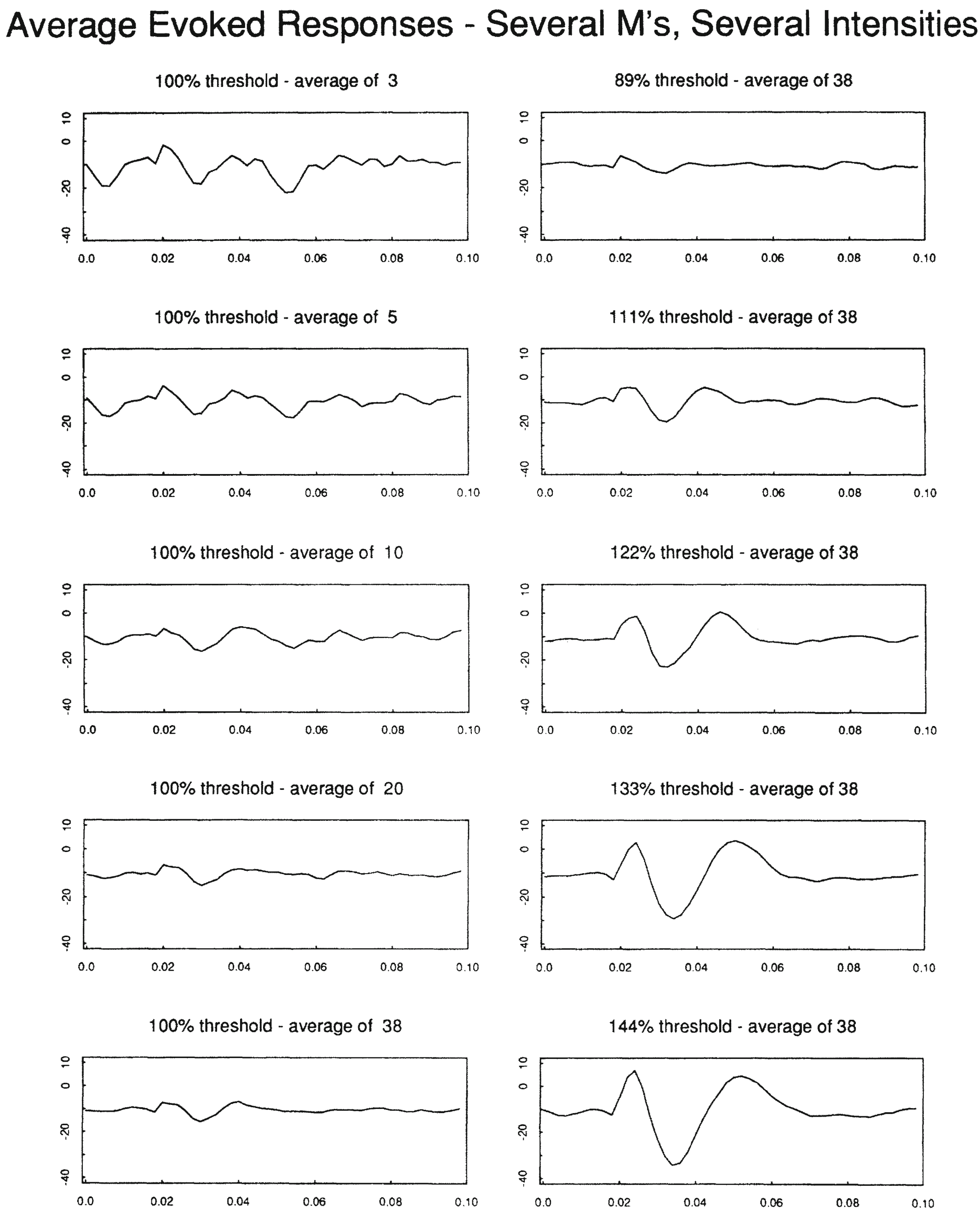


FIG. 22. *The various graphs here are meant to show the effects of changing the number of responses averaged (left column) and the strength of stimulus applied (right column) for data such as that of Figure* 21.

(superposable), then a model for consideration is

$$Y(t) = \mu + \sum_m a(t - \sigma_m) + \varepsilon(t), \tag{15}$$

with $Y(\cdot)$ denoting the ongoing EEG and $\varepsilon(\cdot)$ denoting noise. In the case of the EEG this model seems to have to be empirically verified, rather than being an implication of basic biology. (In the seismological case it came out of a conceptual framework.) For example, the assumption of superposability may be examined as follows for the animal studied. To begin, carry out some single-shock experiments, i.e., apply the shocks at times far enough apart that their individual effects seem likely to have died off. Let $\hat{a}(s)$ denote the average of the responses evoked, with s lag since stimulus application. Now carry out some two-shock experiments, i.e., apply shocks say Δ time units apart. Let $\hat{b}(s, \Delta)$ denote the average of the responses evoked. To examine the assumption of superposability compare $\hat{a}(s) + \hat{a}(s - \Delta)$ with $\hat{b}(s, \Delta)$. The results of carrying out such a check, in an experimental situation, are given in Biedenbach and Freeman (1965). They form averages of $M = 150$ responses and do not note departure from superposability.

We now turn to one formal analysis of the model (15). If one writes

$$X(t) = \sum_m \delta(t - \sigma_m),$$

then (15) takes the form

$$Y(t) = \mu + \int a(t - s)X(s)\,ds + \varepsilon(t),$$

i.e., it is seen to be the model of cross-spectral analysis. Taking Fourier transforms, one has

$$d_Y^T(\lambda) \approx A(\lambda)\,d_X^T(\lambda) + d_\varepsilon^T(\lambda),$$

for $\lambda > 0$, with $A(\lambda)$ denoting the Fourier transform of $a(\cdot)$. Consider a number of frequencies $\lambda_k = 2\pi k/T$ near λ. Then, assuming $A(\cdot)$ smooth, one has the approximate linear model

$$Y_k \approx A(\lambda)X_k + E_k,$$

with

$$Y_k = \sum_{t=0}^{T-1} Y(t)\exp\left\{-i\frac{2\pi kt}{T}\right\},$$

and similar definitions of X_k, E_k. Next, via a central limit theorem for empirical Fourier transforms, the noise variates E_k may be approximated by independent (complex) normals having mean 0 and variance $2\pi T f_{\varepsilon\varepsilon}(\lambda)$. All the inference procedures for the linear model become available. For example, as an estimate of the transfer function $A(\lambda)$, one has

$$\hat{A}(\lambda) = \sum_k Y_k \overline{X}_k \Big/ \sum_k X_k \overline{X}_k$$

and this variate will be approximately distributed as complex normal with mean $A(\lambda)$ and variance $2\pi T f_{\varepsilon\varepsilon}(\lambda)/\Sigma_k |X_k|^2$. This formulation has a variety of convenient properties. It directly extends to the cases of multiple stimuli and multiple responses. It handles stimuli of varying intensity. It allows the individual responses of the separate shocks to overlap. Formal inference procedures, such as tests, are available. Complex experiments may be designed and analyzed—complexities handled such as blocking, rotation, factorial treatment structure, measured covariates. Formal checks for interaction are available. Finally, one can turn to the question of optimal design.

It is sometimes convenient to adopt a different viewpoint for the problem. Suppose that the shocks are applied at times such that $\sigma_{m+1} - \sigma_m > V$ with $a(s) = 0$ for $s > V$ and $s < 0$. Write

$$Y_m(s) = Y(s + \sigma_m).$$

Then $Y_m(s) = \mu + a(s) + \varepsilon_m(s)$ for $0 \leq s \leq V$. The average evoked response is now conveniently denoted $\overline{Y}(s)$. As an example of the use of this formulation, suppose there are I different stimuli and that each are applied J times, then one is led to set down the model

$$Y_{ij}(s) = \mu_{ij} + a(s) + b_i(s) + \varepsilon_{ij}(s),$$

with i indexing stimuli and j indexing replicates. Other methodologies, such as grown curves and discriminant analysis, are seen to become available with this formulation.

It was mentioned that evoked-response data may be contaminated by artifacts. It is perhaps worth noting that robust/resistant estimates are directly available. Suppose one has a measure of distance, such as

$$\|Y - a\|^2 = \int_0^V [Y(s) - a(s)]^2 \, ds$$

and an estimate of scale $\hat{\rho}$. Then a family of robust/resistant estimates is provided by

$$\hat{a}(s) = \sum_m W_m Y_m(s) \Big/ \sum_m W_m,$$

with $W_m = W(\|Y_m - \hat{a}\|/\hat{\rho})$ and $W(\cdot)$ a univariate set of multipliers for robust/resistance. The estimate will need to be computed recursively. An elementary example is provided by the "trimmed mean"

$$\hat{a}(s) = \sum{}' Y_m(s)/\beta M,$$

with Σ' over the βM smallest $\|Y_m - \hat{a}\|$. This class of estimates was proposed in Brillinger (1979, 1981a) and investigated in Folledo (1983). The upper graph of Figure 23 provides an example of this estimate with 50% trimming ($\beta = 0.5$), in the case of data like that of Figure 21 (but with a stimulus of strength 122% of the threshold stimulus). The solid curve denotes the average evoked response, the dashed one the trimmed statistic. The two curves are nearly identical, although when examined, the individual responses are found to differ noticeably. Fifty-percent trimming was employed, because this is usually considered a highly

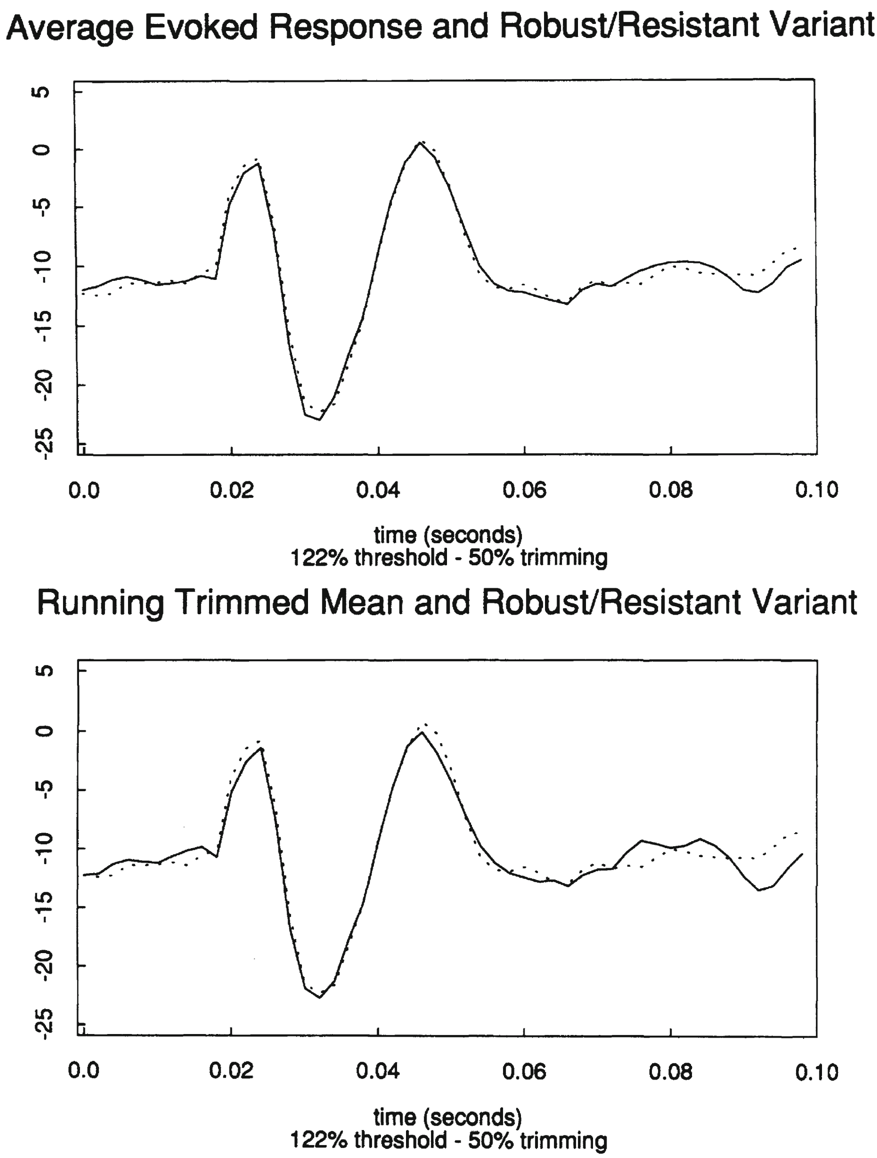


FIG. 23. *The upper graph compares the average evoked response with the 50% trimmed mean for the data taken at 122% of a threshold stimulation value. The lower graph contrasts the 50% trimmed mean statistic with a value computed recursively.*

resistant level in the case of elementary statistics. The fact that the trimming had such little effect on the final answer suggests that there were no substantial outlying curves in the data set. Had a curve been far removed from the rest, then it would have been rejected from the average. It is to be remarked that in this case of present concern, whole curves are being eliminated from the average, not just outlying points that some curves might have.

It is to be noted that a "real-time" version of such a trimmed mean may be computed; see Brillinger (1981a). This statistic is given recursively for $m =$

1, 2, ... by

$$\hat{\rho}_{m+1} = \hat{\rho}_m - \frac{L}{m}\left(\frac{1}{\beta} - 1\right),$$

if $\|Y_{m+1} - \hat{a}_m\| \le \hat{\rho}_m$, and

$$\hat{\rho}_{m+1} = \hat{\rho}_m + L/m$$

otherwise, and by

$$\hat{a}_{m+1}(s) = \hat{a}_m(s) + \frac{1}{\beta m}(Y_{m+1}(s) - \hat{a}_m(s)),$$

if $\|Y_{m+1} - \hat{a}_m\| \le \hat{\rho}_m$, and

$$\hat{a}_{m+1}(s) = \hat{a}_m(s)$$

otherwise. (In preparing a worked example, it was found more convenient in the choice of L to replace ρ by its logarithm.) The lower graph of Figure 23 gives the result for the same data as that of the upper graph. The algorithm was run setting $\hat{a}_1(s) = Y_1(s)$ and $L = 0.15$. The real-time estimate, given by the solid line, has performed virtually as well as the dead-time estimate in this case. One can remark again that had there been some highly dissimilar curves present, then this estimate would have differed from the sample average. Following the advice sometimes given in connection with resistant regression estimates, it would seem sensible to compute both the ordinary and the resistant forms. If the two are similar, then there is no difficulty. If the two differ noticeably, then the situation should be examined in some detail.

Brillinger (1979) proposed the preceding techniques and various others. Brillinger (1981a, b) were based on that lecture and cover some other statistical problems arising from evoked-response methods. Tukey (1978) also addresses statistical issues and proposes some procedures.

15. A confirmed (Fourier) inference. Muscle cells are electrochemical devices. If the chemical acetylcholine is applied at the neuromuscular junction, measurable voltage fluctuations result. Specifically, acetylcholine release causes postsynaptic membrane channels to open leading to voltage fluctuations. Katz and Miledi (1971, 1972) measured voltage fluctuations associated with this phenomenon and found that the power spectrum could be approximated by the functional form $\alpha/(\beta^2 + \lambda^2)$. [An example of the fit of this function to such data and a description of a fitting procedure may be found in Bevan, Kullberg and Rice (1979).] They proposed the model

$$Y(t) = \sum_m a(t - \sigma_m),$$

with the σ_m points of a Poisson process and with $a(t) = \exp\{-\beta t\}$. This $a(\cdot)$ function corresponds to the effectiveness of an open channel decaying exponentially and leads to a power spectrum of the indicated form. Katz and Miledi mentioned that the pulses might actually be rectangular of random duration, but they preferred to deal with the exponential form. Stevens (1972) proposed the specific model

$$Y(t) = \sum_m a_m(t - \sigma_m),$$

also with $\{\sigma_m\}$ Poisson, but now with $a_m(t) = 1$ for $0 < t < T_m$ and $a_m(t) = 0$ otherwise. The T_m are independent exponentials of mean $1/\beta$ and correspond to the lengths of time that the channels are open. Stevens noted that this model also led to a power spectrum of the form $\alpha/(\beta^2 + \lambda^2)$. The models were indistinguishable with the data collected.

The problem was later resolved by improved experimental technique. Neher and Sakman (1976) developed a technique that allowed the opening and closings of individual channels to be seen. They found that the channels remained equally effective and open for time periods of varying lengths. The two proposed models could be distinguished.

Examples of single-channel data and the corresponding estimated power spectrum may be found in Lecar (1981). Jackson and Lecar (1979) present results confirming the exponential duration of the openings.

16. Other topics. Spatial–temporal data are commonly collected by neuroscientists. One form is the electroencephalogram recorded by an array of sensors on the scalp. Figure 19 presented an example of data collected for the olfactory system of the rabbit. The stimulus was release of the odor ethylacetate. An 8×8 array of electrodes was imbedded in the animal. The data, already presented in Figure 19, give the responses for the sensors at the positions with x-coordinates 2, 4, 6 and 8 and y-coordinates 1, 3, 5 and 7 of Figure 23. One procedure that Freeman has found helpful for understanding this type of data is the computing of empirical orthogonal functions; see Freeman (1980). Figure 24 gives an example. These results are derived by stacking the responses into a matrix $\mathbf{X}$ with rows corresponding to sensor and columns to time, and then computing the singular value decomposition $\mathbf{X} = \mathbf{UDV}^{\tau}$, of that matrix. The $\mathbf{U}$ for a particular component, say the first, are then plotted versus sensor location as in the upper graph of Figure 23. The $\mathbf{V}$ values are similarly plotted versus time and appear in the lower graph. The results of Figure 23 are based on 64 series, not just the 16 of Figure 19. The contour plot suggests the presence of a focus of activity. The time-series component elicited may be seen lurking in the individual responses of Figure 19. (It may be mentioned that meteorologists have long computed empirical orthogonal functions for spatial–temporal data and used them in forecasting; see Lorentz (1956), for example. A number of other references are given in Jolliffe (1986).]

Childers has also made use of array data in studying the neural system. In Childers (1977), he estimates the frequency–wavenumber spectrum for responses evoked by visual stimuli (light flashes) in the human EEG. He was concerned with estimating the speed and direction of propagating waves. In the paper cited he first notes an apparent high-velocity wave. After this wave has been "removed," he notes the presence of a pair of waves moving in opposite directions. His research is directed at developing a diagnostic procedure for various visual disorders and obtaining insight concerning how the visual system functions.

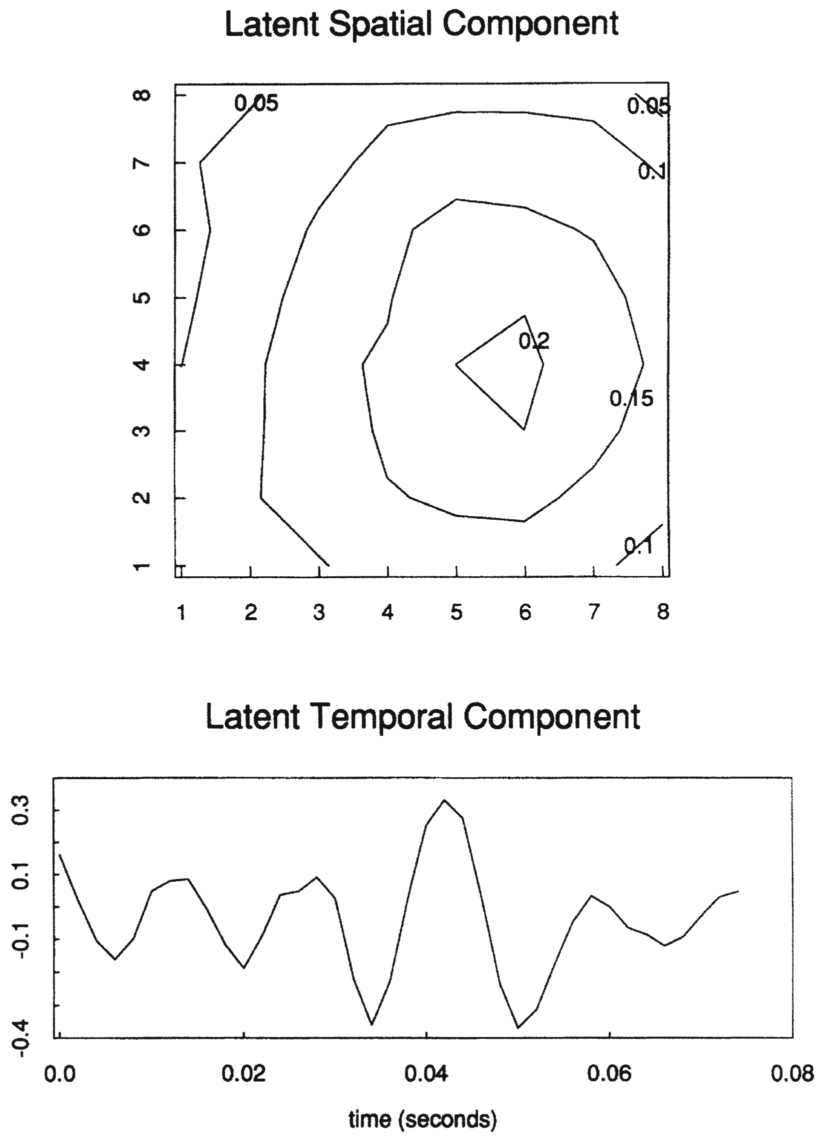


FIG. 24. *The results of a singular value decomposition of the full set of the data from which the bursts of Figure 20 were taken. The values graphed are for the first components. The axes in the upper graph give spatial location.*

The decaying cosine model of Section 2 has also found a use in neurophysiology. In his work with the olfactory system, Freeman (1972, 1975, 1979) found that the average evoked response could be well fitted by the sum of a few decaying cosine terms. He developed a model involving spike-to-wave conversion, involving collections of constant coefficient second-order differential equations, involving feedforward and feedback and involving wave-to-spike conversion. He employed nonlinear regression in the time domain to estimate the unknowns. In one case, involving two cosines, he was led to view the stronger wave as representing intracortical negative feedback and the weaker as representing a second feedback loop. Of interest in this type of work is what happens to the frequencies and the decay rates when the experimental conditions are altered. A second reference to decaying cosines is Childers and Pao (1972). They consider the model

$$Y(t) = \sum_k \alpha_k t \exp\{-\beta_k t\} \cos(\gamma_k t + \delta_k) + \varepsilon(t), \qquad t > 0,$$

for visual evoked responses monitored over the occipital region. In particular, they study the data by complex demodulation.

Brief reference will be made to several other topics. Dumermuth, Huber, Kleiner and Gasser (1971) estimate the bispectrum of human EEGs. de Weerd and Kap (1981) discuss the computation of some time-varying quantities. Marmarelis and Naka (1974) consider the case of biological systems with several inputs. An extreme case of this occurs when the input is varying in both time and space. This circumstance is considered in Yasui, Davis and Naka (1979). The book by Marmarelis and Marmarelis (1978) goes into great detail concerning the identification of systems that are polynomial and time invariant in the input. They emphasize the advantages resulting from employing a Gaussian white-noise input. The dedication of the book is worth mentioning—"To an ambitious new breed: SYSTEMS PHYSIOLOGISTS".

Another area of research activity has been that of control. The works by Poggio and Reichardt (1981) and Wehrhahn, Poggio and Bulthoff (1982) may be noted. They are concerned with data that are three-dimensional trajectories.

17. Discussion. As the examples presented indicate, a broad range of data types arise in the neurosciences. Furthermore, data are collected at both the micro and macro level. The procedures developed often have the opportunity to move on to direct clinical use.

It is particularly interesting to note the evolution of the analysis in the case of the neuronal signaling analysis as presented in Sections 11 and 13. One can recognize the stages of (1) (feature) description; (2) correlation/association; (3) (ad hoc) regression; (4) conceptual model. These stages are usual in many elementary situations.

The field of neurophysiology has the satisfying aspect that in many cases controlled laboratory experiments are possible and repeatable. Furthermore, there are opportunities for the design of experiments. In the field, statistics has been seen to provide techniques for model formation and validation, for measuring uncertainty in conclusions and for addressing questions of causality. Statistical techniques have led to insight concerning the underlying physiology. In this connection it seems important to note the following proviso of my collaborator J. P. Segundo, however, ". . . . The maxim of all of the above is that the power of available mathematics (and of the instrumentation that implements them) should be used exhaustively, guided by an unflagging biological realism, mistrustful and stubborn, and keeping in mind that the ultimate goal is understanding in strictly biological terms." [See Segundo (1984), page 294.]

It seems likely that in the neurosciences, more often than not, notable advances will come from the carrying out of novel experiments, rather than from novel analytic methods. Experiments will be carried out measuring things at new orders of smallness. More complex stimuli will be invoked. Nonlinear systems will be the norm. Neural networks will be a major concern. Luckily, for us statisticians, digital computers have become common in the laboratory and this seems to be bringing a move toward quantization of other aspects of the work beyond the simple recording of the data.

18. Update. The analysis of single ion-channel data, briefly referred to in Section 15, has become a whole industry. Models with several states are now routinely fitted. References include Coloquhoun and Hawkes (1983), Labarca, Rice, Fredkin and Montal (1986) and Milne, Edeson and Madsen (1986). Extending the work of Section 13, Brillinger (1986) presents a number of examples of the maximum likelihood fitting of a neural model employing corresponding spike train input and output data. Smith and Chen (1986) study a more complicated neural model. The chirp signal was propounded as being of substantial importance in seismic exploration. Some use of it has been made recently in physiological studies. In Norcia and Tyler (1985), a 10-s spatial frequency sweep stimulus is employed and the corresponding visual evoked potential measured. Th. Gasser and collaborators have now carried out a substantial number of statistical and substantive analyses of evoked responses. We mention in particular the papers by Mocks, Tuan and Gasser (1984), Gasser, Mocks, Kohler and de Weerd (1986) and Gasser, Mocks and Kohler (1986). Finally, we note that Grajski, Breiman, di Prisco and Freeman (1986) apply modern classification procedures to study the effects of applying different odors on the olfactory bulb EEGs of rabbits and that Gevins, Morgan, Bressler, Cutillo, White, Illes, Greer, Doyle and Zeitlin (1987) relate human performance accuracy to brain electrical patterns just before a task.

IV. Concluding remarks. In this article we have presented a number of examples, drawn mainly from our personal experience, showing the use of the same statistical technique in the rather separate sciences of seismology and neurophysiology. It now seems appropriate to ask what, if anything, have the *three* sciences—statistics, seismology, neurophysiology—gained from each other as a result of connections even though they are indirect? Having in mind a broader class of examples than those discussed in this paper, one can say that: (i) statistics is richer for having been led to develop and study various novel methods to handle specific problems arising in seismology or neurophysiology; (ii) both seismology and neurophysiology are the richer for the other's field having generated a problem for the statistician to abstract sufficiently that the result's applicability to their field became apparent; (iii) either seismology or neurophysiology benefit from a statistical formulation because various of their problems seem necessarily to need to be stated in terms of probabilities (e.g., neither neuron firings nor earthquakes seem deterministic) and because these fields need procedures to validate results and to fit conceptual models. That the methods of statistics can lead to important insight and understanding in substantive problems seems agreed.

It may be remarked that the applicability of statistical procedures to these two substantive fields has further grown in direct consequence of their move to greater quantification and digital data collection. The data sets analyzed were of high quality. The fact that the analyses were informative to an extent here bodes well for the use of such techniques in fields with data of lesser quality. I need to remark how crucial, in working with the data sets discussed, I have found it to be to plot the data in its original form. Something special seemed to be learned in each case from doing so. This is why for the various analyses, I have sought to provide data plots as parts of the presentation.

The reader will have noted that some of the analyses were time-side and some were frequency-side. Each domain has its advantages. It seems worth pointing out specifically that stationarity was not required for some of the frequency-side procedures. It would seem that most time series and point process situations would benefit from carrying out simultaneous time-side and frequency-side analyses.

On review it may be seen that the techniques employed for time-series data and for point process data in many cases are not that different. Brillinger (1978) presents some comparative discussion of the techniques for the two cases. Our presentation is somewhat remiss in the seismological case in not presenting some worked examples of auto- and cross-intensity estimation. Examples could have been provided.

It should be apparent that major data management and computational efforts were required in the derivation of all results presented. I have been impressed by the way that the neuroscientists could turn to their lab book kept during the experiments and pull out crucial details, sometimes many years after the experiments had been completed. My analyses also extend over many years now. I have found it very useful recently to maintain a "Readme" file in the various computer directories for the data sets, wherein I list what the various programs do, future wishes concerning the programs and all of the things that I think I will never forget.

Turning to thoughts concerning developments to come, it seems that the future will see many of the traditional statistical techniques extended to apply to datum of more complicated forms—specifically, to curves, moving surfaces, point clouds and the like. It seems that techniques developed in one field will continue to be transferred (by statisticians?) to other fields. For example, I expect to see the results developed by neuroscientists for arrays on a curved surface (the skull) to be taken up by the seismologists as they need to take specific note of the Earth's curvature. If I have found anything lacking in our current toolkit of statistical methods and devices, it is a collection of techniques that suggest what to do next when a model fails a validation check. Perhaps the future will see such techniques developed in an organized fashion.

Acknowledgments. One wonders if it is possible to work meaningfully on problems from substantive areas unless one is in close contact with researchers of those areas. I doubt it. I can say that these lectures would never have come about, but for the help and encouragement that my collaborators B. A. Bolt and

J. P. Segundo have provided through the years. I cannot thank them enough. The influence of my doctoral supervisor, J. W. Tukey, should be readily apparent as well. In particular, I remember his saying, when I was a student, "One cannot consult with a chemist, unless one becomes a chemist." This has certainly been my own experience with respect to the parts of neurophysiology and seismology on which I have worked. The reverse has also proved to be the case, namely, collaborators from those fields have had to learn parts of statistics as well as I knew them. I also must thank my students who have worked with me on various of the problems described. In particular, I mention M. Folledo and R. Ihaka, whose theses have been detailed.

Finally, I would like to thank J. Rice and R. Zucker for providing some neurophysiological references. I would like to thank N. Abrahamson and R. Daragh for providing the Smart 1 data sets. I would like to thank W. J. Freeman and K. Grajski for providing the rabbit olfactory system data. I would like to thank Finbarr O'Sullivan, the Associate Editor, and the referees for their helpful suggestions for improvement of the initial draft paper.

REFERENCES

ABRAHAMSON, N. A. (1985). Estimation of seismic wave coherency and rupture velocity using the Smart 1 motion array recordings. Ph.D. thesis, Univ. California, Berkeley.

AIDLEY, D. J. (1978). *The Physiology of Excitable Cells*. Cambridge Univ. Press, Cambridge.

AKI, K. and RICHARDS, P. G. (1980). *Quantitative Seismology*. Freeman, San Francisco.

BERGAMINI, L., BERGAMASCO, B., FRA, L., GANDIGLIO, G. and MUTANI, R. (1967). Diagnostic recording of somato-sensory cortical evoked potentials in man. *Electroenceph. Clin. Neurophysiol.* **22** 260–262.

BEVAN, S., KULLBERG, R. and RICE, J. (1979). An analysis of cell membrane noise. *Ann. Statist.* **7** 237–257.

BIEDENBACH, M. A. and FREEMAN, W. J. (1965). Linear domain of potentials for the prepyriform cortex with respect to stimulus parameters. *Experimental Neurology* **11** 400–417.

BOLT, B. A. (1976). Abnormal seismology. *Bull. Seismol. Soc. Amer.* **66** 617–623.

BOLT, B. A. (1982). *Inside the Earth*. Freeman, San Francisco.

BOLT, B. A. and BRILLINGER, D. R. (1979). Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series. *Geophys. J. Roy. Astron. Soc.* **58** 593–603.

BOLT, B. A. and MARUSI, A. (1962). Eigenvibrations of the Earth observed at Trieste. *Geophys. J. Roy. Astron. Soc.* **6** 299–311.

BOLT, B. A., TSAI, Y. B., YEH, K. and HSU, M. K. (1982). Earthquake strong motions recorded by a large near-source array of digital seismographs. *Earthquake Engin. Structural Dynamics* **10** 561–573.

BOORE, D. M. (1977). The motion of the ground in earthquakes. *Scientific American* **237** 68–78.

BRILLINGER, D. R. (1975a). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

BRILLINGER, D. R. (1975b). Statistical inference for stationary point processes. In *Stochastic Processes and Related Topics* (M. L. Puri, ed.) **1** 55–99. Academic, New York.

BRILLINGER, D. R. (1978). Comparative aspects of the study of ordinary time series and of point processes. In *Developments in Statistics* (P. R. Krishnaiah, ed.) **1** 33–133. Academic, New York.

BRILLINGER, D. R. (1979). Some statistical methods in neurophysiology. Special Invited Lecture, Western Regional Meeting, Institute of Mathematical Statistics, Los Angeles.

BRILLINGER, D. R. (1981a). Some aspects of the analysis of evoked response experiments. In *Statistics and Related Topics* (M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, eds.) 155–168. North-Holland, Amsterdam.

BRILLINGER, D. R. (1981b). The general linear model in the design and analysis of evoked response experiments. *J. Theoret. Neurobiol.* **1** 105–119.

BRILLINGER, D. R. (1983). The finite Fourier transform of a stationary process. In *Handbook of Statistics* (D. R. Brillinger and P. R. Krishnaiah, eds.) **1** 21–37. North-Holland, Amsterdam.

BRILLINGER, D. R. (1985). A maximum likelihood approach to frequency–wavenumber analysis. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-33** 1076–1085.

BRILLINGER, D. R. (1986). Maximum likelihood analysis of spike trains of interacting nerve cells. Technical Report 80, Dept. Statistics, Univ. California, Berkeley.

BRILLINGER, D. R., BRYANT, H. L. and SEGUNDO, J. P. (1976). Identification of synaptic interactions. *Biol. Cybernetics* **22** 213–228.

BRILLINGER, D. R. and IHAKA, G. R. (1982). Maximum likelihood estimation of source parameters. *Earthquake Notes* **53** 39–40.

BRILLINGER, D. R. and SEGUNDO, J. P. (1979). Empirical examination of the threshold model of neuron firing. *Biol. Cybernetics* **35** 213–220.

BRILLINGER, D. R., UDIAS, A. and BOLT, B. A. (1980). A probability model for regional focal mechanism solutions. *Bull. Seismol. Soc. Amer.* **70** 149–170.

BRUNE, J. N. (1970 / 1971). Tectonic stress and the spectra of seismic shear waves from earthquakes. *J. Geophys. Res.* **75** 4997–5009. Correction **76** 5002.

BRYANT, H. L., MARCOS, A. R. and SEGUNDO, J. P. (1973). Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs. *J. Neurophysiol.* **36** 205–225.

BRYANT, H. L. and SEGUNDO, J. P. (1976). Spike initiation by transmembrane current: A white noise analysis. *J. Physiol.* **260** 279–314.

BUFORN, E. (1982). Estudio estadistico de la direccion de esfuerzos principales en terremotos. Doctoral thesis, Universidad Complutense de Madrid.

BULLEN, K. E. and BOLT, B. A. (1985). *An Introduction to the Theory of Seismology.* Cambridge Univ. Press, Cambridge.

BURNS, B. D. (1968). *The Uncertain Nervous System.* Arnold, London.

BYERLY, P. (1926). The Montana earthquake of June 28, 1925. *Bull. Seismol. Soc. Amer.* **16** 209–265.

CAPON, J. (1969). High-resolution frequency–wavenumber spectrum analysis. *Proc. IEEE* **57** 1408–1418.

CHILDERS, D. G. (1977). Evoked responses: Electrogenesis, models, methodology, and wavefront reconstruction and tracking analysis. *Proc. IEEE* **65** 611–626.

CHILDERS, D. G. and PAO, M. (1972). Complex demodulation for transient wavelet detection and extraction. *IEEE Trans. Audio Electroacoust.* **AU-20** 295–308.

CHIU, S.-T. (1986). Statistical estimation of the parameters of a moving source from array data. *Ann. Statist.* 14 559–578.

COLOQUHOUN, D. and HAWKES, A. G. (1983). The principles of the stochastic interpretation of ion-channel mechanisms. In *Single-Channel Recording* (B. Sakman and E. Neher, eds.) 135–175. Plenum, New York.

COOK, C. E. and BERNFIELD, M. (1967). *Radar Signals.* Academic, New York.

COPAS, J. B. (1983). Plotting p against x. *Appl. Statist.* **32** 25–31.

DAHLEN, F. A. (1982). The effect of data windows on the estimation of free oscillation parameters. *Geophys. J. Roy. Astron. Soc.* **69** 537–549.

DE WEERD, J. P. C. and KAP, J. I. (1981). Spectro–temporal representations and time-varying spectra of evoked potentials. *Biol. Cybernetics* **41** 101–117.

DONOHO, D. L. (1981). On minimum entropy deconvolution. In *Applied Time Series Analysis* (D. F. Findley, ed.) **2** 565–608. Academic, New York.

DONOHO, D. L., CHAMBERS, R. E. and LARNER, K. L. (1986). Robust estimation of residual statics corrections. Preprint, Dept. Statistics, Univ. California, Berkeley.

DUMERMUTH, G., HUBER, P. J., KLEINER, B. and GASSER, TH. (1971). Analysis of the interrelations between frequency bands of the EEG by means of the bispectrum. A preliminary study. *Electroenceph. Clin. Neurophysiol.* **31** 137–148.

FOLLEDO, M. (1983). Robust/resistant methods in the estimation of the evoked response curve. Ph.D. thesis, Univ. California, Berkeley.

FREEMAN, W. J. (1972). Linear analysis of the dynamics of neural masses. *Ann. Rev. Biophys. Bioeng.* **1** 225–256.

FREEMAN, W. J. (1975). *Mass Action in the Nervous System*. Academic, New York.

FREEMAN, W. J. (1979). Nonlinear dynamics of paleocortex manifested in the olfactory EEG. *Biol. Cybernetics* **35** 21–37.

FREEMAN, W. J. (1980). Measurement of corticalevoked potentials by decomposition of their wave forms. *J. Cybernetics Inform. Sci.* **2** 44–56.

FREEMAN, W. J. and SCHNEIDER, W. (1982). Changes in spatial patterns of rabbit olfactory EEG with conditioning to odors. *Psychophysiology* **19** 44–56.

GASSER, TH., MOCKS, J. and KOHLER, W. (1986). Amplitude probability distribution of noise for flash-evoked potentials and robust response estimates. *IEEE Trans. Biomed. Eng.* **BME-33** 579–583.

GASSER, TH., MOCKS, J., KOHLER, W. and DE WEERD, J. P. C. (1986). Performance and measures of performance estimators of brain potentials using real data. *IEEE Trans. Biomed. Eng.* **BME-33** 949–956.

GEVINS, A. S., MORGAN, N. H., BRESSLER, S. L., CUTILLO, B. A., WHITE, R. M., ILLES, J., GREER, D. S., DOYLE, J. C. and ZEITLIN, G. M. (1987). Human neuroelectric patterns predict performance accuracy. *Science* **235** 580–585.

GIANNAKIS, G. B. and MENDEL, J. M. (1986). Tomographic wavelet estimation via higher-order statistics. *Proc. Fifty-Sixth Internat. Conf. Soc. Explor. Geophys.* 512–514. Houston, Texas.

GLASER, E. M. and RUCHKIN, D. S. (1976). *Principles of Neurobiological Signal Analysis*. Academic, New York.

GRAJSKI, K. A., BREIMAN, L., DI PRISCO, G. V. and FREEMAN, W. J. (1986). Classification of EEG spatial patterns with a tree-structured methodology: CART. *IEEE Trans. Biomed. Eng.* **BME-33** 1076–1086.

HANSEN, R. A. (1982). Simultaneous estimation of terrestrial eigenvibrations. *Geophys. J. Roy. Astron. Soc.* **70** 155–172.

HASAN, T. (1982). Nonlinear time series regression for a class of amplitude modulated cosinusoids. *J. Time Series Anal.* **3** 109–122.

HAUBRICH, R. A. (1965). Earth noise, 5 to 500 millicycles per second. *J. Geophys. Res.* **70** 1415–1427.

HOLDEN, A. V. (1976). *Models for the Stochastic Activity of Neurones*. Springer, New York.

HOPPER, M. J. (1980). *Harwell Subroutine Library*. Atomic Energy Research Establishment, Harwell, England.

HUDSON, J. A. (1981). Mathematics for seismology. *J. Inst. Math. Appl.* **17** 34–39.

IHAKA, R. (1985). Ruaumoko. Ph.D. thesis, Dept. Statistics, Univ. California, Berkeley.

JACKSON, M. B. and LECAR, H. (1979). Single postsnyaptic channel currents in tissue cultured muscle. *Nature* **282** 863–864.

JEFFREYS, H. (1977). *Collected Papers of Sir Harold Jeffreys on Geophysics and Other Sciences*. Gordon and Breach, London.

JEFFREYS, H. and BULLEN, K. E. (1940). *Seismological Tables*. British Association, Gray-Milne Trust.

JOLLIFFE, I. T. (1986). *Principal Component Analysis*. Springer, New York.

KATZ, B. and MILEDI, R. (1971). Further observations on acetylcholine noise. *Nature* **232** 124–126.

KATZ, B. and MILEDI, R. (1972). The statistical nature of the acetylcholine potential and its molecular components. *J. Physiol.* (*London*) **224** 665–700.

KITAGAWA, G. and GERSCH, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. *IEEE Trans. Automat. Control* **AC-30** 48–56.

LABARCA, P., RICE, J. A., FREDKIN, D. R. and MONTAL, M. (1986). Kinetic analysis of channel gating: Application to the cholinergic receptor channel and the chloride channel from *Torpedo californica*. *Biophysical J.* **47** 469–478.

LAPWOOD, E. R. and USAMI, T. (1981). *Free Oscillations of the Earth*. Cambridge Univ. Press, Cambridge.

LECAR, H. (1981). Single-channel conductances and models of transport. In *The Biophysical Approach to Excitable Systems* (W. J. Adelman and D. E. Goldman, eds.) 109–121. Plenum, New York.

LII, K. S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *Ann. Statist.* **10** 1195–1208.

LINDBERG, C. R. (1986). Multiple taper spectral analysis of terrestrial free oscillations. Ph.D. thesis, Univ. California, San Diego.

LORENTZ, E. H. (1956). Empirical orthogonal functions and statistical weather prediction. Report 1, Statistical Forecasting Project, Dept. Meteorology, MIT, Cambridge, Mass.

MARMARELIS, P. Z. and MARMARELIS, V. Z. (1978). *Analysis of Physiological Systems*. Plenum, New York.

MARMARELIS, P. Z. and NAKA, N.-I. (1974). Identification of multi-input biological systems. *IEEE Trans. Biomed. Eng.* **BME-21** 88–101.

MENDEL, J. M. (1983). *Seismic Deconvolution: An Estimation-Based Approach*. Academic, New York.

MENDEL, J. M. (1986). Some modeling problems in reflection seismology. *IEEE ASSP Magazine* 4–17.

MILNE, R. K., EDESON, R. O. and MADSEN, B. W. (1986). Stochastic modeling of a single ion channel: Interdependence of burst length and number of openings per burst. *Proc. Roy. Soc. London Ser. B* **227** 83–102.

MOCKS, J., TUAN, P. D. and GASSER, TH. (1984). Testing for homogeneity of noisy signals evoked by repeated stimuli. *Ann. Statist.* **12** 193–209.

MOORE, G. P., PERKEL, D. H. and SEGUNDO, J. P. (1966). Statistical analysis and functional interpretation of neuronal spike data. *Ann. Rev. Physiol.* **28** 493–522.

NEHER, E. and SAKMAN, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260** 799–801.

NEYMAN, J. (1974). A view of biometry: An interdisciplinary domain concerned with chance mechanisms operating in living organisms; illustration: Urethan carcinogenis. In *Reliability and Biometry* 183–201. SIAM, Philadelphia.

NORCIA, A. M. and TYLER, C. W. (1985). Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Res.* **25** 1399–1408.

OGATA, Y. (1983). Likelihood analysis of point processes and its applications to seismological data. *Bull. Internat. Statist. Inst.* **50** (2), 943–961.

OGATA, Y. and KATSURA, K. (1986). Point-process models with linearly parameterized intensity for application to earthquake data. In *Essays in Time Series and Allied Processes* (J. Gani and M. B. Priestley, eds.) 291–310. Applied Probability Trust, Sheffield.

O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.* **1** 502–527.

POGGIO, T. and REICHARDT, W. (1981). Visual fixation and tracking by flies: Mathematical properties of simple control systems. *Biol. Cybernetics* **40** 101–112.

RAPIN, I. and GRAZIANI, L. J. (1967). Auditory-evoked responses in normal, brain-damaged and deaf infants. *Neurology* **17** 881–891.

ROBINSON, E. A. (1983). *Migration of Geophysical Data*. International Human Resources Development Corp., Boston.

SAGALOVSKY, B. D. (1982). Maximum likelihood and related estimation methods in point processes and point process systems. Ph.D. thesis, Univ. California, Berkeley.

SCHEIMER, J. and LANDERS, T. E. (1974). Short-period coda of a local event at LASA. *Semiannual Technical Summary* 42–44. Lincoln Lab., MIT, Cambridge, Mass.

SEGUNDO, J. P. (1984). La neurofisiologica: algunos supuestos y bases, recovecos e implicaciones. UNAM, Mexico.

SHUMWAY, R. H. and DER, Z. A. (1985). Deconvolution of multiple time series. *Technometrics* **27** 385–393.

SMITH, C. E. and CHEN, C.-L. (1986). Serial dependency in neural point processes due to cumulative after hyperpolarization. Statistics Mimeo Series 1691, North Carolina State Univ.

STEVENS, C. F. (1972). Inferences about membrane properties from electrical noise measurements. *Biophys. J.* **12** 1028–1047.

TJØSTHEIM, D. (1981). Multidimensional discrimination techniques—theory and application. In *Identification of Seismic Sources—Earthquake or Underground Explosion* (E. S. Husebye and S. Mykkeltveit, eds.) 663–694. Reidel, Dordrecht.

TUKEY, J. W. (1978). A data analysts's comments on a variety of points and issues. In *Event-Related Brain Potentials in Man* (E. Callaway, P. Tueting and S. H. Koslow, eds.) 139–154. Academic, New York.

UDIAS, A., MUNOZ, D. and BUFORN, E. (1985). Mecanismo de los terremotos y tectonica. Editorial de la Universidad complutens, Madrid.

VERE-JONES, D. and SMITH, E. G. C. (1981). Statistics in seismology. *Comm. Statist. A—Theory Methods* **10** 1559–1585.

WALD, A. (1952). On the principles of statistical inference. *Notre Dame Mathematical Lectures*, No. 1. Notre Dame, Ind.

WATERS, K. H. (1978). *Reflection Seismology*. Wiley, New York.

WEHRHAHN, C., POGGIO, T. and BULTHOFF, H. (1982). Tracking and chasing in houseflies. (*Musca*) *Biol. Cybernetics* **45** 123–130.

WIENER, N. (1930). Generalized harmonic analysis. *Acta Math.* **55** 117–258.

WOOD, L. C. and TREITEL, S. (1975). Seismic signal processing. *Proc. IEEE* **63** 649–661.

YASUI, S., DAVIS, W. and NAKA, K.-I. (1979). Spatio–temporal receptive field measurement of retinal neurons by random pattern stimulation and cross correlation. *IEEE Trans. Biomed. Eng.* **BME-26** 263–272.

ZADRO, M. B. and CAPUTO, M. (1968). Spectral, bispectral analysis and Q of the free oscillations of the Earth. *Nuovo Cimento Suppl.* **6** 67–81.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

Part II
Time Series Papers

Commentary by Pedro A. Morettin

This volume honors the work of David Brillinger in several areas, but most notably in the fields of point processes and of time series analysis and applications. It has been a privilege to me to have been his PhD student at Berkeley and then become a friend for the past 40 years. It has been quite a journey. David's work on time series has been influential to many people, especially for his students(over forty), many of them who pursued careers in time series and related fields. He is well know in Brazil, for his constant visits and valuable collaboration over the years. His book"Time Series: Data Analysis and Theory" became a classic and I was very fortunate to attend his classes using an earlier draft of the book.

In his assessment of the works of John W. Tukey, David wrote: "Anyone who has been involved with John has indeed been fortunate. They have probably remarked upon his rapid domination of the situation at hand, his extensive knowledge of pertinent physical background, his leaps in unthought of directions to concrete procedures, his vocabulary and possibly even his humor." I could not write anything more perfect that applies to David himself. His works pervaded several areas and disciplines, and some of them will be discussed in this volume.

I will concentrate on some articles in time series analysis, starting from some seminal works on second order spectra and polyspectra. Some of the papers are the kind of work that David likes to write on: they bring examples illustrating the interplay between the theory of time series, point processes, spatial processes and areas of applications, like geophysics, neurophysiology, sports, physics etc. I had read most of the papers, but reading them again was a wonderful experience and I hope this will be the case for the readers of this book.

An Introduction to polyspectra [1965]

I begin with this paper, because I think it is one of the most influential papers written by David. The purpose of the work is to derive: a) certain

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_9, © Springer Science+Business Media, LLC 2012

mathematical properties of polyspectra; b) estimates of polyspectra based on an observed stretch of a time series; c) certain statistical properties of proposed estimates and d) several applications of the results obtained.

The term polyspectrum is due to John Tukey. Polyspectra generalizes spectrum and bispectrum for a single time series and cross-spectra for a pair of time series. The author discusses why to use cumulants and not moments in the definition of polyspectra.

A particular class of discrete or continuous k-dimensional complex-valued processes is defined, and for members of this class the polyspectrum is defined as the Fourier transform of a cumulant of some order, assumed to exist. Some estimation procedures are considered (e. g. moment-type estimators and estimators based on complex demodulation) and asymptotic (complex) normal distributions are derived for the estimators.

Curiously enough the paper concludes with a note of pessimism on the use of polyspectra.

Asymptotic Theory of Estimates of k-th order Spectra [1967]

This work considers a vector of strictly stationary processes, all moments existing. Under mixing conditions given in terms of cumulants, the k-th order cumulant spectral density is defined as the Fourier transform of the corresponding k-th order cumulant of the process. Estimates for the cumulant spectral densities are provided and their properties derived. The proposed estimates are weighted averages of periodograms and asymptotic unbiasedness, joint normality and covariance structure are obtained. Some remarks on aliasing and previous works are made.

Asymptotic Properties of Spectral Estimates of Second Order [1969]

This article considers an r-variate strictly stationary, zero mean, stochastic process, satisfying some mixing condition. From the finite Fourier transform of T observations of the process, the periodogram and other estimates are proposed, namely estimates of the spectral measure, the autocovariance function and spectral density.

The asymptotic unbiasedness and asymptotic distributions of the estimates are derived. Under additional conditions, the asymptotic distribution of the periodogram is complex Wishart, for the spectral measure estimate is multivariate normal, the same for the autocovariance estimates and spectral density estimates. Some departures from the assumptions are commented on.

Fourier Analysis of Stationary Processes [1974]

This is an invited review paper on Fourier analysis (FA) of stationary processes written for the Proceedings of the IEEE. It begins with a description of important procedures in FA, including the estimation of the spectrum, fitting of parametric models and identification of linear systems.

The topics surveyed include: stationary real-valued discrete time series, the finite Fourier transform, the estimation of the spectrum, parametric models, linear models, vector-valued continuous spatial series, stationary point processes and stationary random Schwartz distributions.

Enphasis is on the large sample properties of estimators. Final remarks are made on higher order spectra and nonlinear systems.

The Digital Rainbow: Some History and Applications of Numerical Spectrum Analysis [1993]

This paper focus on the spectrum of a phenomenon. This is viewed as a display of the intensity of the phenomenon versus frequency. Some historical development of the field of spectrum analysis is given, with contributions of Michelson (1892), Schuster (1898), Einstein (1914), Fisher (1929), Bartlet (1950), Tukey (1958) and Yaglom (1987).

Some applications are given on: a) the free oscillations of the Earth, with an example of the 1960 Chilean earthquake; b) seismic surface waves (earthquake waves whose energy is trapped near the Earth's surface) and c) nuclear magnetic resonance (NMR) spectroscopy.

The paper concludes with some discussion on future prospects.

An Investigation of the Second-and Higher-order Spectra of Music [1998]

In this work the authors describe the two basic representations of music (signal and score representations), review some previous investigations and then present results of modelling second and higher order spectra in order to assess Gaussianity and linearity. They also discuss time series and marked point process representations. Four models for the spectra are considered (one of them being the $1/f$ noise model) and these are fitted to some scores, in particular the Bach's Coffee Cantata. Another feature is the fit of $1/f$ model to 12 music scores, ranging from Baroque to Latin music, concluding that it fits well.

Some Examples of Empirical Fourier Analysis in Scientific Problems [1999]

This is another paper describing some interesting applications of Fourier analysis in real problems. The article starts by stating the importance of Fourier analysis and them reviews some physical examples of the methodology. It continues giving some analytical background on Fourier and wavelet analysis and moves to stationary processes, central limit theorems and shrinking. Finally gives four examples, in electron microscopy, seismic surface waves, nuclear magnetic resonance spectroscopy and a wavelet analysis of microtubule movement (linear polymers basic to cell motility). The paper is concluded with some open problems.

Some Examples of Random Process Environmental Data Analysis [2000]

This paper presents examples in the environmental sciences. The processes analyzed range from point and marked point processes, to time series, spatial-temporal processes and particle processes. Preliminary some basic concepts and methods on random processes and inference are set down.

For the point process case an example from space science is given. The data consists of counts of orbiting debris, that may cause problems for space crafts. Statistics as the average periodogram, autointensity and coherence estimates are employed. Interest lie in the type of point process that better represent the data, the size of the particles (marks) and questions of independence of marks and times, and of altitudes and sizes.

In the case of a time series, the example is from Public Health. The data for analysis consists of average number of daily births in Toronto in 1986. The interest is in high level of cesarean deliveries. The counts are modelled by a Poisson with trending mean.

For spatio-temporal processes it is given an example from Neuroscience, concerning the olfactory system, data collected of the response of rabbits sniffing an odor. A random effects model is proposed followed by Fourier transformation.

Finally, for particle processes (path or trajectory of an object moving along a line), an example from Ecology is given, namely the migration path of an elephant seal. The proposed model is a nonlinear state space model.

The paper concludes with a discussion of other types of processes, data and techniques.

AN INTRODUCTION TO POLYSPECTRA[1]

By David R. Brillinger

The London School of Economics and Political Science

1. Introduction and summary. The subject of this paper is the higher-order spectra or polyspectra of multivariate stationary time series. The intent is to derive (i) certain mathematical properties of polyspectra, (ii) estimates of polyspectra based on an observed stretch of time series, (iii) certain statistical properties of the proposed estimates and (iv) several applications of the results obtained.

As might be expected, in lower order cases the polyspectrum reduces to spectra already considered. If one is considering a single time series, the first order polyspectrum is the usual power spectrum considered in [2], [14], [22], while the second order polyspectrum is the bispectrum considered in [12], [23], [28]. Also, if one is considering a pair of time series the first order polyspectrum is the cross-spectrum considered in [6], [10], [15].

For the case of a single time series the idea of a higher-order spectrum occurs in [3]. The idea has since been developed to a higher level of algebraic and analytic detail in [24]. Also in [24] the notion of considering a spectral representation for a cumulant rather than for a product moment occurs and is acknowledged to be due to Kolmogorov. Another related early paper is [18].

The present paper generalizes the definitions of these papers in the sense that k-dimensional time series are considered. Another contribution is a theorem indicating that for a broad class of processes one is wise to restrict consideration to cumulants rather than product moments.

Finally it should be noted that the term polyspectrum is due to J. W. Tukey. I have perhaps used the term in a more restricted sense than he would wish in that I have reserved it for the Fourier transform of a cumulant (at the expense of other functions of moments).

2. General motivation. In a heuristic sense the harmonic analysis of a time series $X(t)$ may be looked upon as the consideration of a representation of the series in the form,

$$X(t) = \sum R_k \exp [i(w_k t + \phi_k)]. \tag{2.1}$$

This consideration gains some validity from a theorem of Cramér's [9] to the effect that any covariance stationary time series $X(t)$ with mean 0, has a representation in the form

Received 13 July 1964; revised 2 April 1965.

[1] This research was commenced when the author had a joint appointment at the Bell Telephone Laboratories and Princeton University. At Princeton the research received the support of National Science Foundation Grant NSF-GS30 and Army Research Office Contract DA 36-034-ORD-2297.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_10, © Springer Science+Business Media, LLC 2012

(2.2) $$X(t) = \int e^{iwt}\, dZ(w)$$

where $Z(w)$ is a stochastic function.

A further aspect of harmonic analysis is that one often acts as if the various terms, $R_k \exp [i(w_k t + \phi_k)]$, appearing in (2.1) are independent of one another. This simplifying thought is possibly instigated by the knowledge that $Z(w)$ appearing in (2.2) is such that

(2.3) $$E\{dZ(w)\, dZ^*(w')\} = 0, \qquad w \neq w',$$

implying independence in the real-valued Gaussian case. Or perhaps it is due to the fact that one often imagines a series as being generated by a number of linear time invariant operations on a Wiener process and one knows that such operations do not mix up frequencies. (See [21], p. 83 for example.)

In practice however the frequency components of a time series do not always appear to be independent. In a study of ocean wave records, [12], Hasselmann, Munk and MacDonald have found empirically various wave components related to one another. In a study of the effect of introducing a signal into the eye [31] Van der Tweel has found that the responses at 5 c/s and 10 c/s are related to one another. Many economists have noted a seasonal effect in economic time series of persistent non-cosinusoidal shape. This finding perhaps indicates that the various harmonics of 1 cycle/year are in some form of fixed relation with one another.

A simple form of tying together of frequency components occurs if a number of independent frequency components, $R_k \exp [i(w_k t + \phi_k)]$, instead of simply adding together to produce a series $X(t)$, as in (2.1), add together and also multiply together in pairs to produce the series

(2.4) $$\sum R_k \exp [i(w_k t + \phi_k)] + \sum A_{jk} \exp (i\alpha_{jk}) R_j R_k \exp [i(w_j + w_k) + i(\phi_j + \phi_k)].$$

That is, we are moving away from an additive model to a model containing second order product interactions.

The reader will note that, in the expression (2.4), the correlation between the product of the components at frequencies w_j and w_k with the component at frequency $w_j + w_k$ is one, provided the sum of no other pair of frequencies present is $w_j + w_k$. This observation will later lead us directly to the polyspectrum.

Continuing to consider (2.4), a simple means of producing a time series containing terms such as those in (2.4) is to take a series $X(t)$ with a simple harmonic analysis and then to form the series

(2.5) $$Y(t) = f[X(t)]$$

where f is a non-linear function. In a situation in which one is given the series $Y(t)$, one would like to find the function f in order to be able to remove the non-additivity that it has introduced. A coefficient will be proposed for this purpose in Section 6.

As a final point, people often introduce the power spectrum by noting the ease with which linear time invariant operations may be described in terms of it. If one wishes to describe easily the effect of a multilinear or polynomial (in the sense of [20]) operation, one finds oneself led to higher order spectra. Tukey in [30] has commenced the development of a calculus relating polynomial operations to higher order spectra.

3. Definitions. Various classes of stochastic processes have been introduced in order to deal with higher order spectral moments. Specifically the classes $T^{(k)}$, $S^{(k)}$, and $\Phi^{(k)}$ defined below have been discussed in [24]. However we shall require a new class. Before defining this new class $\Psi^{(k)}$, let us present the definitions of $T^{(k)}$, $S^{(k)}$, and $\Phi^{(k)}$ as they will also be needed in the paper.

Let $U(t)$ be a real measurable random process, $-\infty < t < \infty$. Then

$T^{(k)}$ denotes the class of processes $U(t)$ for which

$$E\,|U(t)|^k \leqq C_k < \infty; \tag{3.1}$$

$S^{(k)}$ denotes the class of processes $U(t)$ belonging to $T^{(k)}$ and such that for $1 \leqq j \leqq k$, $-\infty < u < \infty$,

$$EU(t_1) \cdots U(t_j) = EU(t_1 + u) \cdots U(t_j + u); \tag{3.2}$$

$\Phi^{(k)}$ denotes the class of processes $U(t)$ belonging to $T^{(k)}$ and such that for $1 \leqq j \leqq k$, there exist functions $M^{(j)}(w_1, \cdots, w_j)$ of bounded variation such that

$$EU(t_1) \cdots U(t_j) = \int \cdots \int \exp\,[i(w_1 t_1 + \cdots + w_j t_j)]\, dM^{(j)}(w_1, \cdots, w_j). \tag{3.3}$$

Before defining $\Psi^{(k)}$ the following notation will be required:

(i) $(v_1, \cdots, v_j)$ denotes a grouping of the integers $1, 2, \cdots, k$ into j groups $v_1, \cdots, v_j$;

(ii) $t_v = (t_{h_1}, \cdots, t_{h_n})$ when v corresponds to the grouping $(h_1, \cdots, h_n)$. For example if $v = (1, 8, 9)$ then $t_v = (t_1, t_8, t_9)$;

(iii) $\{X_1(t), \cdots, X_k(t)\}$ stands for a k-dimensional complex-valued stochastic process;

(iv) $m_{1 \cdots k}(t_1, \cdots, t_k)$ denotes the kth order product moment $EX_1(t_1) \cdots X_k(t_k)$;

(v) $c_{1 \cdots k}(t_1, \cdots, t_k)$ denotes the kth order cumulant

$$\sum (-1)^{p-1}(p-1)!\, m_{v_1}(t_{v_1}) \cdots m_{v_p}(t_{v_p}) \tag{3.4}$$

where the summation extends over all groupings of the integers $1, \cdots, k$.

$\Psi^{(k)}$ is now defined as the class of discrete or continuous time k-dimensional complex-valued processes $\{X_1(t), \cdots, X_k(t)\}$ such that

(a) for $1 \leqq j \leqq k$ and $1 \leqq h_1, \cdots, h_j \leqq k$, $m_{h_1 \cdots h_j}(t_1, \cdots, t_j)$ exists,

(b) for $1 \leqq j \leqq k$ and $-\infty < u < \infty$ in the continuous case or $u = 0, \pm 1, \pm 2, \cdots$ in the discrete case

$$m_{h_1 \cdots h_j}(t_1 + u, \cdots, t_j + u) = m_{h_1 \cdots h_j}(t_1, \cdots, t_j),$$

(c) for $1 \leqq j \leqq k$ there exist measures $\delta(w_1 + \cdots + w_j)C_{h_1\cdots h_j}(w_1, \cdots, w_j)\, dw_1 \cdots dw_j$ absolutely continuous with respect to Lebesgue measure on the plane $w_1 + \cdots + w_j = 0$ such that $c_{h_1\cdots h_j}(t_1, \cdots, t_j)$ equals

$$(3.5) \quad \int \cdots \int \exp\,[i(w_1 t_1 + \cdots + w_j t_j)] \cdot \delta(w_1 + \cdots + w_j)C_{h_1\cdots h_j}(w_1, \cdots, w_j)\, dw_1 \cdots dw_j .$$

(Throughout the paper $\delta(w)$ denotes the Dirac delta function. The reader may easily see that the condition (b) above implies that the measure must in fact have support on the plane $w_1 + \cdots + w_j = 0$.)

If a process $\{X_1(t), \cdots, X_k(t)\}$ belongs to $\Psi^{(k)}$ then $C_{1\cdots k}(w_1, \cdots, w_k)$ is defined to be the $(k - 1)$th order *polyspectrum* of the process.

A number of comments may be made about $\Psi^{(k)}$.

(i) In the integrals above, the range of the arguments w is $-\pi \leqq w \leqq \pi$ in the discrete case and $-\infty < w < \infty$ in the continuous case.

(ii) If the series involved are real

$$(3.6) \qquad C^*_{1\cdots k}(w_1, \cdots, w_k) = C_{1\cdots k}(-w_1, \cdots, -w_k).$$

(iii) If the series are identical, $C_{1\cdots k}(w_1, \cdots, w_k)$ is symmetric in its arguments.

(iv) If a process $\{X_1(t), \cdots, X_k(t)\}$ satisfies (a) and (b) above and for ϕ equal both c and C,

$$\int \cdots \int |\phi_{h_1\cdots h_j}(t_1, \cdots, t_{j-1}, 0)|\, dt_1 \cdots dt_{j-1} < \infty$$

in the continuous case, or

$$\sum \cdots \sum |\phi_{h_1\cdots h_j}(t_1, \cdots, t_{j-1}, 0)| < \infty$$

in the discrete case, then the process belongs to $\Psi^{(k)}$. In this case the Fourier relation (3.5) may in fact be inverted;

(v) $C_{1\cdots k}(w_1, \cdots, w_k)$ being a complex number, for some purposes it may be useful to express it in terms of an amplitude and phase;

(vi) If $\{X_1(t), \cdots, X_k(t)\}$ is in fact $\{X(t), \cdots, X(t)\}$, i.e. all of the components are identical, and if $X(t)$ is real then $\Psi^{(k)}$ reduces to the class $\Delta^{(k)}$ introduced by Kolmogorov (see [24].)

This section will be concluded with a number of examples of polyspectra.

Example 1. Suppose $\{X(t), X(t)\}$ denotes a two dimensional real process with identical components, then the first order polyspectrum $C_{11}(w_1, w_2)$ reduces to the power spectrum of $X(t)$.

Example 2. Suppose $\{X_1(t), X_2(t)\}$ denotes a two dimensional real process, then the first order polyspectrum $C_{12}(w_1, w_2)$ reduces to the cross-spectrum of the two series $X_1(t)$ and $X_2(t)$.

Example 3. Suppose $X(t) = \int g(t - u)\, dY(u)$ where $\int |g(u)|\, du < \infty$ and $Y(u)$ is a process with stationary and independent increments. Denote the jth cumulant of $Y(1) - Y(0)$ by K_j (it being assumed to exist). The $(j - 1)$th order polyspectrum of $X(t)$ (really of $\{X(t), \cdots, X(t)\}$) is given by

$$K_j G(w_1) \cdots G(w_j) \tag{3.7}$$

where $G(w) = \int e^{-iwu} g(u)\, du$ and $w_1 + \cdots + w_j = 0$.

This result may be demonstrated by making use of the characteristic functional of the process as derived in [1], p. 148 for example.

EXAMPLE 4. Suppose

$$X(t) = \int a(t - u)\, dW(u) + \iint b(t - u, t - v)\, dW(u)\, dW(v) \tag{3.8}$$

where $W(t)$ is a Wiener process, $a(u)$ and $b(u, v)$ have Fourier transforms $A(w)$, $B(w_1, w_2)$ respectively and $b(u, v)$ is assumed symmetric in u and v for convenience. In this case the second order polyspectrum or bispectrum of $X(t)$ is given by,

$$\begin{aligned} &2[A(w_1)A(w_2)B(-w_1, -w_2) + A(w_2)A(w_3)B(-w_2, -w_3) \\ &\qquad + A(w_3)A(w_1)B(-w_3, -w_1)] \\ &\qquad + 8 \int \overline{B(w, w_1 - w)B(w_2 + w, -w)B(w - w_1, -w - w_2)}\, dw \end{aligned} \tag{3.9}$$

where the bar denotes the mean of all permutations of (w_1, w_2, w_3).

This result may be demonstrated by making use of the formula for the kth order product moment of a Wiener process (see [32]).

EXAMPLE 5. Suppose

$$\begin{aligned} X_1(t) &= \int a(t - u)\, dW_1(u), \\ X_2(t) &= \int b(t - u)\, dW_2(u), \\ X_3(t) &= \iint c(t - u, t - v)\, dW_1(u)\, dW_2(v), \end{aligned} \tag{3.10}$$

where $W_1(t)$ and $W_2(t)$ are independent Wiener processes and where $a(u)$, $b(u)$ and $c(u, v)$ have Fourier transforms $A(w)$, $B(w)$ and $C(w_1, w_2)$ respectively. In this case the second order polyspectrum of $\{X_1(t), X_2(t), X_3(t)\}$ is given by,

$$A(w_1)B(w_2)C(-w_1, -w_2). \tag{3.11}$$

If in fact $W_1(t)$ and $W_2(t)$ are not independent, but are completely dependent, $W_1(t) = W_2(t)$, then the polyspectrum is given by,

$$A(w_1)B(w_2)\{C(-w_1, -w_2) + C(-w_2, -w_1)\} \tag{3.12}$$

or if $c(u, v)$ is symmetric in u and v by,

$$2A(w_1)B(w_2)C(-w_1, -w_2). \tag{3.13}$$

4. Estimation. In this section it will be supposed that an observed stretch $\{X_1(t), \cdots, X_k(t); 0 \leqq t \leqq T\}$ of a real discrete time series belonging to $\Psi^{(k)}$, $k \geqq 2$, is available. (The corresponding procedures for a stretch of a continuous time series are immediately apparent.) Three distinct techniques for the estimation of polyspectra will be proposed.

The first technique follows directly from the definition (3.5) which indicates that the polyspectrum $C_{1\cdots k}(w_1, \cdots, w_k)$, $\sum w_j = 0$, is the $(k-1)$-dimensional Fourier transform of $c_{1\cdots k}(t_1, \cdots, t_{k-1}, 0)$. The technique is the following; estimate the product moment $m_{1\cdots k}(t_1, \cdots, t_{k-1}, 0)$, and all necessary lower order product moments by formulae of the form,

$$(4.1) \quad \hat{m}_{1\cdots k}(t_1, \cdots, t_{k-1}, 0) = T^{-1} \textstyle\sum_{t=0}^{s} X_1(t_1 + t) \cdots X_{k-1}(t_{k-1} + t) X_k(t)$$

where $s = T - \max \{t_j\}$, and for $(t_1, \cdots, t_{k-1})$ in a set I to be specified later. The joint cumulant $c_{1\cdots k}(t_1, \cdots, t_{k-1}, 0)$ may now be estimated by substituting in (3.4), i.e. by forming

$$(4.2) \quad \hat{c}_{1\cdots k}(t_1, \cdots, t_{k-1}, 0) = \sum (-1)^{p-1}(p-1)!\, \hat{m}_{v_1}(t_{v_1}) \cdots \hat{m}_{v_p}(t_{v_p})$$

where the summation extends over all grouping of the integers $1, \cdots, k$ and $t_k = 0$. (Some workers may wish to divide by $T - s + 1$ rather than T in the expression (4.1). For a discussion of this point in the first order case see [29]. Also some workers may perhaps wish to substitute into the formulas for Fisher's k-statistics. For a definition of these latter see [16].) The estimate (4.2) has one undesirable property, namely it is not invariant under changes $X_j(t) \to X_j(t) + h_j$, whereas the corresponding population cumulant is. This defect may be remedied by first subtracting the sample means from the series before calculating the estimate. In this case the summation in (3.4) extends only over groupings containing no first-order elements.

From (3.5) we see that $\hat{c}_{1\cdots k}(t_1, \cdots, t_{k-1}, 0)$ is estimating

$$(4.3) \quad \int \cdots \int \exp [i(w_1 t_1 + \cdots + w_{k-1} t_{k-1})] C_{1\cdots k}(w_1, \cdots, w_k)\, dw_1 \cdots dw_{k-1}$$

where $\sum w_j = 0$. That is it is estimating the coefficient of a term in the Fourier series expansion of $C_{1\cdots k}(w_1, \cdots, w_k)$, $\sum w_j = 0$. The sum of a number of such terms may be used to approximate the function itself; however classical Fourier analysis indicates that the use of a summability technique may well improve the approximation (see [13] for example). This leads one to consider estimates of the form,

$$(4.4) \quad (1/2\pi)^{k-1} \textstyle\sum_I \lambda^{(n)}_{t_1 \cdots t_{k-1}} \exp [-i(w_1 t_1 + \cdots + w_{k-1} t_{k-1})] \cdot \hat{c}_{1\cdots k}(t_1, \cdots, t_{k-1}, 0)$$

where the $\lambda^{(n)}_{t_1 \cdots t_{k-1}}$ are the convergence factors of a summability method.

Convergence factors that seem appropriate for this situation include;

(a) a product of one dimensional convergence factors, i.e.

$$(4.5) \quad \lambda^{(n)}_{t_1 \cdots t_{k-1}} = \lambda^{(n)}_{t_1} \cdots \lambda^{(n)}_{t_{k-1}}$$

where for example (Féjer summability)

$$(4.6) \quad \lambda^{(n)}_t = 1 - |t|/n, \quad 0 \leqq |t| \leqq n$$
$$= 0 \quad \text{otherwise}$$

or (Tukey summability)

$$(4.7) \qquad \lambda_t^{(n)} = .54 + .46 \cos \pi t/n, \qquad 0 \leqq |t| \leqq n$$
$$= 0 \quad \text{otherwise};$$

(b) a genuine multidimensional factor such as (Bochner-Riesz summability, see [4])

$$(4.8) \qquad \lambda_{t_1 \cdots t_{k-1}}^{(n)} = (1 - |t|^2/n^2)^m, \qquad 0 \leqq |t| \leqq n$$
$$= 0 \quad \text{otherwise}$$

where $|t|^2 = t_1^2 + \cdots + t_{k-1}^2$.

This last factor has the advantage that the convergence of the approximating series at a specified point depends only on the behaviour of the function in the neighborhood of the point.

We see that the set I mentioned earlier is in fact determined by the non-zero values of the convergence factors.

Before describing the second estimation technique, let us introduce another means of looking at the polyspectrum. Because $\Psi^{(k)} \subseteq \Psi^{(2)}$, the series $\{X_j(t)\}$ has a Cramér representation

$$(4.9) \qquad \{\int e^{iwt}\, dZ_j(w)\}$$

where $Z_j(w)$ is a stochastic function.

In terms of this representation the cumulant $c_{1\cdots k}(t_1, \cdots, t_k)$ may be written

$$(4.10) \qquad \int \cdots \int \exp\, [i(w_1 t_1 + \cdots + w_k t_k)] \mathcal{C}(dZ_1(w_1), \cdots, dZ_k(w_k))$$

where $\mathcal{C}(x_1, \cdots, x_k)$ denotes the joint cumulant of $x_1, \cdots, x_k$. (This results from the fact that the joint cumulant of $y_1, \cdots, y_m$ where $y_k = \sum a_{i_k k} x_{i_k k}$ is given by

$$(4.11) \qquad \sum \cdots \sum a_{i_1 1} \cdots a_{i_m m} \mathcal{C}(x_{i_1 1}, \cdots, x_{i_m m}).$$

In fact (4.11) would appear to be one of the main reasons why cumulants prove so useful. It states that one can write down immediately the joint cumulant of a number of linear combinations of independent or dependent random variables, in terms of their joint cumulants.)

Comparing (3.5) and (4.10) and assuming that the Fourier transform is unique almost everywhere,

$$(4.12) \qquad \delta(w_1 + \cdots + w_k) C_{1\cdots k}(w_1, \cdots, w_k)\, dw_1 \cdots dw_k$$
$$= \mathcal{C}(dZ_1(w_1), \cdots, dZ_k(w_k)) \quad \text{(almost everywhere)}.$$

This indicates that with realizations of the spectral functions $dZ_j(w_j)$ and a proper normalization one can estimate the polyspectrum.

After this introduction, it can be stated that the second proposed technique of estimating polyspectra is based upon obtaining realizations of the spectral functions by means of the procedure of complex demodulation ([8], [29]). Given an observed stretch of series $\{X_j(t), 0 \leqq t \leqq T\}$ the steps are as follows:

(i) form $X_j(t) \cos w_0 t$ and $X_j(t) \sin w_0 t$, $0 \leqq t \leqq T$,

(ii) form the series,

$$U_j(t, w_0) = (2k + 1)^{-1} \sum_{s=-k}^{k} \lambda_{t-s}^{(k)} X_j(s) \cos w_0 s, \tag{4.13}$$

$$U_j^H(t, w_0) = (2k + 1)^{-1} \sum_{s=-k}^{k} \lambda_{t-s}^{(k)} X_j(s) \sin w_0 s, \tag{4.14}$$

$k \leqq t \leqq T - k$, where for example $\lambda_t^{(k)}$ is given by (4.6) or (4.7). $\Delta Z_j(w_j)$ may now be approximated by $U_j(t, w_j) + iU_j^H(t, w_j)$.

(iii) $C_{1 \cdots k}(w_1, \cdots, w_k)$, $\sum w_j = 0$, is now estimated by forming

$$\sum (-1)^{p-1}(p-1)! \, \hat{m}_{v_1} \cdots \hat{m}_{v_p}, \tag{4.15}$$

the summation extending over all groupings of $1, 2, \cdots, k$ and $\hat{m}_v$ is given by,

$$T^{-1} \sum_{t=k}^{T-k} \{U_{h_1}(t, w_{h_1}) + iU_{h_1}^H(t, w_{h_1})\} \cdots \{U_{h_p}(t, w_{h_p}) + iU_{h_p}^H(t, w_{h_p})\} \tag{4.16}$$

where $v = (h_1, \cdots, h_p)$.

The final technique proposed for the estimation of a polyspectrum is based upon the fact that the expression (4.12) is also equal to

$$\mathcal{C}(e^{iw_1 t} \, dZ_1(w_1), \cdots, e^{iw_k t} \, dZ_k(w_k)). \tag{4.17}$$

The polyspectrum can consequently be estimated by obtaining realizations of the frequency components $e^{iw_j t} \Delta Z_j(w_j)$. These realizations may be obtained by deriving estimates of $X_j(t, w_0)$, the component of frequency w_0 in the series $X_j(t)$ and $X_j^H(t, w_0)$ the corresponding Hilbert transform (see [8]). A useful technique for obtaining $X_j(t, w_0)$ and $X_j^H(t, w_0)$ is described below:

$$e^{iw_j t} \Delta Z_j(w_j) \quad \text{may be estimated by} \quad X_j(t, w_j) + iX_j^H(t, w_j).$$

$C_{1 \cdots k}(w_1, \cdots, w_k)$, $\sum w_j = 0$, may be estimated by forming the expression (4.15) where U and U^H in (4.16) are replaced by X and X^H respectively.

The promised technique for obtaining $X_j(t, w_0)$ and $X_j^H(t, w_0)$ evolves from a procedure suggested in [11], pp. 77–78. Define

$$a_m(t) = N^{-1} \sum_{s=-N}^{N} X(t + s) \cos \pi m s / N, \qquad m = 0, 1, \cdots, N, \tag{4.18}$$

$$b_m(t) = N^{-1} \sum_{s=-N}^{N} X(t + s) \sin \pi m s / N, \qquad m = 1, \cdots, N - 1, \tag{4.19}$$

where $w_0 = \pi m / N$. The advantage of this definition is that the $a_m(t)$, $b_m(t)$ may be generated by recursion,

$$a_m(t + 1) = a_m(t) \cos \pi m / N + b_m(t) \sin \pi m / N + [(-1)^m / N][X(N + 1 + t) - X(-N + t)], \tag{4.20}$$

$$b_m(t + 1) = -a_m(t) \sin \pi m / N + b_m(t) \cos \pi m / N. \tag{4.21}$$

Use of these recursion relations greatly decreases the number of arithmetical operations involved.

The proposed estimates of $X(t, w_0)$ and $X^H(t, w_0)$ are now,

$$.23a_{m-1}(t) + .54a_m(t) + .23a_{m+1}(t), \tag{4.22}$$

$$.23b_{m-1}(t) + .54b_m(t) + .23b_{m+1}(t), \tag{4.23}$$

respectively. (The coefficients .23 and .54 used here are derived from Tukey's weights.)

For later reference it is noted here that in terms of the spectral representation $a_m(t, w_0) + ib_m(t, w_0)$ may be written,

$$\int \exp\,(iwt + i\tfrac{1}{2}\theta)\,[\sin N\theta/N \sin \tfrac{1}{2}\theta]\,dZ(w) \tag{4.24}$$

where $\theta = w + w_0$. This means that $X(t, w_0) + iX^H(t, w_0)$ equals

$$\int e^{iwt}Q(w + w_0)\,dZ(w) \tag{4.25}$$

with an elementary function $Q(w)$.

The final two estimation techniques proposed above have several advantages over the first. They are easily adapted to obtain running estimates of the polyspectrum and so the presence of nonstationarities may be investigated. Also once an initial effort has been made to obtain the series $X(t, w_0) + iX^H(t, w_0)$ or $U(t, w_0) + iU^H(t, w_0)$, they may be put to a variety of uses with few additional calculations; for example polyspectra of various orders, involving various series may be calculated. These series should have to be calculated only once in the history of a series, provided enough foresight is shown in the bandwidths of the filters employed. The series $U + iU^H$ has a further advantage; typically it is fairly smooth so not every value need necessarily be retained.

5. Some statistical properties of the proposed estimates. The discussion in this section will be restricted to the discrete case; however the continuous case follows in an identical manner, sums in the time domain being replaced by integrals, and integrals in the frequency domain having their range increased from $-\pi, \pi$ to $-\infty, \infty$.

Suppose a stretch of a time series $\{X_1(t), \cdots X_k(t); -T' \leqq t \leqq T'\}$ is available. When the second and third estimates of Section 4 are examined in detail for this case, it is seen that they have the form,

$$\hat{C}_{1\cdots k} = \sum (-1)^{p-1}(p-1)!\,\hat{n}_{\nu_1} \cdots \hat{n}_{\nu_p} \tag{5.1}$$

where when $\nu = (j_1, \cdots, j_m)$,

$$\hat{n}_\nu = (2T+1)^{-1} \textstyle\sum_{t=-T}^{T} Y_{j_1}(t) \cdots Y_{j_m}(t) \tag{5.2}$$

with

$$Y_j(t) = \sum g_j(t-u)X_j(u) \tag{5.3}$$

for (complex valued) functions $g_j(t)$ related to the filters employed and where $T' > T > 0$.

We will restrict consideration to estimates of this form throughout this section.

The $g_j(t)$ appearing in (5.3) will be said to be *absolutely summable* if they are such that

(5.4) $$\sum_{t=-\infty}^{\infty} |g_j(t)| < \infty.$$

Also if the time series $\{X_1(t), \cdots, X_k(t)\}$ is such that for joint cumulants of all orders and all sets of subscripts $(i_1, \cdots, i_m)$,

(5.5) $$\sum_{t_2} \cdots \sum_{t_m} |c_{i_1 \cdots i_m}(0, t_2, \cdots, t_m)| < \infty.$$

Then the series is said to satisfy *Condition A*.

Expressions will be required for the joint cumulants of a number of non-elementary random variables. Before presenting these expressions however, let us introduce some terminology of [17].

Consider the two-way table

(5.6) $$\begin{matrix} (1, 1) & \cdots & (1, k_1) \\ \vdots & & \vdots \\ (j, 1) & \cdots & (j, k_j) \end{matrix}$$

and a partition $P_1 \cup P_2 \cup \cdots \cup P_m$ of its elements. We shall say that the sets P_{i^1} and P_{i_2} of the partition *hook* if there exist $(i_1, j_1) \varepsilon P_{i_1}$ and $(i_2, j_2) \varepsilon P_{i_2}$ such that $i_1 = i_2$. We shall say that the sets $P_{i'}$ and $P_{i''}$ *communicate* if there exists a sequence of sets $P_{i_1} = P_{i'}, P_{i_2}, \cdots, P_{i_r} = P_{i''}$ such that P_{i_j} and $P_{i_{j+1}}$ hook for each j. A partition is said to be *indecomposable* if and only if all its sets communicate.

If the rows of table (5.6) are denoted by $R_1, \cdots, R_j$ then $\{P_1, \cdots, P_m\}$ is indecomposable if and only if there exist no sets $P_{i_1}, \cdots, P_{i_r} (r < m)$ and rows $R_{j_1}, \cdots, R_{j_s} (s < j)$ with

(5.7) $$P_{i_1} \cup \cdots \cup P_{i_r} = R_{j_1} \cup \cdots \cup R_{j_s}.$$

The indecomposable partitions correspond to the arrays of [16], Rule 3, p. 283, when the rule is extended to the higher dimensional case.

LEMMA 5.1. *Consider a (not necessarily rectangular) array* $||x_{mn}||$ *of random variables* x_{mn}. *Consider the* j *random variables*

(5.8) $$y_m = \prod_{n=1}^{k_m} x_{mn}.$$

The joint jth *order cumulant* $\mathcal{C}(y_1, \cdots, y_j)$ *is given by*

(5.9) $$\sum_v \mathcal{C}_{v_1} \cdots \mathcal{C}_{v_p}$$

where $\mathcal{C}_v = \mathcal{C}(x_{a_1}, \cdots, x_{a_m})$ *when* $v = (a_1, \cdots, a_m)$, *(the a's are pairs of integers selected from table* (5.6)*), and the summation in* (5.9) *extends over all the indecomposable partitions of* (5.6).

PROOF. This result follows immediately from a theorem of [17].

LEMMA 5.2. *Consider series* $Z_1(t), \cdots, Z_k(t)$ *of the form*

(5.10) $$Z_j(t) = \sum h_j(t - u) X_j(u)$$

with $h_j(t)$ *complex-valued, bounded*[2] *and absolutely summable. Suppose that the*

[2] The boundness follows from the absolute summability in fact.

series $\{X_1(t), \cdots, X_k(t)\}$ satisfies Condition A and let $(v_1, \cdots, v_p)$ denote an indecomposable partition of table (5.6) *and $f_{v_i}(s_{v_i})$, $(s_1 = 0)$, the joint cumulant of elements selected from the table*

$$(5.11)\qquad \begin{matrix} Z_1(0) & \cdots & Z_k(0) \\ Z_1(s_2) & \cdots & Z_k(s_2) \\ \vdots & \cdots & \vdots \\ Z_1(s_j) & \cdots & Z_k(s_j) \end{matrix}$$

in accordance with v_i. Under these conditions

$$\sum_{s_2=-\infty}^{\infty} \cdots \sum_{s_j=-\infty}^{\infty} |f_{v_1}(s_{v_1}) \cdots f_{v_p}(s_{v_p})| < \infty.$$

PROOF. If $v = \{(i_1, j_1), \cdots, (i_m, j_m)\}$, define $\tilde{h}_v(t_v)$ as

$$(5.12)\qquad h_{j_1}(t_{i_1, j_1}) \cdots h_{j_m}(t_{i_m, j_m}).$$

Now $\sum_{s_2} \cdots \sum_{s_j} f_{v_1}(s_{v_1}) \cdots f_{v_p}(s_{v_p})$

$$(5.13)\qquad = \sum_{t_{11}} \cdots \sum_{t_{jk}} \sum_{s_2} \cdots \sum_{s_j} \tilde{h}_{v_1}(s_{v_1} - t_{v_1}) \cdots \tilde{h}_{v_p}(s_{v_p} - t_{v_p})\, c_{v_1}(t_{v_1}) \cdots c_{v_p}(t_{v_p})$$

$$(5.14)\qquad = \sum_{u_{v_1}} \cdots \sum_{u_{v_p}} \sum_{t_1} \cdots \sum_{t_p} \sum_{s_2} \cdots \sum_{s_j} \tilde{h}_{v_1}(s_{v_1} - u_{v_1} - t_1) \cdots \tilde{h}_{v_p}(s_{v_p} - u_{v_p} - t_p) c_{v_1}(u_{v_1}) \cdots c_{v_p}(u_{v_p}),$$

where t_i is one of the arguments of t_{v_i} and $u_{v_i} = t_{v_i} - t_i$. (We are here taking advantage of the stationarity of the process.) Now in the arguments of the h's there occur a variety of $s_i - t_m$. Since the partition is indecomposable, there exist $p + j - 1$ of these such that the relationship

$$(5.15)\qquad s_{i_n} - t_{m_n} = a_n,$$

$n = 1, \cdots, p + j - 1$ is non-singular.

Let us substitute the a's into (5.14) retaining $p + j - 1$ h's with arguments of the form $a_m - u_n$ and note that the remaining h's are bounded. Thus the absolute value of (5.13) is

$$(5.16)\qquad \leqq M \sum_{u_{v_1}} \cdots \sum_{u_{v_p}} |c_{v_1}(u_{v_1}) \cdots c_{v_p}(u_{v_p})| \cdot \sum_{a_1} \cdots \sum_{a_{p+j-1}} |h(a_1 - u_{m_1}) \cdots h(a_{p+j-1} - u_{m_{p+j-1}})|.$$

That (5.16) is bounded now follows from the discrete analog of Theorem 33 of [5]. The interchanges of the various summations in this lemma may be justified by Fubini's theorem.

Define a_T to be

$$(5.17)\qquad (2T + 1)^{-1} \sum_{-T}^{T} Y_1(t) \cdots Y_k(t) - \sum_{p>1} \prod_{j=1}^{p} \sum \cdots \sum \tilde{g}_{v_j}(-t_{v_j}) c_{v_j}(t_{v_j}).$$

THEOREM 5.1. *Consider the real-valued random variable $\alpha a_T + \alpha^* a_T^*$ where the $g_j(t)$ are bounded*[3] *and absolutely summable and the series $\{X_1(t), \cdots, X_k(t)\}$*

[3] The boundness follows from the absolute summability in fact.

satisfies Condition A. Under these conditions the jth *cumulant of* $(\alpha a_T + \alpha^* a_T^*) \cdot (2T+1)^{1-j^{-1}}$ *approaches*

$$\sum_v \sum_\epsilon \sum_{s_2} \cdots \sum_{s_j} \alpha^{\epsilon_1} \cdots \alpha^{\epsilon_j} d_{v_1}(s_{v_1}) \cdots d_{v_p}(s_{v_p}) \tag{5.18}$$

where $\epsilon = (\epsilon_1, \cdots, \epsilon_j)$ *ranges over* $\epsilon_i =$ "*blank*" or "*", v *ranges over the indecomposable partitions* $(v_1, \cdots, v_p)$ *of the table*

$$\begin{matrix} (1,1) & \cdots & (1,k) \\ \vdots & \cdots & \vdots \\ (j,1) & \cdots & (j,k) \end{matrix} \tag{5.19}$$

and $d_{v_i}(s_{v_i})$ *denotes the joint cumulant of elements selected from the table*

$$\begin{matrix} Y_1^{\epsilon_1}(0) & \cdots & Y_k^{\epsilon_1}(0) \\ Y_1^{\epsilon_2}(s_2) & \cdots & Y_k^{\epsilon_2}(s_2) \\ \vdots & \cdots & \vdots \\ Y_1^{\epsilon_j}(s_j) & \cdots & Y_k^{\epsilon_j}(s_j) \end{matrix} \tag{5.20}$$

in accordance with v_i.

PROOF. Consider the case $j = 1$ first.

$$\begin{aligned} \mathfrak{C}_1(\alpha a_T + \alpha^* a_T^*) &= E(\alpha a_T + \alpha^* a_T^*) \\ &= \alpha E a_T + \alpha^*(E a_T)^* \end{aligned} \tag{5.21}$$

where

$$\mathrm{E} a_T = \sum_{t_1} \cdots \sum_{t_k} g_1(-t_1) \cdots g_k(-t_k) c_{1\cdots k}(t_1, \cdots, t_k), \tag{5.22}$$

giving the stated result. Next, if $j > 1$, using the result of Lemma 5.1 $\mathfrak{C}_j(\alpha a_T + \alpha^* a_T^*)$ equals

$$(2T+1)^{-j} \sum_{-T}^{T} \cdots \sum_{-T}^{T} \sum_v \sum_\epsilon \alpha^{\epsilon_1} \cdots \alpha^{\epsilon_j} d_{v_1}(t_{v_1}) \cdots d_{v_p}(t_{v_p}). \tag{5.23}$$

Taking advantage of the stationarity of the series involved, (5.23) may be written

$$\begin{aligned} (2T+1)^{-j} \sum_v \sum_\epsilon \sum_{s_2=-\infty}^{\infty} \cdots \sum_{s_j=-\infty}^{\infty} \sum_{t_1=-\infty}^{\infty} \alpha^{\epsilon_1} \cdots \alpha^{\epsilon_j} d_{v_1}(s_{v_1}) \\ \cdots d_{v_p}(s_{v_p}) \phi(t_1/T) \phi[(t_1+s_2)/T] \cdots \phi[(t_1+s_j)/T] \end{aligned} \tag{5.24}$$

where $\phi(x) = 1$ for $|x| \leqq 1$ and $= 0$ otherwise, and where $s_j = t_j - t_1$. In turn (5.24) equals

$$\begin{aligned} (2T+1)^{-j+1} \sum_v \sum_\epsilon \sum_{s_2} \cdots \sum_{s_j} \alpha^{\epsilon_1} \cdots \alpha^{\epsilon_j} d_{v_1}(s_{v_1}) \\ \cdots d_{v_p}(s_{v_p}) \Phi_T(s_2/T, \cdots, s_j/T) \end{aligned} \tag{5.25}$$

where $\Phi_T(s_2/T, \cdots, s_j/T)$ is given by

$$(2T+1)^{-1} \sum_t \phi(t/T) \phi[(t+s_2)/T] \cdots \phi[(t+s_j)/T]. \tag{5.26}$$

(5.26) may be seen to be measurable, uniformly bounded in T and convergent to 1. Taking advantage of the absolute summability result of Lemma 5.2, Lebesgue's bounded convergence theorem may be applied and the stated result seen to be true.

COROLLARY 5.1.1. *Under the assumptions of the theorem, a_T is asymptotically complex Gaussian with mean (5.22) and variance covariance matrix*

$$(2T+1)^{-1}\begin{Vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{Vmatrix} \tag{5.27}$$

where

$$r_{11} = \tfrac{1}{4}\textstyle\sum_v \sum_\epsilon \sum_s d_{v_1}(s_{v_1}) \cdots d_{v_p}(s_{v_p}), \tag{5.28}$$

$$r_{12} = (1/4i)\textstyle\sum_v \sum_s [d_{v_1}(s_{v_1}) \cdots d_{v_p}(s_{v_p}) - d^*_{v_1}(s_{v_1}) \cdots d^*_{v_p}(s_{v_p})], \tag{5.29}$$

$$r_{22} = \tfrac{1}{4}\textstyle\sum_v \sum_\epsilon \sum_s d_{v_1}(s_{v_1}) \cdots d_{v_p}(s_{v_p})\alpha^{\epsilon_1}\alpha^{\epsilon_2}, \tag{5.30}$$

with $\alpha = 1/i$ in (5.30). In (5.28), (5.30) $\epsilon = (\epsilon_1, \epsilon_2)$ extends over $\epsilon_i =$ "blank" or "". In (5.28), (5.29), (5.30), v extends over all indecomposable partitions selected from the table*

$$\begin{matrix} (1, 1) & \cdots & (1, k) \\ (2, 1) & \cdots & (2, k). \end{matrix} \tag{5.31}$$

In (5.28), (5.30) $d_{v_i}(s_{v_i})$ denotes the joint cumulant of elements selected from the table

$$\begin{matrix} Y_1^{\epsilon_1}(0) & \cdots & Y_k^{\epsilon_1}(0) \\ Y_1^{\epsilon_2}(s) & \cdots & Y_k^{\epsilon_2}(s) \end{matrix} \tag{5.32}$$

in accordance with v_i. In (5.29) $d_{v_i}(s_{v_i})$ denotes the joint cumulant of elements selected from the table

$$\begin{matrix} Y_1(0) & \cdots & Y_k(0) \\ Y_1(s) & \cdots & Y_k(s) \end{matrix} \tag{5.33}$$

in accordance with v_i.

(As in the case of this corollary table (5.31) has but two rows, v extends over partitions such that at least one set of the partition has an element from both row 1 and row 2.)

This corollary results from the fact that the cumulants of order >2 of $(2T+1)^{\frac{1}{2}}a_T$ tend to 0.

Let

$$b_T = (2T+1)^{-1}\textstyle\sum_{-T}^{T} Y_{i_1}(t) \cdots Y_{i_r}(t), \tag{5.34}$$

$$c_T = (2T+1)^{-1}\textstyle\sum_{-T}^{T} Y_{j_1}(t) \cdots Y_{j_s}(t). \tag{5.35}$$

COROLLARY 5.1.2. *Under the assumptions of the theorem, b_T and c_T are asymptotically joint complex Gaussian.*

This result may be demonstrated by considering the joint cumulants of b_T

and c_T. The result obviously extends to the case of more than two estimates as well.

COROLLARY 5.1.3. *Under the conditions of the theorem*

$$(2T+1)^{-1}\sum_{-T}^{T} Y_j(t) \tag{5.36}$$

is asymptotically complex Gaussian for all j.

LEMMA 5.3. *The estimate* $\hat{C}_{1\cdots k}$ *given in* (5.1) *may be written in the form*

$$(2T+1)^{-1}\sum_{-T}^{T} Y_1(t) \cdots Y_k(t) - \sum_{p>1} \hat{C}_{v_1} \cdots \hat{C}_{v_p}. \tag{5.37}$$

PROOF. $\hat{C}_{1\cdots k}$ and $\hat{n}$ (as defined in (5.2)) are actually the cumulant and moments of the random variable taking on the values

$$\{Y_1(t), \cdots, Y_k(t)\}, \qquad -T \leqq t \leqq T, \tag{5.38}$$

with probability $(2T+1)^{-1}$. Applying the moment-cumulant relation inverse to (3.4) to the variable (5.38) yields

$$\hat{n}_{1\cdots k} = \sum_{p\geqq 1} \hat{C}_{v_1} \cdots \hat{C}_{v_p} \tag{5.39}$$

where the summation extends over all partitions of the integers 1, 2, $\cdots$, k. The stated result is now evident.

THEOREM 5.2. *Consider the estimate* $\hat{C}_{1\cdots k}$ *given at* (5.1) *where the* $g_j(t)$ *are bounded*[4] *and absolutely summable and where the series* $\{X_1(t), \cdots, X_k(t)\}$ *satisfies Condition A.* $\hat{C}_{1\cdots k}$ *is asymptotically complex Gaussian with mean* (5.22) *and variance-covariance matrix* (5.27).

PROOF. Corollary 5.1.3 indicates that the stated result is true for $k = 1$. Lemma 5.4 yields the representation (5.37). We will use this representation to prove the stated result by means of induction. Suppose that the result is true for $K \leqq k-1$; therefore the $\hat{C}_v$ appearing on the right hand side of (5.37) are asymptotically normal. Now on consideration of $(2T+1)^{\frac{1}{2}}\hat{C}_{v_1} \cdots \hat{C}_{v_p}$, and the rate at which the $\hat{C}_v$ are tending to asymptotic normality, one sees that

$$\hat{C}_{v_1} \cdots \hat{C}_{v_p} = \prod_{j=1}^{p} \sum \cdots \sum \tilde{g}_{v_j}(-t_{v_j}) c_{v_j}(t_{v_j}) + o_p(2T+1)^{-\frac{1}{2}}. \tag{5.40}$$

Thus the asymptotic distribution of $\hat{C}_{1\cdots k}$ is the same as that of

$$(2T+1)^{-1}\sum_{-T}^{T} Y_1(t) \cdots Y_k(t) - \sum_{p>1} \prod_{j=1}^{p} \cdot\sum \cdots \sum \tilde{g}_{v_j}(-t_{v_j}) c_{v_j}(t_{v_j}). \tag{5.41}$$

The distribution of (5.41) was derived in Theorem 5.1.

COROLLARY 5.2.1. *Consider a pair of estimates* $\hat{C}_{v_1}$, $\hat{C}_{v_2}$ *of lower order polyspectra. Under the conditions of the theorem these estimates have asymptotically a joint complex Gaussian distribution.*

THEOREM 5.3. *In the frequency domain expressions* (5.22), (5.28), (5.29), (5.30) *take the form*

[4] The boundness follows from the absolute summability in fact.

(5.42) $$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \delta(w_1 + \cdots + w_k) G_1(w_1) \cdots G_k(w_k) C_{1\cdots k}(w_1, \cdots, w_k)\, dw_1 \cdots dw_k,$$

(5.43) $$\tfrac{1}{4} \int_{w_1} \int_{w_2} G_1(w_{11}) \cdots G_k(w_{1k}) G_1(w_{21}) \cdots G_k(w_{2k}) \sum_\epsilon \sum_v D_{v_1}(w_{v_1}) \cdots D_{v_p}(w_{v_p}),$$

(5.44) $$(1/4i) \int_{w_1} \int_{w_2} [G_1(w_{11}) \cdots G_k(w_{1k}) G_1(w_{21}) \cdots G_k(w_{2k})] \cdot [\textstyle\sum_v \delta(\tilde{w}_{v_1}) C_{v_1}(w_{v_1}) \cdots \delta(\tilde{w}_{v_p}) C_{v_p}(w_{v_p}) - \delta(-\tilde{w}_{v_1}) C_{v_1}(-w_{v_1}) \cdots \delta(-\tilde{w}_{v_p}) C_{v_p}(-w_{v_p})]\, dw_{11} \cdots dw_{2k},$$

(5.45) $$\tfrac{1}{4} \int_{w_1} \int_{w_2} G_1(w_{11}) \cdots G_k(w_{1k}) G_1(w_{21}) \cdots G_k(w_{2k}) \sum_\epsilon \sum_v D_{v_1}(w_{v_1}) \cdots D_{v_p}(w_{v_p}) \alpha^{\epsilon_1} \alpha^{\epsilon_2}, \quad \alpha = 1/i.$$

In (5.43), (5.45) $\epsilon = (\epsilon_1, \epsilon_2)$ *extends over* $\epsilon_i =$ "+1" or "−1". *In* (5.43), (5.44), (5.45), v *extends over all indecomposable partitions selected from table* (5.31). *In* (5.43) *and* (5.45) $D_{v_i}(w_{v_i})$ *denotes the joint cumulant of elements selected from the table*

(5.46) $$\begin{matrix} dZ_1(\epsilon_1 w_{11}) \cdots dZ_k(\epsilon_1 w_{1k}) \\ dZ_1(\epsilon_2 w_{21}) \cdots dZ_k(\epsilon_2 w_{2k}) \end{matrix}$$

in accordance with v_i. $Z_i(w)$ *comes from the spectral representation of* $X_i(t)$, *and* $g_j(t)$ *is given by*

(5.47) $$(2\pi)^{-1} \int_{-\pi}^{\pi} e^{iwt} G_j(w)\, dw$$

with $G_j(w)$ *real.*

Proof. This result may be proved by noting that

(5.48) $$Y_j(t) = \int_{-\pi}^{\pi} e^{iwt} G_j(w)\, dZ_j(w).$$

The reader will have noted that throughout this section a limiting process leading to estimates of averaged polyspectra (as in (5.42)) was used rather than one actually leading to $C_{1\cdots k}(w_1, \cdots, w_k)$. One could obtain $C_{1\cdots k}$ in the limit by letting the g_j's employed depend on T; however it is felt that the limiting procedure employed yields results more representative of the finite T case.

It is perhaps of interest to mention the paper [1] where it is shown that the moment estimates derived from a stationary Gaussian process are asymptotically Gaussian.

6. Applications of the theory. The intention of this section is to present a number of situations in which the estimation of polyspectra or associated polyspectral coefficients may be of use.

Suppose that we are interested in a real-valued time series $X(t)$. Are we wise to carry out a harmonic analysis of $X(t)$ or does some function of $X(t)$, say $\log X(t)$, have a simpler harmonic analysis? This question may be answered to a limited extent by evaluating certain polyspectra. To begin, note that many functional relationships may be approximated by relationships of the form

$$x = y + \alpha y^{k-1}, \tag{6.1}$$

where α is small. Consider the time series relationship

$$X(t) = Y(t) + \alpha[Y(t)]^{k-1}, \tag{6.2}$$

where α is small and $Y(t)$ is a simpler series than $X(t)$; simpler in the sense that cumulants of order j, $2 < j \leqq k$ are negligible. Evaluating the $(k - 1)$th order polyspectrum of $X(t)$ in terms of the polyspectra of $Y(t)$, using (6.2) and retaining terms of first and lower order gives

$$C(w_1, \cdots, w_k) = \alpha(k-1)\sum f(w_{j_1}) \cdots f(w_{j_{k-1}}) \tag{6.3}$$

where $C(w_1, \cdots, w_k)$ denotes the $(k-1)$th order polyspectrum of $X(t)$, $f(w)$ denotes the power spectrum of $X(t)$, and in (6.3) the summation extends over the indices $1, \cdots, k$ taken $(k-1)$ at a time. (Remember that in the case of a single series $X(t)$, when we are considering the $(k-1)$th order polyspectrum we are really thinking of the series as $\{X(t), \cdots, X(t)\}$.) Thus we see that if a relationship of the form (6.2) holds, α is given approximately by

$$C(w_1, \cdots, w_k)/(k-1)\sum f(w_{j_1}) \cdots f(w_{j_{k-1}}). \tag{6.4}$$

This coefficient may be estimated by substituting estimates of the $(k-1)$th order polyspectrum and the power spectrum of $X(t)$ into (6.4).

In this connection we have,

Theorem 6.1. *Let $X(t)$ denote a time series satisfying the conditions of Theorems 5.2, 5.3. Let $\hat{C}(w_1, \cdots, w_k)$, $\hat{f}_j(w_j)$ denote estimates of $C(w_1, \cdots, w_k)$, $f(w_j)$ respectively of the form of the estimates of Theorem 5.2. The random variable*

$$\hat{C}(w_1, \cdots, w_k)/(k-1)\sum \hat{f}_{j_1}(w_{j_1}) \cdots \hat{f}_{j_{k-1}}(w_{j_{k-1}}) \tag{6.5}$$

tends to

$$\frac{\int \cdots \int \delta(w_1 + \cdots + w_k) G_1(w_1) \cdots G_k(w_k) C(w_1, \cdots, w_k)\, dw_1 \cdots dw_k}{(k-1) \sum h_{j_1}(w_{j_1}) \cdots h_{j_{k-1}}(w_{j_{k-1}})} \tag{6.6}$$

in probability, where

$$h_j(w_j) = \int_{-\pi}^{\pi} |G_j(w)|^2 f(w)\, dw. \tag{6.7}$$

(6.5) is also asymptotically complex Gaussian.

Proof. This theorem results from Theorem 5 and Corollary 3 of [19] and Theorems 5.2 and 5.3 of this paper.

Turning to another application of the theory, consider the following heuristic model of a frequency component being produced by the beating or multiplication together of a number of individual frequency components. Suppose we are considering real-valued time series $X_1(t), \cdots, X_k(t)$ with spectral representations

$$X_j(t) = \int e^{iwt}\, dZ_j(w). \tag{6.8}$$

Consider the following question, does the component at frequency w_k in the series $X_k(t)$ come about as the product of the components at frequencies $-w_j$ in the series $X_j(t)$, $j = 1, \cdots, k-1$, where $\sum w_j = 0$?

In terms of the spectral functions $Z_j(w)$, we are wondering if $\Delta Z_k(w_k)$ is of the approximate form

$$\beta \Delta Z_1(-w_1) \cdots \Delta Z_{k-1}(-w_{k-1}) \tag{6.9}$$

for some constant β where

$$\Delta Z_j(w_j) = Z_j(w_j + \Delta w_j) - Z_j(w_j). \tag{6.10}$$

Since the series involved are real, (6.9) may be written

$$\beta \Delta Z_1^*(w_1) \cdots \Delta Z_{k-1}^*(w_{k-1}). \tag{6.11}$$

The linear regression coefficient of $\Delta Z_k(w_k)$ on (6.10) is therefore

$$E\Delta Z_1(w_1) \cdots \Delta Z_k(w_k)/E|\Delta Z_1(w_1) \cdots \Delta Z_{k-1}(w_{k-1})|^2 \tag{6.12}$$

and the coefficient of determination is

$$\frac{|E\, \Delta Z_1(w_1) \cdots \Delta Z_k(w_k)|^2}{E\, |\Delta\, Z_1(w_1) \cdots \Delta Z_{k-1}(w_{k-1})|^2\, E\, |\Delta Z_k(w_k)|^2}. \tag{6.13}$$

If the w_j satisfy no relation of the form

$$\textstyle\sum_{i=1}^{s} w_{j_i} = 0 \qquad (s < k) \tag{6.14}$$

and the Δw_j are small then (6.12) and (6.13) are given by

$$C_{1\cdots k}(w_1, \cdots, w_k)/f_1(w_1) \cdots f_{k-1}(w_{k-1}) \tag{6.15}$$

and

$$[|C_{1\cdots k}(w_1, \cdots, w_k)|^2/f_1(w_1) \cdots f_k(w_k)]\cdot[|\Delta w_1 \cdots \Delta w_{k-1}|/|\Delta w_k|], \tag{6.16}$$

respectively.

Thus when one is considering the question of frequency components beating together, one is led to consider the coefficients (6.15) and

$$|C_{1\cdots k}(w_1, \cdots, w_k)|^2/f_1(w_1) \cdots f_k(w_k). \tag{6.17}$$

This latter represents the relative appropriateness at various polyfrequencies of the beating together of frequency components model. (Relative because of the additional factor in (6.16).)

These coefficients may be estimated by substituting estimates of the required polyspectra, and we may prove,

THEOREM 6.2. *Let* $\{X_1(t), \cdots, X_k(t)\}$ *denote a time series satisfying the conditions of Theorems* 5.2, 5.3. *Let* $\hat{C}_{1\cdots k}(w_1, \cdots, w_k)$, $\hat{f}_j(w_j)$ *denote estimates of*

$C_{1\cdots k}(w_1, \cdots, w_k)$, $f_j(w_j)$ *respectively of the form of the estimates considered in Theorem* 5.2. *The random variables*

$$\hat{C}_{1\cdots k}(w_1, \cdots, w_j)/\hat{f}_1(w_1) \cdots \hat{f}_{k-1}(w_{k-1}) \tag{6.18}$$

and

$$|\hat{C}_{1\cdots k}(w_1, \cdots, w_k)|^2/\hat{f}_1(w_1) \cdots \hat{f}_k(w_k) \tag{6.19}$$

tend in probability to

$$\frac{\int \cdots \int \delta(w_1 + \cdots + w_k) G_1(w_1) \cdots G_k(w_k) C_{1\cdots k}(w_1, \cdots, w_k)\, dw_1 \cdots dw_k}{h_1(w_1) \cdots h_{k-1}(w_{k-1})} \tag{6.20}$$

and

$$\frac{\left|\int \cdots \int \delta(w_1 + \cdots + w_k) G_1(w_1) \cdots G_k(w_k) C_{1\cdots k}(w_1, \cdots w_k)\, dw_1 \cdots dw_k\right|^2}{h_1(w_1) \cdots h_k(w_k)}, \tag{6.21}$$

respectively where

$$h_j(w_j) = \int |G_j(w)|^2 f_j(w)\, dw. \tag{6.22}$$

Moreover, asymptotically the estimates are joint complex Gaussian.

PROOF. The proof proceeds on the same lines as the proof of Theorem 6.1.

It is perhaps of interest to point out the values of (6.15) and (6.17) in the case of one of the examples considered earlier. Suppose $X(t)$ is the process of Example 3, Section 3. In this case (6.15) and (6.17) are given by

$$[K_k/(K_2)^{k-1}] \cdot G_k(w_k)/|G_1(w_1) \cdots G_{k-1}(w_{k-1})| \tag{6.23}$$

and

$$K_k^2/K_2^k \tag{6.24}$$

and we see that an examination of the coefficients (6.19) for constancy provides a test for the model of this example.

The reader will have noted that in the derivation of the coefficients (6.15) and (6.17) it was assumed that the w_j satisfy no relation of the form (6.14). This assumption is reasonable in view of the fact that if the process satisfies an ergodicity requirement to be presented in the next section, then components whose frequencies are such that (6.14) is true, are uncorrelated with the remaining components and a relation of the form (6.9) is then inconsistent.

7. Moments or cumulants? At this point the reader is no doubt wondering why the polyspectrum was defined as the Fourier transform of the cumulant rather than of the product moment or of the central product moment. In this section a justification of this definition will be provided for a class of processes. The essential property that these processes have is a form of ergodicity.

Let us begin by noting that the Fourier transform of at most one of the product moment, central product moment or cumulant can be "nice" in the sense of being a proper function. Suppose for example that the polyspectra are proper functions and consider the Fourier transform of $m_{1\cdots k}(t_1, \cdots, t_k)$. As derived from the relation inverse to (3.4) it is,

$$(7.1) \quad \delta(w_1 + \cdots + w_k)M(w_1, \cdots, w_k) = \sum \delta(\tilde{w}_{v_1}) \cdots \delta(\tilde{w}_{v_p})C_{v_1}(w_{v_1}) \cdots C_{v_p}(w_{v_p})$$

where $\tilde{w}_v = w_{i_1} + \cdots + w_{i_j}$ if v denotes the grouping $(i_1, \cdots, i_j)$. $M(w_1, \cdots, w_k)$ is seen to contain many delta functions if the lower order polyspectra do not vanish (as the ordinary power spectrum must not). Thus we see that if the polyspectra are proper functions, then the Fourier transforms of the product moments are not. The converse of this statement may be seen to be true by considering the expansion (3.4). By considering similar expansions involving central moments, we are led to the conclusion that at most one of the definitions may lead to proper functions.

It will now be shown that for processes satisfying a form of ergodicity requirement, the property of having a proper function as a polyspectrum is not evidently inconsistent, whereas the corresponding property for moments and central moments is inconsistent. The class $\Psi^{(k)}$ introduced earlier is thus perhaps a reasonable one so far as ergodic type processes are concerned.

The following notation will be adhered to in the remainder of this section:

(a) if v denotes a group of distinct integers $(i_1, \cdots, i_j)$ selected from $(1, \cdots, k)$, then $\tilde{X}_v(t_v)$ denotes the product $X_{i_1}(t_{i_1}) \cdots X_{i_j}(t_{i_j})$,

(b) if u, v denote distinct groupings, then the refinement grouping obtained by inserting the subdivisions of u into v will be denoted by $u \otimes v$,

(c) if u is the grouping (u_1, u_2) and $t = (t_1, \cdots, t_k)$, then $\bar{t}$ will denote $(\bar{t}_1, \cdots, \bar{t}_k)$ where $\bar{t}_i = t_i + \tau$ if $i \varepsilon u_1$ and $\bar{t}_i = t_i$ if $i \varepsilon u_2$.

The process $X(t) = \{X_1(t), \cdots, X_k(t)\}$ is said to satisfy *Condition I*(k) if the joint moments of order $\leqq k$ exist, and for all groupings u and v (u consisting of two subgroups), and the X's corresponding to the different subgroups v_j of v being from independent realizations of $X(t)$,

$$(7.2) \qquad (2T)^{-1} \int_{-T}^{T} U(\tau)\, d\tau$$

approaches

$$(7.3) \qquad m_{r_1}(t_{r_1}) \cdots m_{r_p}(t_{r_p})$$

in probability where $r = u_1 \otimes v$ and $U(\tau)$ denotes the product of the individual X terms in $\tilde{X}_{v_1}(\bar{t}_{v_1}) \cdots \tilde{X}_{v_p}(\bar{t}_{v_p})$ involving τ.

Condition I(k) is seen to be a form of ergodicity requirement. In fact if we are concerned with a univariate weakly mixing process $X(t)$ belonging to $\Phi^{(\infty)} \cap S^{(\infty)}$, then $X(t)$ satisfies Condition I(k) for every k (see [7]).

The process $X(t) = \{X_1(t), \cdots, X_k(t)\}$ is said to satisfy *Condition II*(k) if,

(i) there exists $\delta > 0$ such that for $j \leqq k$ and distinct indices $i_1, \cdots, i_k$

selected from $1, \cdots, k$,

$$E|X_{i_1}(t_1) \cdots X_{i_j}(t_j)|^{1+\delta} < \infty, \tag{7.4}$$

(ii) there exists T_0 and $M > 0$ such that for $T > T_0$,

$$(2T)^{-1} \int_{-T}^{T} E_1|\tilde{X}_{v_1}(\bar{t}_{v_1})|^{1+\delta} \cdots E_p|\tilde{X}_{v_p}(\bar{t}_{v_p})|^{1+\delta} \, d\tau < M \tag{7.5}$$

for all groupings v, where the subscripts on the expected value operators denote independent realizations of the process.

The following lemma will be required.

LEMMA 7.1. *Let $\{U_n\}$ be a sequence of random variables tending to μ in probability. Let V be a random variable such that* (i) *for some $\delta > 0$, $E|V|^{1+\delta}$ exists,* (ii) *there exist N and $M > 0$ such that for $n > N$, $E|U_nV|^{1+\delta} < M$, then $EU_nV \to \mu EV$.*

PROOF.

$$\begin{aligned} |EU_nV - \mu EV| &= |E(U_n - \mu)V| \leqq E|(U_n - \mu)V| \\ &= \int |U_n - \mu| \cdot |V| \, dP_n(U, V) \\ &= \int_{|U-\mu| \leqq \epsilon} |U - \mu| \cdot |V| \, dP_n(U, V) \\ &\quad + \int_{|U-\mu| > \epsilon} |U - \mu| \cdot |V| \, dP_n(U, V) \end{aligned} \tag{7.6}$$

where $P_n(U, V)$ denotes the joint cdf of U_n and V. The first term in (7.6) is $\leqq \epsilon E|V|$ and consequently may be made arbitrarily small by a choice of ϵ. The second term is less than or equal to

$$\{\int_{|U-\mu|>\epsilon} dP_n(U, V)\}^{\delta/(1+\delta)} \{\int |U - \mu|^{1+\delta} |V|^{1+\delta} \, dP_n(U, V)\}^{1/(1+\delta)}. \tag{7.7}$$

The first term in (7.7) may be made arbitrarily small as a result of the convergence in probability of $\{U_n\}$ to μ, while the second term remains bounded. Consequently (7.6) may be made arbitrarily small and the lemma follows.

THEOREM 7.1. *Consider the process $X(t) = \{X_1(t), \cdots, X_k(t)\}$ that satisfies Conditions $I(k)$ and $II(k)$. For any groupings $(v_1, \cdots, v_p)$ and (u_1, u_2) of $(1, \cdots, k)$,*

$$\begin{aligned} \lim_{T\to\infty} (2T)^{-1} \int_{-T}^{T} m_{v_1}(\bar{t}_{v_1}) \cdots m_{v_p}(\bar{t}_{v_p}) \, d\tau \\ = m_{r_1}(t_{r_1}) \cdots m_{r_p}(t_{r_p}) m_{s_1}(t_{s_1}) \cdots m_{s_p}(t_{s_p}) \end{aligned} \tag{7.8}$$

where $r = u_1 \otimes v$ and $s = u_2 \otimes v$.

PROOF.

$$\begin{aligned} (2T)^{-1} \int_{-T}^{T} m_{v_1}(\bar{t}_{v_1}) \cdots m_{v_p}(\bar{t}_{v_p}) \, d\tau \\ = (2T)^{-1} \int_{-T}^{T} E_1\tilde{X}_{v_1}(\bar{t}_{v_1}) \cdots E_p\tilde{X}_{v_p}(\bar{t}_{v_p}) \, d\tau \end{aligned} \tag{7.9}$$

$$= E_1 \cdots E_p(2T)^{-1} \int_{-T}^{T} \tilde{X}_{v_1}(\bar{t}_{v_1}) \cdots \tilde{X}_{v_p}(\bar{t}_{v_p}) \, d\tau, \tag{7.10}$$

since under the stated conditions Tonelli's theorem applies.

The result now follows from the lemma taking U_T to be (7.2) and $U(\tau)$ to be the product of the individual X terms in $\tilde{X}_{v_1}(\bar{t}_{v_1}) \cdots \tilde{X}_{v_p}(\bar{t}_{v_p})$ involving τ.

Let us next prove,

THEOREM 7.2. *Consider the process* $X(t) = \{X_1(t), \cdots, X_k(t)\}$ *satisfying Conditions* $I(k)$ *and* $II(k)$.

$$\lim_{T\to\infty} (2T)^{-1} \int_{-T}^{T} c_{1\cdots k}(\bar{t})\, d\tau = 0 \tag{7.11}$$

for any grouping (u_1, u_2).

PROOF.

$$c_{1\cdots k}(\bar{t}) = \sum (-1)^{p-1}(p-1)!\, m_{v_1}(\bar{t}_{v_1}) \cdots m_{v_p}(\bar{t}_{v_p}). \tag{7.12}$$

Thus,

$$\begin{aligned}(2T)^{-1} &\int_{-T}^{T} c_{1\cdots k}(\bar{t})\, d\tau \\ &= \sum (-1)^{p-1}(p-1)!\, (2T)^{-1} \int_{-T}^{T} m_{v_1}(\bar{t}_{v_1}) \cdots m_{v_p}(\bar{t}_{v_p})\, d\tau.\end{aligned} \tag{7.13}$$

From the preceding theorem this tends to

$$\sum (-1)^{p-1}(p-1)!\, m_{r_1}(t_{r_1}) \cdots m_{r_p}(t_{r_p}) m_{s_1}(t_{s_1}) \cdots m_{s_p}(t_{s_p}) \tag{7.14}$$

where $r = u_1 \otimes v$ and $s = u_2 \otimes v$.

We note that (7.14) is the joint kth order cumulant of the process $X(t) = \{X_1(t), \cdots, X_k(t)\}$ wherein the components with subscripts in u_1 are statistically independent of those with subscripts in u_2. The expression must consequently be 0 as this cumulant is 0.

Before proceeding to the next theorem, let us make one last definition:

$\Phi_k^{(k)}$ denotes the class of k-dimensional processes $X(t) = \{X_1(t), \cdots, X_k(t)\}$ with finite kth order absolute moments and such that for $v = (i_1, \cdots, i_j)$ any group of j distinct integers from $1, \cdots, k$ there exist complex totally finite measures $M_v(\Omega)$ such that

$$\begin{aligned}EX_{i_1}(t_1) &\cdots X_{i_j}(t_j) \\ &= \int \cdots \int \exp[i(w_1t_1 + \cdots + w_jt_j)] M_v(dw_1, \cdots, dw_j).\end{aligned} \tag{7.15}$$

As in [24] it is possible to introduce in an obvious manner a polyspectral measure

$$\sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(\Omega) \tag{7.16}$$

for this class where Ω is a Borel set of R^k.

THEOREM 7.3. *Consider the process* $X(t) = \{X_1(t), \cdots, X_k(t)\}$ *belonging to* $\Phi_k^{(k)}$ *and satisfying Conditions* $I(k)$ *and* $II(k)$. *Given the grouping* (u_1, u_2), *let* Ω_1 *be the flat* $\tilde{w}_{u_1} = 0$, Ω_2 *the flat* $\tilde{w}_{u_2} = 0$, *and* Ω *a measurable subset of* $\Omega_1 \times \Omega_2$, *then*

$$\sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(\Omega) = 0. \tag{7.17}$$

PROOF.

$$\begin{aligned}c_{1\cdots k}(t_1, \cdots, t_k) = \int \cdots \int &\exp[i(t_1w_1 + \cdots + t_kw_k)] \\ &\cdot \sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(dw),\end{aligned} \tag{7.18}$$

$$(7.19)\quad (2T)^{-1}\int_{-T}^{T} c_{1\cdots k}(t_1, \cdots, t_k)\, d\tau = \int \cdots \int \exp\,[i(t_1w_1 + \cdots + t_kw_k)] \cdot \sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(dw)(2T)^{-1}\int_{-T}^{T} \exp\,[i\tilde{w}_{v_1}\tau]d\tau.$$

But

$$(7.20)\quad \lim_{T\to\infty} (2T)^{-1}\int_{-T}^{T} \exp\,[i\tilde{w}_{v_1}\tau]d\tau = \epsilon(\tilde{w}_{v_1}) = 1 \qquad \tilde{w}_{v_1} = 0$$
$$= 0 \qquad \tilde{w}_{v_1} \neq 0.$$

Thus we have

$$(7.21)\quad \int \cdots \int \exp\,[i(t_1w_1 + \cdots + t_kw_k)]\epsilon(\tilde{w}_{v_1}) \cdot \sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(dw) = 0 \quad \text{for all } t$$

and

$$(7.22)\quad \int \cdots \int \chi(\Omega)\epsilon(\tilde{w}_{v_1}) \sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(dw) = 0$$

for any measurable set Ω of the flat $\sum w_i = 0$, where $\chi(\Omega)$ is the characteristic function of Ω, and we see that

$$\sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(\Omega) = 0$$

if in fact $\Omega \subseteq \Omega_1 \times \Omega_2$.

This argument parallels an argument in [25].

Now the Lebesgue measure of the set Ω of this theorem is 0, consequently the measure $\sum (-1)^{p-1}(p-1)!\, M_{v_1} \times \cdots \times M_{v_p}(dw)$ satisfies a necessary condition for it to be absolutely continuous with respect to $(k-1)$-dimensional Lebesgue measure. We conclude that the polyspectrum, which is an attempt to provide a density of this measure with respect to Lebesgue measure, is not evidently inconsistent.

The ergodicity of stationary processes is also investigated in [26].

A different type of justification of the use of cumulants is the following; in the Gaussian case all the information is contained in the first two moments. Consequently a kth order product moment $k > 2$, has no new information to provide, nor does its Fourier transform. The kth order cumulant is a function of the product moments of orders k and less which is zero in the Gaussian case. The consideration of the cumulant in this case is not liable to deceive one into believing that he has gained some information. In the non-Gaussian case the cumulant provides an indication of the non-Gaussianity. The cumulants appear to provide a form of harmonic analysis of the distribution in fact.

It seems appropriate to end the paper on a note of pessimism. Experience with real random variables indicates that higher order moments are typically not efficient estimates of scientifically relevant parameters; consequently as the specifications of stochastic processes become tighter, polyspectra are likely to prove less pertinent in a similar manner.

8. Acknowledgment. The author would like to acknowledge many conversations with Professor John W. Tukey and Dr. Michael Godfrey that were most helpful in the preparation of this paper.

REFERENCES

[1] BARTLETT, M. S. (1961). *An Introduction to Stochastic Processes.* Cambridge Univ. Press.

[2] BLACKMAN, R. B. and TUKEY, J. W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering.* Dover, New York.

[3] BLANC-LAPIERRE, A. and FORTET, R. (1953). *Theorie des Fonctions Aleatoires.* Masson, Paris.

[4] BOCHNER, S. (1936). Summation of multiple Fourier series by spherical means. *Trans. Amer. Math. Soc.* **40** 175–207.

[5] BOCHNER, S. and CHANDRASEKHARAN, K. (1949). *Fourier Transforms.* Princeton Univ. Press.

[6] BRILLINGER, D. R. (1964). A technique for measuring the spectral density matrix of two signals. *Proc. I.E.E.E.* **52** 103–104.

[7] BRILLINGER, D. R. (1965). The ergodicity of stationary processes. Submitted to *Theor. Prob. Appl.*

[8] BRILLINGER, D. R. and HATANAKA, M. (1965). An harmonic analysis of non-stationary time series with special reference to the Permanent Income Hypothesis. To be submitted to *Econometrica.*

[9] CRAMÉR, H. (1942). On harmonic analysis in certain functional spaces. *Ark. Mat. Astr. Fys.* **28B**.

10] GOODMAN, N. R. (1957). On the joint estimation of the spectra, cospectrum and quadrature spectrum of a two-dimensional stationary Gaussian process. Scientific Paper No. 10, Engineering Statistics Laboratory, New York University.

[11] HAMMING, R. W. (1962). *Numerical Methods for Scientists and Engineers.* McGraw-Hill, New York.

[12] HASSELMAN, K., MUNK, W., and MACDONALD, G. (1963). Bispectrum of ocean waves. *Time Series Analysis,* (M. Rosenblatt ed.) 125–139. Wiley, New York.

[13] JACKSON, D. (1930). *The Theory of Approximation.* Amer. Math. Soc., Providence.

[14] JENKINS, G. M. (1961). General considerations in the analysis of spectra. *Technometrics* **3** 133–166.

[15] JENKINS, G. M. (1963). Cross-spectral analysis and the estimation of linear open loop transfer functions. *Time Series Analysis,* (M. Rosenblatt ed.) 267–278. Wiley, New York.

[16] KENDALL, M. G. and STUART, A. (1958). *The Advanced Theory of Statistics,* **1**. Griffin, London.

[17] LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. *Theor. Prob. Appl.* **4** 319–329.

[18] MAGNESS, T. A. (1954). Spectral response of a quadratic device to non-Gaussian noise. *J. Appl. Phys.* **25** 1357–1365.

[19] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[20] MAZUR, S. and ORLICZ, W. (1934). Grundlegende Eigenschaften der polynomischen Operationen. *Studia Math.* **5** 50–68 and 170–189.

[21] PAPOULIS, P. (1962). *The Fourier Integral and Its Applications.* McGraw-Hill, New York.

[22] PARZEN, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics* **3** 167–190.

[23] ROSENBLATT, M. and VAN NESS, J. W. (1964). Estimates of the bispectrum of stationary random processes. Brown University Technical Report. Contract NONR562(29).

[24] SHIRYAEV, A. N. (1960). Some problems in the spectral theory of higher-order moments, I. *Theor. Prob. Appl.* **5** 265–284.

[25] SHIRYAEV, A. N. (1963). On conditions for ergodicity of stationary processes in terms of higher order moments. *Theor. Prob. Appl.* **8** 436–439.

[26] SINAI, YA. G. (1963). On higher order spectral measures of ergodic stationary processes. *Theor. Prob. Appl.* **8** 429–436.

[27] SUN, T. C. (1963). A central limit theorem for non-linear functions of a normal stationary processes. *J. Math. Mech.* **12** 945–978.

[28] TUKEY, J. W. (1959). An introduction to the measurement of spectra. *Probability and Statistics*, (U. Grenander ed.) 300–330. Wiley, New York.

[29] TUKEY, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3** 191–219.

[30] TUKEY, J. W. (1963). *An Introduction to the Frequency Analysis of Time Series*. Notes on Mathematics 596, course offered at Princeton University.

[31] VAN DER TWEEL, L. H. (1961). Some problems in vision regarded with respect to linearity and frequency response. *Ann. New York Acad. Sci.* **89** 829–856.

[32] WIENER, N. (1958). *Nonlinear Problems in Random Theory*. M.I.T. Press, Cambridge.

ASYMPTOTIC THEORY OF ESTIMATES OF kTH-ORDER SPECTRA*

BY D. R. BRILLINGER AND M. ROSENBLATT†

LONDON SCHOOL OF ECONOMICS, UNIVERSITY OF CALIFORNIA, SAN DIEGO (LA JOLLA), AND UNIVERSITY COLLEGE, LONDON

Communicated by John W. Tukey, December 5, 1966

1. *Notation and Assumptions.*—Let $X(t)$ be a strictly stationary r-vector valued process with real-valued components. All moments are assumed to exist. While t may be discrete or continuous, principal results are stated only in the discrete case (t running through the integers) though the obvious parallels in the continuous case are valid, provided $X(t)$ is assumed to be continuous in mean square.

Existence of second-order moments implies that $X(t)$ has a vector-valued Fourier representation

$$X(t) = \int e^{it\lambda}\, dZ(\lambda) \tag{1.1}$$

in mean square with $Z(\lambda)$ an r-vector valued process with orthogonal increments. Let $c(Z) = c(Z_1,\ldots, Z_k)$ denote the kth-order cumulant of $Z = (Z_1,\ldots,Z_k)$, and for $a = (a_1,\ldots,a_k)$, $t = (t_1,\ldots,t_k)$ let

$$c_{k,a}(t) = c_{a_1,\ldots,a_k}(t_1,\ldots,t_k) = c_{a_1,\ldots,a_k}(\tau + t_1,\ldots,\tau + t_k) = c(X_a(t)) \tag{1.2}$$

be the kth-order cumulant of $X_a(t) = (X_{a_1}(t_1),\ldots,X_{a_k}(t_k))$. Using stationarity, write its asymmetric form as

$$c'_{k,a}(t') = c_{a_1,\ldots,a_k}(\tau + t_1,\ldots,\tau + t_{k-1},\tau), \tag{1.3}$$

where $t' = (t_1,t_2,\ldots,t_{k-1})$.

ASSUMPTION I. *For each $j = 1,\ldots,k - 1$ and any k-tuple $a_1,\ldots,a_k$ let*

$$\sum_{t'} \left| t_j c'_{k,a}(t')\right| < \infty, \tag{1.4}$$

where $k = 2,3,\ldots$.

Assumption I implies that all cumulant spectral densities

$$f'_{k,a}(\omega') = f'_{a_1,\ldots,a_k}(\omega_1,\ldots,\omega_{k-1}) = f_{k,a}(\omega) = f_{a_1,\ldots,a_k}(\omega_1,\ldots,\omega_k)$$
$$= (2\pi)^{-k+1} \sum_{t'} c'_{k,a}(t') \exp\left(-i \sum_{j=1}^{k-1} t_j\omega_j\right) \tag{1.5}$$

exist where it is understood that

$$\sum_{j=1}^{k} \omega_j \equiv 0 \mod 2\pi \quad j = 1,\ldots,k \tag{1.6}$$

in formula (1.5). Further, Assumption I implies that all the cumulant spectral densities are continuous and continuously differentiable. The cumulant

$$c(dZ_a(\omega)) = f_{k,a}(\omega)\eta\left(\sum_1^k \omega_j\right)d\omega, \tag{1.7}$$

where $\eta(x) = \sum_{-\infty}^{\infty} \delta(x + 2j\pi)$ with $\delta(x)$ the Dirac delta function and $d\omega = \prod_1^k d\omega_j$.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_11, © Springer Science+Business Media, LLC 2012

Let

$$d_{a_j}^{(T)}(\lambda_j) = \sum_{t=0}^{T-1} X_{a_j}(t) \exp(-i\lambda_j t) \quad (1.8)$$

and

$$I_{k,a}^{(T)}(\lambda) = (2\pi)^{-k+1}T^{-1} \prod_{j=1}^{k} d_{a_j}^{(T)}(\lambda_j) \quad (1.9)$$

with $\sum_{j=1}^{k} \lambda_j \equiv 0 \mod 2\pi$. The function $I_{k,a}^{(T)}(\lambda)$ is a kth-order analogue of the second-order periodogram and cross-periodogram. The following lemma is useful in analyzing the asymptotic behavior of a kth-order periodogram and estimates of kth-order spectra based on the periodogram.

LEMMA. *Suppose Assumption I is satisfied. Then the cumulant*

$$c(d_{a_1}^{(T)}(\lambda_1),\ldots d_{a_k}^{(T)}(\lambda_k)) = (2\pi)^{-k+1}f'_{k,a}(\lambda') \sum_{t=0}^{T-1} \exp(-i \sum_{1}^{k} \lambda_j) + 0(1), \quad (1.10)$$

where the error term $0(1)$ *is uniform in* $\lambda_1,\ldots,\lambda_k$ *as* $T \rightarrow \infty$.

2. *A Class of Estimates.*—Let $W(u)$ be a bounded continuous weight function on the plane $\sum_{1}^{k} u_j = 0$ symmetric about zero, $W(u) = W(-u)$ with

$$\int W(u)\delta(\sum_{1}^{k} u_j)du = 1. \quad (2.1)$$

ASSUMPTION II. *Let* $W(u)$ *be continuously differentiable with*

$$|u_j W(u)|, \left|\frac{\partial}{\partial u_j} W(u)\right| \leq M(1 + ||u||)^{-k+1-\epsilon}, \quad (2.2)$$

$||u|| = (\sum_{1}^{k} u_j^2)^{1/2}$, *uniformly in* j *and* $u = (u_1,\ldots,u_k)$ *where* $M, \epsilon > 0$. *Set*

$$W_T(u) = B_T^{-k+1} \Sigma' W(B_T^{-1}(u + 2\pi j)), \quad (2.3)$$

where the summation is over $j = (j_1,\ldots,j_k)$ *such that* $\sum_{1}^{k} (j_\alpha + u_\alpha) = 0$ *and* B_T *is chosen so that* $B_T \rightarrow 0$ *as* $T \rightarrow \infty$ *but* $B_T^{k-1}T \rightarrow \infty$ *as* $T \rightarrow \infty$.

The estimates of $f_{k,a}(\lambda)$, $\sum_{1}^{k} \lambda_j \equiv 0 \mod 2\pi$, that we consider take the form

$$f_{k,a}^{(T)}(\lambda) = \left(\frac{2\pi}{T}\right)^{k-1} \sum_{s_1,\ldots,s_k=0}^{T-1} W_T\left(\lambda_1 - \frac{2\pi s_1}{T},\ldots,\lambda_k - \frac{2\pi s_k}{T}\right) \times \Phi\left(\frac{2\pi s_1}{T},\ldots,\frac{2\pi s_k}{T}\right) I_{k,a}^{(T)}\left(\frac{2\pi s_1}{T},\ldots,\frac{2\pi s_k}{T}\right), \quad (2.4)$$

where $\Phi(u)$ is zero unless u is on the manifold $\sum_{1}^{k} u_j \equiv 0 \mod 2\pi$, but not on a proper submanifold $\sum_{j \in I} u_j \equiv 0 \mod 2\pi$ with I a proper nonvacuous subset of $(1,\ldots,k)$, in which case it is one.

THEOREM 1. *Let $X(t)$ be an r-vector valued strictly stationary process satisfying Assumption I. Let $f_{k,a}^{(T)}(\lambda)$ be an estimate of $f_{k,a}(\lambda)$ of the type given in (2.4) with weight function W satisfying Assumption II. If $B_T T \to \infty$ as $B_T \to 0$ and $T \to \infty$, then*

$$Ef_{k,a}^{(T)}(\lambda) = f_{k,a}(\lambda) + 0(B_T) + 0(B_T^{-1}T^{-1}). \tag{2.5}$$

Thus we have a class of asymptotically unbiased estimates under the assumption made in Theorem 1. The following results describe the asymptotic behavior of the covariance of two such estimates. It is remarkable that to the first order the asymptotic behavior of covariances of kth-order cumulant spectral density estimates depends only on second-order spectra under the assumptions made. The covariance of two complex-valued random variables X,Y is taken to be $\text{cov}\,(X,Y) = EX\bar{Y} - (EX)(E\bar{Y})$.

THEOREM 2. *Let $X(t)$ be an r-vector valued strictly stationary process satisfying Assumption I. Let $f_{k,a}^{(T)}(\lambda)$ and $f_{k,b}^{(T)}(\mu)$ be estimates of $f_{k,a}(\lambda)$ and $f_{k,b}(\mu)$, respectively, of the type given in formula (2.4) with weight function W_T satisfying Assumption II. Then*

$$\text{cov}\,[f_{k,a}^{(T)}(\lambda), f_{k,b}^{(T)}(\mu)] = 2\pi\, T^{-1} \sum_P \int W_T(\lambda - \alpha)\, W_T(\mu - \beta)\eta(\sum_1^k \alpha_j)$$
$$\times (\prod_1^k \eta(\alpha_j - (P\beta)_j) f'_{a_j,(Pb)_j}(\alpha_j))d\alpha\, d\beta + 0(B_T^{-k+2}T^{-1}), \tag{2.6}$$

where the summation is over all permutations P on the integers $1,\ldots,k$ $(Pb = (b_{P1},\ldots,b_{Pk}))$, and the error term is uniform in λ and μ subject to $\sum_1^k \lambda_j, \sum_1^k \mu_j \equiv 0 \mod 2\pi$.

COROLLARY. *Under the assumptions of Theorem 2,*

$$\lim_{T \to \infty} B_T^{k-1}T\, \text{cov}\,[f_{k,a}^{(T)}(\lambda), f_{k,b}^{(T)}(\mu)] = 2\pi \sum_P (\prod_{j=1}^k \eta\{\lambda_j - (P\mu)_j\} f'_{a_j,(Pb)_j}(\lambda_j))$$
$$\int W(\beta)W(P\beta)\delta(\sum_1^k \beta_j)d\beta, \tag{2.7}$$

where the summation is over all permutations P on the integers $1,\ldots,k$. Here $\eta\{\lambda\} = \sum_{-\infty}^{\infty} \delta\{\lambda + 2j\pi\}$ with $\delta\{\lambda\}$ the Kronecker delta.

Consider estimates of cumulant spectra of orders $k_1 \le k_2 \le \ldots \le k_m$ of the form given in (2.4) with scale factors $B_T^{(1)} \le \ldots \le B_T^{(m)}$. Write the jth such estimate in the form

$$f_{A_j}^{(T)}(\lambda^{(j)}) = \left(\frac{2\pi}{T}\right)^{k_j - 1} \sum W_T^{(j)}\left(\lambda^{(j)} - \frac{2\pi s^{(j)}}{T}\right) \Phi\left(\frac{2\pi s^{(j)}}{T}\right) I_{A_j}^{(T)}\left(\frac{2\pi s^{(j)}}{T}\right), \tag{2.8}$$

where A_j denotes the indices of the k_j series involved in the jth estimate. The scale factors of two estimates of the same order will be taken to be the same.

THEOREM 3. *Let $X(t)$ be an r-vector valued strictly stationary process satisfying As-*

sumption I. Let $f_{A_j}^{(T)}(\lambda^{(j)})$, $j = 1,\ldots,m$, be estimates as given by (2.8) whose weight functions $W_T^{(j)}$ satisfy Assumption II. Then the estimates are asymptotically jointly normally distributed as $T \to \infty$ with estimates of different orders asymptotically independent and estimates of the same order having covariance structure given by (2.7).

The results given above are for discrete parameter processes. Using the analogue of Assumption I for a continuous parameter process and Assumption II, parallel results for a continuous parameter process continuous in the mean are valid, with η in formulas (2.8) and (2.9) replaced by δ.

3. *Aliasing.*—Often a continuous time parameter strictly stationary process $X(t)$ is sampled discretely at time points jh, $j = 0, \pm 1,\ldots$, where $h > 0$. The analogue of Assumption I for a continuous time parameter process implies that the corresponding cumulant spectral densities $g_{k,a}(\lambda)$, $\sum_1^k \lambda_\alpha = 0$, $-\infty < \lambda_\alpha < \infty$, exist, are continuous and continuously differentiable. In addition, we will require

ASSUMPTION III. *Let*

$$|g_{k,a}(\lambda)| \leq M(1 + \|\lambda\|)^{-k+1-\epsilon} \tag{3.1}$$

and

$$|c'_{k,a}(t')| \leq M(1 + \|t'\|)^{-k+1-\epsilon} \tag{3.2}$$

uniformly in λ and t' with $M, \epsilon > 0$.

Under Assumption III the cumulants for the continuous time parameter process $X(t)$ have the following Fourier representation in terms of the cumulant spectral densities

$$c_{k,a}(t) = \int \exp\,(i \sum_{j=1}^k t_j\lambda_j) g_{k,a}(\lambda)\delta(\sum_1^k \lambda_j)d\lambda. \tag{3.3}$$

The discretely sampled process $X(jh)$ has cumulants with the corresponding Fourier representation

$$C_{a_1,\ldots,a_k}(j_1h,\ldots,j_kh)$$
$$= \int \exp\,(i \sum_{\alpha=1}^k j_\alpha h\lambda_\alpha)\delta(\sum_1^k \lambda_j) g_{k,a}(\lambda)d\lambda$$
$$= \int_{-\pi/h}^{\pi/h}\cdots\int \exp\,(i \sum_{\alpha=1}^k j_\alpha\lambda_\alpha h)\eta(\sum_{\alpha=1}^k h\lambda_\alpha) f_{k,a}(\lambda)d\lambda \tag{3.4}$$

in terms of the corresponding cumulant spectral density of $X(jh)$

$$f_{k,a}(\lambda) = \sum{}' g_{k,a}\left(\lambda + \frac{2\pi j}{h}\right), \tag{3.5}$$

where the sum in (3.5) is over $j = (j_1,\ldots,j_k)$ such that $\sum_1^k j_\alpha = -(h/2\pi)\sum^k \lambda_\alpha$.

4. *Previous Work.*—Moments of order k of the $dZ_a(\omega)$ have been considered by Blanc-Lapierre and Fortet.[1] The third-order spectral density, or bispectrum, is defined in Tukey,[2] and asymptotic properties of a class of estimates are given in Rosenblatt and Van Ness.[3] Asymptotic properties of a class of estimates of the kth-order cumulant spectra have been considered in Brillinger.[4]

Summary.—Under appropriate assumptions the asymptotic variance and bias of a class of estimates of the kth-order cumulant spectra of a stationary random process are obtained. The estimates are shown to be distributed asymptotically as complex-valued Gaussian variables. Remarks are made on aliasing.

* Research partially supported by Office of Naval Research.

† John Simon Guggenheim Fellow, 1965–1966.

[1] Blanc-Lapierre, A., and R. Fortet, *Théorie des Fonctions Aléatoires* (Paris: Masson, 1953).

[2] Tukey, J. W., "An introduction to the measurement of spectra," in *Probability and Statistics*, ed. U. Grenander (New York: Wiley, 1959), the Harald Cramér volume.

[3] Rosenblatt, M., and J. W. Van Ness, "Estimation of the bispectrum," *Ann. Math Statist.*, **36**, 1120–1136 (1965).

[4] Brillinger, D. R., "An introduction to polyspectra," *Ann. Math. Statist.*, **36**, 1351–1374 (1965).

Biometrika (1969), **56**, 2, *p*. 375
Printed in Great Britain

Asymptotic properties of spectral estimates of second order

BY DAVID R. BRILLINGER
London School of Economics and Political Science

SUMMARY

Let $\mathbf{X}(t)$ $(t=0, \pm 1, \ldots)$ be a zero mean, r vector-valued, strictly stationary time series satisfying a particular assumption about the near-independence of widely separated values. Given the values $\mathbf{X}(t)$ $(t=0, 1, \ldots, T-1)$, we construct the statistics: $\mathbf{I}_{XX}^{(T)}(\lambda)$ $(-\infty < \lambda < \infty)$, the matrix of second-order periodograms, $\mathbf{F}_{XX}^{(T)}(\lambda)$, the matrix of sample spectral measures, $\mathbf{f}_{XX}^{(T)}(\lambda)$, the matrix of sample spectral densities and $\mathbf{c}_{XX}^{(T)}(u)$ $(u=0, \pm 1, \ldots)$, the matrix of sample covariances. In the paper expressions are derived for the first- and second-order moments and the asymptotic distributions of $\mathbf{I}_{XX}^{(T)}(\lambda)$, $\mathbf{F}_{XX}^{(T)}(\lambda)$, $\mathbf{f}_{XX}^{(T)}(\lambda)$ and $\mathbf{c}_{XX}^{(T)}(u)$. Our purpose is to determine the form of these moments and to indicate the appearance of the Wishart distribution as an exact limiting distribution for $\mathbf{f}_{XX}^{(T)}(\lambda)$. It has previously been suggested as an approximation.

1. INTRODUCTION

We consider asymptotic properties of second-order statistics based on sample values from a strictly stationary vector-valued time series. The series is assumed to possess moments of all orders and to be such that values of the series, well separated in time, are nearly stochastically independent. This weak span of dependence requirement is formulated as Assumption I. It is a principal and unifying assumption of the theorems presented.

The statistics considered are based on the matrix of second-order periodograms. Our method of proceeding is to derive a general theorem on the asymptotic behaviour of the periodogram, including a necessary uniform error term, and then to deduce the behaviour of the other statistics from this. In fact, the periodograms are based on the discrete Fourier transform of the sample. A lemma of Brillinger & Rosenblatt (1967) indicates the elementary asymptotic sampling properties of this transform. The work of Tukey (1967) indicates the extreme rapidity with which it may be calculated and the consequent quick calculation of the statistics of this paper. In addition the pleasant analytic properties of Fourier transforms are well known. We have therefore been led to take the periodogram as the basis of our work for three distinct and important reasons. Our work differs from much of the previous work in giving the periodogram such an important position. A further important distinction from previous work is that no assumption about the linearity of the underlying process is required for the results presented here.

We prove that distinct values of the periodogram tend to be asymptotically independent and have Wishart distributions. The sample spectral measure, $\mathbf{F}_{XX}^{(T)}(\lambda)$, tends to be Gaussian with a spectrum of order four, a trispectrum, appearing in its distribution. The sample autocovariance function, $\mathbf{c}_{XX}^{(T)}(u)$, is also seen to be asymptotically Gaussian, the distribution again involving a trispectrum. We demonstrate the convergence of these statistics, considered as random functions of λ and u respectively, to limiting Gaussian processes.

Two limiting distributions are seen to appear in the case of the sample spectral density

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_12,

matrix. Under one limiting process it tends to be Gaussian and under a second it tends to have a Wishart distribution.

We commence to set down notation. Let $\mathbf{X}(t)$ $(t=0, \pm 1, \ldots)$ be a strictly stationary r vector-valued time series all of whose moments exist. Set

$$E\{\mathbf{X}(t)\} = \mathbf{c}_X, \tag{1·1}$$

$$E[\{\mathbf{X}(t+u)-\mathbf{c}_X\}\{\mathbf{X}(t)-\mathbf{c}_X\}'] = \mathbf{c}_{XX}(u) \quad (t, u=0, \pm 1, \ldots). \tag{1·2}$$

Suppose

$$\sum_{u=-\infty}^{\infty} |\mathbf{c}_{XX}(u)| < \infty. \tag{1·3}$$

Here $|\mathbf{c}_{XX}(u)|$ denotes the matrix of absolute values. We may then define $\mathbf{f}_{XX}(\lambda)$, the $r \times r$ matrix of second-order spectral densities, by

$$\mathbf{f}_{XX}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} \mathbf{c}_{XX}(u) \exp(-i\lambda u) \quad (-\infty < \lambda < \infty) \tag{1·4}$$

and $\mathbf{F}_{XX}(\lambda)$, the matrix of second-order spectral measures, by

$$\mathbf{F}_{XX}(\lambda) = \int_0^\lambda \mathbf{f}_{XX}(\alpha)\, d\alpha \quad (0 \leqslant \lambda \leqslant \pi). \tag{1·5}$$

We suppose that $\mathbf{X}(t)$ has a weak span of time dependence as indicated by Assumption I.

We construct estimates $\mathbf{c}_{XX}^{(T)}(u)$, $\mathbf{f}_{XX}^{(T)}(\lambda)$ and $\mathbf{F}_{XX}^{(T)}(\lambda)$ of $\mathbf{c}_{XX}(u)$, $\mathbf{f}_{XX}(\lambda)$ and $\mathbf{F}_{XX}(\lambda)$. These estimates are based on $\mathbf{I}_{XX}^{(T)}(\lambda)$, the matrix of second-order periodograms. This last is derived from the finite Fourier transform of an observed stretch of data, $\mathbf{X}(t)$ $(t=0, 1, \ldots, T-1)$.

We determine asymptotic expressions for the cumulants of $\mathbf{c}_{XX}^{(T)}(u)$, $\mathbf{I}_{XX}^{(T)}(\lambda)$, $\mathbf{F}_{XX}^{(T)}(\lambda)$ and $\mathbf{f}_{XX}^{(T)}(\lambda)$ and from these cumulants are able to identify the limiting distributions of the appropriately standardized estimates. We also consider the weak convergence of the sequences of stochastic processes

$$\{\mathbf{c}_{XX}^{(T)}(u)\ (u=0, \pm 1, \ldots)\}, \quad \{\mathbf{F}_{XX}^{(T)}(\lambda)\ (0 \leqslant \lambda \leqslant \pi)\} \quad \text{and} \quad \{\mathbf{f}_{XX}^{(T)}(\lambda)\ (-\infty < \lambda < \infty)\}.$$

We do not assume that $\mathbf{X}(t)$ is a linear process.

In the paper $W_r(\nu, \mathbf{\Sigma})$ will denote an $r \times r$ symmetric matrix-valued Wishart variate with variance–covariance matrix $\mathbf{\Sigma}$ and ν degrees of freedom. Let $W_r^C(\nu, \mathbf{\Sigma})$ denote an $r \times r$ Hermitian matrix-valued complex Wishart variate with variance–covariance matrix $\mathbf{\Sigma}$ and ν degrees of freedom. This last distribution is discussed by Goodman (1963). For real matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{Z} = \mathbf{A} + i\mathbf{B}$, write

$$\mathbf{Z}^R = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & \mathbf{A} \end{bmatrix},$$

then the two are connected by $W_r^C(\nu, \mathbf{\Sigma})^R = W_{2r}(\nu, \mathbf{\Sigma}^R)$. We set

$$\Delta^{(T)}(\lambda) = \sum_{t=0}^{T-1} \exp(-i\lambda t), \tag{1·6}$$

$$\eta(\lambda) = \begin{cases} 1 & (\lambda \equiv 0, \bmod 2\pi), \\ 0 & \text{otherwise.} \end{cases} \tag{1·7}$$

For $(Y_1, Y_2, \ldots, Y_k)$ a random variable with real or complex components, we denote its joint cumulant of order k by

$$\operatorname{cum}(Y_1, Y_2, \ldots, Y_k). \tag{1·8}$$

This is the coefficient of $t_1 t_2 \ldots t_k$ in the expansion of its cumulant generating function. For X, Y complex-valued, $\operatorname{cov}(X, Y) = E[\{(X - E(X)\}\{\overline{Y} - E(\overline{Y})\}']$.

2. Parameters and estimates

Let the r vector-valued series $\mathbf{X}(t)$ have real-valued components $X_a(t)$ $(a=1,2,\ldots,r)$. All moments are assumed to exist and we set

$$c_{a_1,\ldots,a_k}(t_1,\ldots,t_{k-1}) = \mathrm{cum}\{X_{a_1}(t_1+\tau),\ldots,X_{a_{k-1}}(t_{k-1}+\tau),X_{a_k}(\tau)\}$$
$$(a_1,\ldots,a_k = 1,2,\ldots,r;\ t_1,\ldots,t_{k-1},\tau = 0,\ \pm 1,\ldots;\ k=1,2,\ldots) \quad (2\cdot 1)$$

using the assumed stationarity. We then set down

Assumption I. *$\mathbf{X}(t)$ is a strictly stationary series all of whose moments exist. For each $j=1,2,\ldots,k-1$ and any k-tuple $a_1,a_2,\ldots,a_k$ we have*

$$\sum_{t_1,\ldots,t_{k-1}} |t_j c_{a_1,\ldots,a_{k-1}}(t_1,\ldots,t_{k-1})| < \infty \quad (k=2,3,\ldots). \quad (2\cdot 2)$$

Because cumulants are measures of the joint dependence of random variables, (2·2) is seen to be a form of mixing or asymptotic independence requirement for values of $\mathbf{X}(t)$ well separated in time. In the case of a Gaussian series, because cumulants of order greater than 2 vanish, Assumption I is satisfied if one requires only

$$\sum_{t=-\infty}^{\infty} |tc_{aa}(t)| < \infty, \quad (2\cdot 3)$$

where $c_{aa}(t)$ is the autocovariance function of $X_a(t)$ $(a=1,2,\ldots,r)$.

If $\mathbf{X}(t)$ satisfies Assumption I we may define its cumulant spectral densities by

$$f_{a_1\ldots,a_k}(\lambda_1,\ldots,\lambda_{k-1}) = (2\pi)^{-k+1} \sum_{t_1,\ldots,t_{k-1}} c_{a_1,\ldots,a_k}(t_1,\ldots,t_{k-1}) \exp\left(-i\sum_{j=1}^{k-1}\lambda_j t_j\right)$$
$$(-\infty<\lambda<\infty;\ a_1,\ldots,a_k=1,2,\ldots,r;\ k=1,2,\ldots). \quad (2\cdot 4)$$

If $k=2$, the cross-spectra $f_{a_1a_2}(\lambda)$ are collected together in the matrix $\mathbf{f}_{XX}(\lambda)$ of (1·4).

Suppose now that a stretch, $\mathbf{X}(t)$ $(t=0,1,\ldots,T-1)$ of the series $\mathbf{X}(t)$ is available. For $-\infty<\lambda<\infty$, we define

$$\mathbf{d}_X^{(T)}(\lambda) = \sum_{t=0}^{T-1} \exp(-i\lambda t)\,\mathbf{X}(t), \quad (2\cdot 5)$$

the finite Fourier transform of the given stretch of data. Denote the entries of $\mathbf{d}_X^{(T)}(\lambda)$ by $d_a^{(T)}(\lambda)$ $(a=1,2,\ldots,r)$. Following Brillinger & Rosenblatt (1967) one has

Lemma 2·1. *Suppose Assumption I is satisfied, then*

$$\mathrm{cum}\{d_{a_1}^{(T)}(\lambda_1),\ldots,d_{a_1}^{(T)}(\lambda_k)\} = (2\pi)^{k-1} f_{a_1,\ldots,a_k}(\lambda_1,\ldots,\lambda_{k-1})\,\Delta^{(T)}\left(\sum_{j=1}^{k}\lambda_j\right) + O(1). \quad (2\cdot 6)$$

The error term $O(1)$ is uniform in $\lambda_1,\ldots,\lambda_k$ as $T\to\infty$.

Suppose that $E\{\mathbf{X}(t)\}=0$; then this lemma indicates that one might base estimates of $\mathbf{f}_{XX}(\lambda)$ upon

$$\mathbf{I}_{XX}^{(T)}(\lambda) = (2\pi T)^{-1}\,\mathbf{d}^{(T)}(\lambda)\,\overline{\mathbf{d}}^{(T)\prime}(\lambda), \quad (2\cdot 7)$$

the matrix of second-order periodograms; the bar denotes complex conjugate. As an estimate of $\mathbf{F}_{XX}(\lambda)$ we consider

$$\mathbf{F}_{XX}^{(T)}(\lambda) = \int_0^{\lambda} \mathbf{I}_{XX}^{(T)}(\alpha)\,d\alpha \quad (0\leqslant\lambda\leqslant\pi). \quad (2\cdot 8)$$

24-2

As an estimate of $\mathbf{c}_{XX}(u)$, in this case where $E\{\mathbf{X}(t)\} = 0$, we consider

$$\mathbf{m}_{XX}^{(T)}(u) = \sum_{0 \leqslant t, t+u \leqslant T-1} \mathbf{X}(t+u)\,\mathbf{X}'(t)$$

$$= \int_{-\pi}^{\pi} \mathbf{I}_{XX}^{(T)}(\alpha) \exp(iu\alpha)\, d\alpha. \tag{2·9}$$

Before constructing an estimate of $\mathbf{f}_{XX}(\lambda)$, we set down

ASSUMPTION II. *Let $H(\alpha)$ $(-\pi < \alpha \leqslant \pi)$ be a weight function that is bounded, is symmetric about 0, has a bounded first derivative and is such that*

$$\int_{-\pi}^{\pi} H(\alpha)\, d\alpha = 1. \tag{2·10}$$

Given $B_T > 0$, we then set

$$H^{(T)}(\alpha) = B_T^{-1} H(B_T^{-1}\alpha). \tag{2·11}$$

In later sections we will consider the cases: $B_T = K/T$; $B_T \to 0$, $B_T T \to \infty$ as $T \to \infty$; B_T constant with respect to T.

As an estimate of $\mathbf{f}_{XX}(\lambda)$, we take

$$\mathbf{f}_{XX}^{(T)}(\lambda) = \int_{-\pi}^{\pi} H^{(T)}(\alpha)\, \mathbf{I}_{XX}^{(T)}(\lambda - \alpha)\, d\alpha$$

$$= (2\pi)^{-1} \sum_{u=-T+1}^{T-1} \mathbf{m}_{XX}^{(T)}(u) \exp(-iu\lambda) \int_{-\pi}^{\pi} H^{(T)}(\alpha) \exp(-iu\alpha)\, d\alpha. \tag{2·12}$$

We have been led to consider a variety of statistics based on $\mathbf{I}_{XX}^{(T)}(\lambda)$, the matrix of second-order periodograms, and therefore turn to an investigation of its asymptotic properties.

We note that Bartlett (1966, p. 337) has suggested handling the sampling theory of vector-valued series $\mathbf{X}(t)$ by means of arbitrary linear combinations $\boldsymbol{\alpha}'\mathbf{X}(t)$, with $\boldsymbol{\alpha}$ an r vector.

3. THE PERIODOGRAM

Because all of the moments of $\mathbf{X}(t)$ are finite, all of the moments of $\mathbf{I}_{XX}^{(T)}(\lambda)$ will be finite. We turn to a determination of the asymptotic cumulants of $\mathbf{I}_{XX}^{(T)}(\lambda)$. We do this by using the rules developed by Leonov & Shiryaev (1959) for determining the joint cumulants of polynomial functions of random variables.

Denote the entry in the ath row and bth column of $\mathbf{I}_{XX}^{(T)}(\lambda)$ by $I_{ab}^{(T)}(\lambda)$ $(a, b = 1, 2, \ldots, r)$. We then have

THEOREM 3·1. *Let $\mathbf{X}(t)$ satisfy Assumption I and have mean 0. Then*

$$E\{I_{a_1 b_1}^{(T)}(\lambda_1)\} = f_{a_1 b_1}(\lambda_1) + O(T^{-1}), \tag{3·1}$$

with $O(T^{-1})$ uniform in λ_1,

$$\begin{aligned}\operatorname{cov}\{I_{a_1 b_1}^{(T)}(\lambda_1),\ I_{a_2 b_2}^{(T)}(\lambda_2)\} = {} & T^{-2}|\Delta^{(T)}(\lambda_1 - \lambda_2)|^2 f_{a_1 a_2}(\lambda_1) f_{b_1 b_2}(-\lambda_1) \\ & + T^{-2}|\Delta^{(T)}(\lambda_1 + \lambda_2)|^2 f_{a_1 b_2}(\lambda_1) f_{b_1 a_2}(-\lambda_1) \\ & + 2\pi T^{-1} f_{a_1 b_1 a_2 b_2}(\lambda_1, -\lambda_1, -\lambda_2) + T^{-2} R^{(T)}(\lambda_1, \lambda_2),\end{aligned} \tag{3·2}$$

where there is a finite K such that

$$|R^{(T)}(\lambda_1, \lambda_2)| \leqslant K\{|\Delta^{(T)}(\lambda_1 + \lambda_2)| + |\Delta^{(T)}(\lambda_1 - \lambda_2)|\} \tag{3·3}$$

and $\operatorname{cum}\{I^{(T)}_{a_1b_1}(\lambda_1), \ldots, I^{(T)}_{a_kb_k}(\lambda_k)\}$

$$= T^{-k}\Sigma\Delta^{(T)}(\mu_1+\nu_1)\ldots\Delta^{(T)}(\mu_k+\nu_k)f_{c_1d_1}(\mu_1)\ldots f_{c_kd_k}(\mu_k)+O(T^{-1}). \quad (3\cdot4)$$

Here the summation in (3·5) *extends over all partitions*

$$\{(c_1,\mu_1),(d_1,\nu_1)\},\ldots,\{(c_k,\mu_k),(d_k,\nu_k)\}, \quad (3\cdot5)$$

into pairs, of the quantities

$$(a_1,\lambda_1),(b_1,-\lambda_1),\ldots,(a_k,\lambda_k),(b_k,-\lambda_k) \quad (3\cdot6)$$

excluding the cases with $\mu_j = -\nu_j = \lambda_m$ *for some* j, m. *The error term,* $O(T^{-1})$, *in* (3·4) *is uniform in* $\lambda_1, \ldots, \lambda_k$.

The proofs are given in § 8.

From (3·4) we can derive the following corollary.

Corollary. *Under the conditions of the theorem, if one of* $\lambda_j+\lambda_m \equiv 0 \pmod{2\pi}$ *or* $\lambda_j-\lambda_m \equiv 0 \pmod{2\pi}$ *is not true for each* $j, m = 1, \ldots, r$, *then*

$$\operatorname{cum}\{I^{(T)}_{a_1b_1}(\lambda_1),\ldots,I^{(T)}_{a_kb_k}(\lambda_k)\} \to 0 \quad \text{as} \quad T\to\infty \quad (k=3,4,\ldots). \quad (3\cdot7)$$

Let us now determine the limiting distribution of $\mathbf{I}^{(T)}_{XX}(\lambda)$ on the basis of the limiting values of its cumulants. We have,

Theorem 3·2. *Let* $\mathbf{X}(t)$ *satisfy Assumption I and have mean* 0.

If $0 \leqslant \lambda_1 < \lambda_2 < \ldots < \lambda_k \leqslant \pi$, *then* $\mathbf{I}^{(T)}_{XX}(\lambda_1), \ldots, \mathbf{I}^{(T)}_{XX}(\lambda_k)$ *are asymptotically independent. If* $\lambda \not\equiv 0 \pmod{\pi}$, *then* $\mathbf{I}^{(T)}_{XX}(\lambda)$ *tends, in distribution, to* $W^C_r\{1,\mathbf{f}_{XX}(\lambda)\}$. *If* $\lambda \equiv 0 \pmod{\pi}$, *then it tends in distribution to* $W_r\{1,\mathbf{f}_{XX}(\lambda)\}$.

The different asymptotic distributions in the cases $\lambda \not\equiv 0 \pmod{\pi}$ and $\lambda \equiv 0 \pmod{\pi}$ reflect the fact that $\mathbf{f}_{XX}(\lambda)$ and $\mathbf{I}^{(T)}_{XX}(\lambda)$ are real-valued in the latter case.

The asymptotic behaviour of the periodogram, $I^{(T)}_{aa}(\lambda)$, of $X_a(t)$ has been considered by Bartlett (1966, p. 304), Grenander & Rosenblatt (1957), Hannan (1960, p. 52) and Kawata (1959). Walker (1965) determines the asymptotic distribution of $I^{(T)}_{aa}(\lambda)$ for $X_a(t)$ a linear process. Rao (1967) considers asymptotic properties of the cross-periodogram $I^{(T)}_{ab}(\lambda)$, $a \neq b$ of $X_a(t)$ and $X_b(t)$; see also Slutsky (1934) and Olshen (1967).

4. The spectral measure

If the series $\mathbf{X}(t)$ $(t=0, \pm1, \ldots)$ has spectral density matrix $\mathbf{f}_{XX}(\lambda)$, then $\mathbf{F}_{XX}(\lambda)$, the matrix of spectral measures, is given by

$$\mathbf{F}_{XX}(\lambda) = \int_0^\lambda \mathbf{f}_{XX}(\alpha)\,d\alpha \quad (0\leqslant\lambda\leqslant\pi). \quad (4\cdot1)$$

Outside this range $\mathbf{F}_{XX}(\lambda)$ is taken to have period 2π and satisfy $\mathbf{F}_{XX}(-\lambda) = \overline{\mathbf{F}}'_{XX}(\lambda)$. In view of (3·1) one can consider estimating $\mathbf{F}_{XX}(\lambda)$ by

$$\mathbf{F}^{(T)}_{XX}(\lambda) = \int_0^\lambda \mathbf{I}^{(T)}_{XX}(\alpha)\,d\alpha \quad (0\leqslant\lambda\leqslant\pi). \quad (4\cdot2)$$

Because of its elementary dependence on $\mathbf{I}^{(T)}_{XX}(\lambda)$, we can determine the asymptotic moments of $\mathbf{F}^{(T)}_{XX}(\lambda)$ directly from Theorem 3·1. We have

THEOREM 4·1. *Let* $\mathbf{X}(t)$ *satisfy Assumption I and have mean* 0. *Then*

$$E\{F^{(T)}_{a_1b_1}(\lambda_1)\} = F_{a_1b_1}(\lambda_1) + O(T^{-1}), \tag{4·3}$$

$$\operatorname{cov}\{F^{(T)}_{a_1b_1}(\lambda_1), F^{(T)}_{a_2b_2}(\lambda_2)\} = 2\pi T^{-1}\left\{\int_0^{\min(\lambda_1,\lambda_2)} f_{a_1a_2}(\alpha) f_{b_1b_2}(-\alpha)\,d\alpha + \int_0^{\lambda_1}\int_0^{\lambda_2} f_{a_1b_1a_2b_2}(\alpha,-\alpha,-\beta)\,d\alpha\,d\beta\right\} + O(T^{-1}\log T) \tag{4·4}$$

and

$$\operatorname{cum}\{F^{(T)}_{a_1b_1}(\lambda_1), \ldots, F^{(T)}_{a_kb_k}(\lambda_k)\} = O(T^{-k+1}) \quad (k=1,2,\ldots). \tag{4·5}$$

We see that $\mathbf{F}^{(T)}_{XX}(\lambda)$ is an asymptotically unbiased and consistent estimate of $\mathbf{F}_{XX}(\lambda)$. In fact one has

COROLLARY. *Under the conditions of the theorem*

$$\operatorname{prob}\{\lim_{T\to\infty} \mathbf{F}^{(T)}_{XX}(\lambda) = \mathbf{F}_{XX}(\lambda)\} = 1 \quad (0 \leqslant \lambda \leqslant \pi) \tag{4·6}$$

and, in fact,

$$\operatorname{prob}\{\lim_{T\to\infty} \sup_{0\leqslant\lambda\leqslant\pi} |\mathbf{F}^{(T)}_{XX}(\lambda) - \mathbf{F}_{XX}(\lambda)| = 0\} = 1. \tag{4·7}$$

We see that $\mathbf{F}^{(T)}_{XX}(\lambda)$ is a strongly consistent estimate of $\mathbf{F}_{XX}(\lambda)$ with the convergence uniform in λ.

Let us now turn to the consideration of the asymptotic distribution of $\mathbf{F}^{(T)}_{XX}(\lambda)$. We may use Theorem 4·1 to evaluate the limits of the cumulants of $T^{\frac{1}{2}}\{\mathbf{F}^{(T)}_{XX}(\lambda) - \mathbf{F}_{XX}(\lambda)\}$. The limits of cumulants of order greater than two are seen to vanish and we may conclude

THEOREM 4·2. *Let* $\mathbf{X}(t)$ $(t = 0, \pm 1, \ldots)$ *satisfy Assumption I and have mean* 0. *Then*

$$T^{\frac{1}{2}}\{\mathbf{F}^{(T)}_{XX}(\lambda_1) - \mathbf{F}_{XX}(\lambda_1)\}, \ldots, T^{\frac{1}{2}}\{\mathbf{F}^{(T)}_{XX}(\lambda_k) - \mathbf{F}_{XX}(\lambda_k)\}$$

are asymptotically jointly multivariate normal with covariance structure given by

$$\lim_{T\to\infty} \operatorname{cov}[T^{\frac{1}{2}}\{F^{(T)}_{a_1b_1}(\mu_1) - F_{a_1b_1}(\mu_1)\}, T^{\frac{1}{2}}\{F^{(T)}_{a_2b_2}(\mu_2) - F_{a_2b_2}(\mu_2)\}]$$
$$= 2\pi\int_0^{\min(\mu_1,\mu_2)} f_{a_1a_2}(\alpha) f_{b_1b_2}(-\alpha)\,d\alpha + 2\pi\int_0^{\mu_1}\int_0^{\mu_2} f_{a_1b_1a_2b_2}(\alpha,-\alpha,\beta)\,d\alpha\,d\beta$$
$$(\mu_1, \mu_2 = \lambda_1, \ldots, \lambda_k;\ a_j, b_j = 1, 2, \ldots, r;\ j = 1, 2, \ldots, k;\ k = 1, 2, \ldots). \tag{4·8}$$

This theorem indicates the asymptotic normality of finite collections of the $\mathbf{F}^{(T)}_{XX}(\lambda)$. We turn to stronger results concerning the convergence of the stochastic process

$$[T^{\frac{1}{2}}\{\mathbf{F}^{(T)}_{XX}(\lambda) - \mathbf{F}_{XX}(\lambda)\} \quad (0 \leqslant \lambda \leqslant \pi)]$$

to a certain Gaussian process. We first set down some terminology.

For $0 < \alpha \leqslant 1$, $\operatorname{Lip}^{r\times r}_\alpha(0,\pi)$ will denote the Banach space of $r \times r$ matrix-valued functions $\mathbf{Y}(\lambda)$ $(0 < \lambda \leqslant \pi)$, $\mathbf{Y}(0) = 0$, with norm

$$\|\mathbf{Y}(\lambda)\| = \sup_{0\leqslant\lambda<\pi} |\mathbf{Y}(\lambda)| + \sup_{0\leqslant\lambda,\,\lambda+\epsilon\leqslant\pi} |\epsilon|^{-\alpha}|\mathbf{Y}(\lambda+\epsilon) - \mathbf{Y}(\lambda)|. \tag{4·9}$$

In the case $r = 1$, this space is discussed by Lamperti (1962).

A sequence $\{\mathbf{Y}^{(T)}(\lambda)\,(0 \leqslant \lambda \leqslant \pi)\}\,(T=1,2,\ldots)$ of stochastic processes, with values in $\mathrm{Lip}_\alpha^{r\times r}(0,\pi)$, is said to *converge weakly in the topology of* $\mathrm{Lip}_\alpha^{r\times r}$ to a process $\mathbf{Y}(\lambda)$, with values in $\mathrm{Lip}_\alpha^{r\times r}(0,\pi)$, if

$$\lim_{T\to\infty} E[\mathscr{F}\{\mathbf{Y}^{(T)}\}] = E\{\mathscr{F}(\mathbf{Y})\} \tag{4·10}$$

for any bounded real-valued function $\mathscr{F}(\,.\,)$, continuous on $\mathrm{Lip}_\alpha^{r\times r}(0,\pi)$.

We may now state

THEOREM 4·3. *Let* $\mathbf{X}(t)$ $(t=0,\pm1,\ldots)$ *satisfy Assumption I and have mean* 0. *Then, for any* α *with* $0<\alpha<\frac{1}{2}$, *the sequence of processes*

$$[T^{\frac{1}{2}}\{\mathbf{F}_{XX}^{(T)}(\lambda)-\mathbf{F}_{XX}(\lambda)\}\quad (0\leqslant\lambda\leqslant\pi)]$$

converges weakly in the topology of $\mathrm{Lip}_\alpha^{r\times r}$ *to an* $r\times r$ *matrix-valued Gaussian process*

$$\{\mathbf{Y}(\lambda)\quad (0\leqslant\lambda\leqslant\pi)\}$$

with mean 0 *and*

$$\operatorname{cov}\{Y_{a_1b_1}(\lambda_1), Y_{a_2b_2}(\lambda_2)\} = 2\pi\int_0^{\min(\lambda_1,\lambda_2)} f_{a_1a_2}(\alpha)f_{b_1b_2}(-\alpha)\,d\alpha$$
$$+2\pi\int_0^{\lambda_1}\int_0^{\lambda_2} f_{a_1b_1a_2b_2}(\alpha,-\alpha,-\beta)\,d\alpha\,d\beta \quad (a_j,b_j=1,2,\ldots,r;\ j=1,2). \tag{4·11}$$

If $\mathbf{X}(t)$ satisfies the condition of the theorem, we are now able to assert the convergence in distribution of functionals such as

$$T^{\frac{1}{2}}\sup_{0\leqslant\lambda\leqslant\pi}|\mathbf{F}_{XX}^{(T)}(\lambda)-\mathbf{F}_{XX}(\lambda)|, \tag{4·12}$$

$$T\int_0^\pi\int_0^\pi |\alpha-\beta|^{-\frac{3}{2}}\,|\mathbf{F}_{XX}^{(T)}(\alpha)-\mathbf{F}_{XX}^{(T)}(\beta)-\mathbf{F}_{XX}(\alpha)+\mathbf{F}_{XX}(\beta)|^2\,d\alpha\,d\beta \tag{4·13}$$

to corresponding functionals based on the Gaussian process $\mathbf{Y}(\lambda)$ of the theorem.

Because

$$\mathbf{I}_{XX}^{(T)}(\lambda) = \frac{d}{d\lambda}\{\mathbf{F}_{XX}^{(T)}(\lambda)\}\quad (0\leqslant\lambda\leqslant\pi), \tag{4·14}$$

we may expect $\mathbf{I}_{XX}^{(T)}(\lambda)$ to exhibit some of the properties of the (generalized) derivative of the process $\mathbf{Y}(\lambda)$.

If $r=1$ and $X(t)$ is a linear process, then Grenander & Rosenblatt (1957) demonstrated the weak convergence of $T^{\frac{1}{2}}|F^{(T)}(\lambda)-F(\lambda)|$ to a Gaussian process in the coarser topology of uniform convergence. Ibragimov (1963) and Malevich (1964, 1965) have considered the weak convergence of $T^{\frac{1}{2}}|F^{(T)}(\lambda)-F(\lambda)|$ in the case that $X(t)$ is Gaussian with square integrable spectral density.

5. The autocovariance function

Let $\mathbf{X}(t)$ $(t=0,\pm1,\ldots)$ denote an r vector-valued stationary series with autocovariance function

$$\mathbf{c}_{XX}(u) = E([\mathbf{X}(t+u)-E\{\mathbf{X}(t+u)\}]\,[\mathbf{X}(t)-E\{\mathbf{X}(t)\}]')\quad (u=0,\pm1,\ldots). \tag{5·1}$$

If $E\{\mathbf{X}(t)\}=0$, and the values $\mathbf{X}(t)$ $(t=0,1,\ldots,T-1)$ are available, then we can consider estimating $\mathbf{c}_{XX}(u)$ by

$$\mathbf{m}_{XX}^{(T)}(u) = T^{-1}\sum_{0\leqslant t,\,t+u\leqslant T-1}\mathbf{X}(t+u)\,\mathbf{X}'(t)\quad (u=0,\pm1,\ldots). \tag{5·2}$$

We have seen that
$$\mathbf{m}_{XX}^{(T)}(u) = \int_{-\pi}^{\pi} \mathbf{I}_{XX}^{(T)}(\alpha)\exp(iu\alpha)\,d\alpha \quad (5\cdot3)$$

and so we may determine the statistical properties of $\mathbf{m}_{XX}^{(T)}(u)$ from those of $\mathbf{I}_{XX}^{(T)}(\alpha)$. Theorem 3·1 gives directly

THEOREM 5·1. *Let* $\mathbf{X}(t)$ $(t=0, \pm1, \ldots)$ *satisfy Assumption I and have mean* 0. *Then*
$$E\{m_{a_1b_1}^{(T)}(u_1)\} = m_{a_1b_1}(u_1) + O(T^{-1}), \quad (5\cdot4)$$

$$\begin{aligned}\operatorname{cov}&\{m_{a_1b_1}^{(T)}(u_1), m_{a_2b_2}^{(T)}(u_2)\}\\ &= T^{-1}\Bigg[\int_0^{2\pi}\exp\{i\alpha(u_1-u_2)\}f_{a_1a_2}(\alpha)f_{b_1b_2}(-\alpha)\,d\alpha\\ &\quad+\int_0^{2\pi}\exp\{i\alpha(u_1+u_2)\}f_{a_1b_2}(\alpha)f_{b_1a_2}(-\alpha)\,d\alpha\\ &\quad+2\pi\int_0^{2\pi}\int_0^{2\pi}\exp\{i(\alpha_1u_1+\alpha_2u_2)\}f_{a_1b_1a_2b_2}(\alpha_1,-\alpha_1,\alpha_2)\,d\alpha_1\,d\alpha_2\Bigg]+O(T^{-2}\log T) \quad (5\cdot5)\end{aligned}$$

and
$$\operatorname{cum}\{m_{a_1b_1}^{(T)}(u_1), \ldots, m_{a_kb_k}^{(T)}(u_k)\} = O(T^{-k+1}) \quad (5\cdot6)$$

for $u_j = 0, \pm1, \ldots;\ a_j, b_j = 1, 2, \ldots, r;\ j = 1, \ldots, k$ *and* $k = 1, 2, \ldots$. *The error terms are uniform in each case.*

COROLLARY. *Under the conditions of the theorem*
$$\operatorname{prob}\{\lim_{T\to\infty}\mathbf{m}_{XX}^{(T)}(u) = \mathbf{m}_{XX}(u)\} = 1 \quad (u=0, \pm1, \ldots). \quad (5\cdot7)$$

Also
$$\operatorname{prob}[\lim_{T\to\infty}\sup_{u\neq0}|u^{-1}\{\mathbf{m}_{XX}^{(T)}(u) - \mathbf{m}_{XX}(u)\}| = 0] = 1. \quad (5\cdot8)$$

Let us now turn to an investigation of the asymptotic distribution of finite collections of the $\mathbf{m}_{XX}^{(T)}(u)$.

THEOREM 5·2. *Let* $X(t)$ $(t=0, \pm1, \ldots)$ *satisfy Assumption I and have mean* 0. *Then*
$$T^{\frac{1}{2}}\{\mathbf{m}_{XX}^{(T)}(u_1) - \mathbf{m}_{XX}(u_1)\}, \ldots, T^{\frac{1}{2}}\{\mathbf{m}_{XX}^{(T)}(u_k) - \mathbf{m}_{XX}(u_k)\}$$
are asymptotically jointly multivariate normal with covariance structure given by
$$\begin{aligned}\lim_{T\to\infty}&\operatorname{cov}[T^{\frac{1}{2}}\{m_{a_1b_1}^{(T)}(u_1) - m_{a_1b_1}(u_1)\}, T^{\frac{1}{2}}\{m_{a_2b_2}^{(T)}(u_2) - m_{a_2b_2}(u_2)\}]\\ &= \int_0^{2\pi}\exp\{i\alpha(u_1-u_2)\}f_{a_1a_2}(\alpha)f_{b_1b_2}(-\alpha)\,d\alpha + \int_0^{2\pi}\exp\{i\alpha(u_1+u_2)\}f_{a_1b_2}(\alpha)f_{b_1a_2}(-\alpha)\,d\alpha\\ &\quad+2\pi\int_0^{2\pi}\int_0^{2\pi}\exp\{i(\alpha_1u_1+\alpha_2u_2)\}f_{a_1b_1a_2b_2}(\alpha_1,-\alpha_1,\alpha_2)\,d\alpha_1\,d\alpha_2. \quad (5\cdot9)\end{aligned}$$

Let us turn to the derivation of a theorem concerning the asymptotic behaviour of the function $\{\mathbf{m}_{XX}^{(T)}(u)\ (u=0, \pm1, \ldots)\}$. We first introduce $\mathscr{A}_\alpha^{r\times r}$. This is the image of $\mathrm{Lip}_\alpha^{r\times r}(0,\pi)$ by the Fourier–Stieltjes transform, that is the space of $r\times r$ matrix-valued sequences $\{\mathbf{y}(u)\ (u=0, \pm1, \ldots)\}$ of the form
$$\mathbf{y}(u) = \int_0^{\pi}\exp(iu\alpha)\,d\mathbf{Y}(\alpha), \quad (5\cdot10)$$
where $\mathbf{Y}(\alpha)\in\mathrm{Lip}_\alpha^{r\times r}(0,\pi)$ for $0 < \alpha \leqslant 1$. As norm of this $\mathbf{y}(u)$ we take $\|\mathbf{Y}(\lambda)\|$.

Elementary calculations indicate that $\mathscr{A}_\alpha^{r\times r}$ is a subspace of the Banach space of $r \times r$ matrix-valued sequences, $\mathbf{y}(u)$ ($u = 0, \pm 1, \ldots$), with norm

$$\|\mathbf{y}(u)\|_\alpha = \sup_u |\mathbf{y}(u)| + \sup_{u \neq 0} |u^{-1+\alpha}\,\mathbf{y}(u)|. \tag{5·11}$$

We have

THEOREM 5·3. *Let* $\mathbf{X}(t)$ *satisfy Assumption I and have mean* 0. *Then, for any* α *with* $0 < \alpha < \frac{1}{2}$, *the sequence of processes* $\{T^{\frac{1}{2}}\{\mathbf{m}_{XX}^{(T)}(u) - \mathbf{m}_{XX}(u)\}\ (u = 0, \pm 1, \ldots)\}$ *converges weakly in the topology of* $\mathscr{A}_\alpha^{r\times r}$ *to a zero mean Gaussian process* $\{\mathbf{y}(u)\ (u = 0, \pm 1, \ldots)\}$ *with*

$$\begin{aligned}
&\operatorname{cov}\{y_{a_1b_1}(u_1), y_{a_2b_2}(u_2)\} \\
&= \int_0^{2\pi} \exp\{i\alpha(u_1-u_2)\} f_{a_1a_2}(\alpha) f_{b_1b_2}(-\alpha)\,d\alpha + \int_0^{2\pi} \exp\{i\alpha(u_1+u_2)\} f_{a_1b_2}(\alpha) f_{b_1a_2}(-\alpha)\,d\alpha \\
&\quad + 2\pi \int_0^{2\pi}\int_0^{2\pi} \exp\{i(\alpha_1u_1+\alpha_2u_2)\} f_{a_1b_1a_2b_2}(\alpha_1, -\alpha_1, \alpha_2)\,d\alpha_1\,d\alpha_2.
\end{aligned} \tag{5·12}$$

Turning to related work and the case of a real-valued series $X_a(t)$, we note that Slutsky (1934) considered asymptotic properties of $m_{aa}^{(T)}(u)$ in the Gaussian case. Parzen (1961) gave conditions for the convergence of $m_{aa}^{(T)}(u)$ with probability 1. Bartlett (1946; 1966, p. 285) and Parzen (1957*b*) developed formulae for the asymptotic variance of $m_{aa}^{(T)}(u)$. Walker (1954), Lomnicki & Zaremba (1957, 1959), Parzen (1957*a*) and Anderson & Walker (1964) considered the asymptotic normality of $m_{aa}^{(T)}(u)$ in the case where $X_a(t)$ was a linear process. Rosenblatt (1962) considered asymptotic normality in the case where $X_a(t)$ is Gaussian.

Bartlett (1966, p. 286) noted that, if instead of the autocovariance function $m_{aa}^{(T)}(u)$, one considered the autocorrelation function $m_{aa}^{(T)}(u)/m_{aa}^{(T)}(0)$, and $X_a(t)$ was a linear process, then only second-order spectra appear in the asymptotic variance formula. Elementary calculations based on (5·9) indicate that this result does not continue to hold in the case that $X_a(t)$ is not a linear process.

6. THE SPECTRAL DENSITY

We turn to the investigation of estimates of $\mathbf{f}_{XX}(\lambda)$ the matrix of second-order spectral densities. Let $H(\alpha)$ $(-\pi < \alpha \leqslant \pi)$ be a weight function satisfying Assumption II. Let B_T be a scale factor depending on T. Suppose $\mathbf{X}(t)$ $(t = 0, \pm 1, \ldots)$ has mean $\mathbf{0}$. Let

$$h^{(T)}(u) = \int_{-\pi}^{\pi} H^{(T)}(\alpha) \exp(-iu\alpha)\,d\alpha, \tag{6·1}$$

where $H^{(T)}(\alpha)$ is given by (2·11). As an estimate of $\mathbf{f}_{XX}^{(T)}(\alpha)$, we propose

$$\begin{aligned}
\mathbf{f}_{XX}^{(T)}(\lambda) &= (2\pi)^{-1} \sum_{u=-T+1}^{T-1} h^{(T)}(u)\,\mathbf{m}_{XX}^{(T)}(u) \exp(-iu\lambda) \\
&= \int_{-\pi}^{\pi} H^{(T)}(\alpha)\,\mathbf{I}_{XX}^{(T)}(\lambda - \alpha)\,d\alpha,
\end{aligned} \tag{6·2}$$

that is a weighted average of the periodogram. We may prove

THEOREM 6·1. *Let* $\mathbf{X}(t)$ $(t = 0, \pm 1, \ldots)$ *satisfy Assumption I and have mean* 0. *Let* $\mathbf{f}_{XX}^{(T)}(\lambda)$ *be constructed in the manner of* (6·2), *where* $H(\alpha)$ *satisfies Assumption II. Then*

$$E\{f_{a_1b_1}^{(T)}(\lambda_1)\} = \int_{-\pi}^{\pi} H(\alpha) f_{a_1b_1}(\lambda - B_T\alpha)\,d\alpha + O(T^{-1}), \tag{6·3}$$

$$\operatorname{cov}\{f^{(T)}_{a_1b_1}(\lambda_1), f^{(T)}_{a_2b_2}(\lambda_2)\}$$
$$= T^{-1}\left\{\int_{-\pi}^{\pi} H(\alpha)\, H(\lambda_2-\lambda_1-\alpha) f_{a_1a_2}(\lambda_1-\alpha) f_{b_1b_2}(\alpha-\lambda_1)\, d\alpha\right.$$
$$\left. + \int_{-\pi}^{\pi} H(\alpha)\, H(\lambda_2+\lambda_1-\alpha) f_{a_1b_2}(\lambda_1-\alpha) f_{b_1a_2}(\alpha-\lambda_1)\, d\alpha\right\}$$
$$+ 2\pi T^{-1} \iint_{-\pi}^{\pi} H(\alpha_1)\, H(\alpha_2) f_{a_1b_1a_2b_2}(\lambda_1-\alpha_1, \alpha_1-\lambda_1, \alpha_2-\lambda_2)\, d\alpha_1\, d\alpha_2$$
$$+ O(T^{-2}\log T) \quad (B_T = 1), \tag{6·4}$$

$$\operatorname{cov}\{f^{(T)}_{a_1b_1}(\lambda_1), f^{(T)}_{a_2b_2}(\lambda_2)\} = B_T^{-1}T^{-1}\left\{\int_{-\pi}^{\pi} H(\alpha)^2\, d\alpha\right\}[\eta(\lambda_1-\lambda_2) f_{a_1a_2}(\lambda_1) f_{b_1b_2}(-\lambda_1)$$
$$+ \eta(\lambda_1+\lambda_2) f_{a_1b_2}(\lambda_1) f_{b_1a_2}(-\lambda_1)]$$
$$+ O(B_T^{-2}T^{-2}) \quad (B_T \to 0,\ B_T T \to \infty \text{ as } T \to \infty).$$

Also $\operatorname{cum}\{f^{(T)}_{a_1b_1}(\lambda_1), \ldots, f^{(T)}_{a_kb_k}(\lambda_k)\} = \begin{cases} O(T^{-k+1}) & (B_T = 1), \\ O(B_T^{-k+1}T^{-k+1}) & (B_T \to 0,\ B_T T \to \infty \text{ as } T \to \infty). \end{cases}$

Turning to the asymptotic distribution of $\mathbf{f}^{(T)}_{XX}(\lambda)$, we have

THEOREM 6·2. *Let* $\mathbf{X}(t)$ $(t = 0, \pm 1, \ldots)$ *satisfy Assumption I and have mean* 0. *Let*

$$\mathbf{f}^{(T)}_{XX}(\lambda_1), \ldots, \mathbf{f}^{(T)}_{XX}(\lambda_k)$$

be constructed in the manner of (6·2), *where* $H(\alpha)$ *satisfies Assumption II. If* $B_T T \to \infty$ *as* $T \to \infty$, *then*

$$(B_T T)^{\frac{1}{2}}[\mathbf{f}^{(T)}_{XX}(\lambda_1) - E\{\mathbf{f}^{(T)}_{XX}(\lambda_1)\}], \ldots, (B_T T)^{\frac{1}{2}}[\mathbf{f}^{(T)}_{XX}(\lambda_k) - E\{\mathbf{f}^{(T)}_{XX}(\lambda_k)\}] \quad (k = 1, 2, \ldots)$$

is asymptotically normal with mean 0 *and covariance structure indicated by* (6·4).

On occasion an alternative form of asymptotic distribution may prove relevant. Suppose we estimate $\mathbf{f}_{XX}(\lambda)$ by a simple average of periodograms. For example, with $s(T)$, m integers and $2\pi s(T)/T$ near λ, consider

$$(2m+1)^{-1} \sum_{s=-m}^{m} \mathbf{I}^{(T)}_{XX}[2\pi\{s(T)+s\}/T] \quad (\lambda \not\equiv 0, \bmod \pi) \tag{6·5}$$

and
$$(2m+2)^{-1}\left\{\mathbf{I}^{(T)}_{XX}(\lambda) + \sum_{s=-m}^{m} \mathbf{I}^{(T)}_{XX}(\lambda + 2\pi s/T)\right\} \quad (\lambda \equiv 0, \bmod \pi). \tag{6·6}$$

Then one has

THEOREM 6·3. *Let* $\mathbf{X}(t)$ $(t = 0, \pm 1, \ldots)$ *satisfy Assumption I and have mean* 0. *Let* m *be fixed and* $2\pi s(T)/T \to \lambda$ *as* $T \to \infty$. *If* $\lambda \not\equiv 0$ $(\bmod \pi)$, (6·5) *tends in distribution to*

$$(2m+1)^{-1}\, W_r^C\{2m+1, \mathbf{f}_{XX}(\lambda)\}.$$

If $\lambda \equiv 0$ $(\bmod \pi)$, (6·6) *tends in distribution to* $(2m+1)^{-1} W_r\{2m+1, f_{XX}(\lambda)\}$.

In the notation of (6·2), the estimates (6·5), (6·6) correspond to a B_T of order T^{-1}.

One may prove a theorem concerning the weak convergence of the process

$$\{\mathbf{f}^{(T)}_{XX}(\lambda) \quad (-\infty < \lambda < \infty)\}$$

in the case $B_T = 1$. The theorem follows directly from the weak convergence of

$$\{\mathbf{F}^{(T)}_{XX}(\lambda) \quad (0 \leqslant \lambda \leqslant \pi)\}$$

and the representation

$$f_{ab}^{(T)}(\lambda) = \int_0^{\pi} H(\alpha)\, d\{F_{ab}^{(T)}(\lambda-\alpha) + F_{ba}^{(T)}(\lambda-\alpha)\}. \tag{6·7}$$

The theorem involves the weak convergence of $T^{\frac{1}{2}}[\mathbf{f}_{XX}^{(T)}(\lambda) - E\{\mathbf{f}_{XX}^{(T)}(\lambda)\}]$ to a zero mean Gaussian process with covariance structure indicated by (6·4) and is clear in view of our previous results.

The asymptotic mean and variance of power spectral estimates were investigated by Grenander & Rosenblatt (1957), Parzen (1957*a*, *b*) and Blackman & Tukey (1958, p. 16). Asymptotic normality has been demonstrated, under various conditions, by Rosenblatt (1959) and Brillinger (1965, 1968). Bartlett (1950) made use of the χ^2 distribution for smoothed periodogram estimates. The approximation of the distribution of $\mathbf{f}_{XX}^{(T)}(\lambda)$ by a complex Wishart was suggested by Goodman (1963). Wahba (1968) proves that expression (6·5) has the form $(2m+1)^{-1} W_r^C\{2m+1, \mathbf{f}_{XX}(\lambda)\} + O_p(T^{-1}) + O_p(m^{-1})$. This does not yield our Theorem 6·3, however.

7. Departures from assumptions

The most common departure from the assumptions of this paper will be for the series to have non-zero mean.

Let $\mathbf{Y}(t)$ satisfy Assumption I. Let $\mathbf{X}(t) = \mathbf{Y}(t) - E\{\mathbf{Y}(t)\}$, then $\mathbf{X}(t)$ will have zero mean and the results of the paper will apply to it. Suppose $\mathbf{Y}(t)$ $(t = 0, 1, \ldots, T-1)$ are available. Set

$$\bar{\mathbf{Y}}^{(T)} = T^{-1} \sum_{t=0}^{T-1} \mathbf{Y}(t) \quad (t = 0, 1, \ldots, T-1) \tag{7·1}$$

and

$$\mathbf{Y}^{(T)}(t) = \mathbf{Y}(t) - \bar{\mathbf{Y}}^{(T)} = \mathbf{X}(t) + [E\{\mathbf{Y}(t)\} - \bar{\mathbf{Y}}^{(T)}] \quad (t = 0, 1, \ldots, T-1). \tag{7·2}$$

Our procedure will be to replace the statistics, of the paper, based on $\mathbf{X}(t)$ $(t = 0, 1, \ldots, T-1)$ by statistics based on $\mathbf{Y}^{(T)}(t)$ $(t = 0, 1, \ldots, T-1)$.

In many cases the difference $\bar{\mathbf{Y}}^{(T)} - E\{\mathbf{Y}(t)\}$ is asymptotically negligible and the results of the paper continue to hold. See the discussion of Walker (1965), Parzen (1957*a*, *b*), Hannan (1967), for example.

There are immediate extensions of the theorems of this paper to apply to the case of a continuous time process $\mathbf{X}(t)$ $(-\infty < t < \infty)$ satisfying

$$\int \cdots \int_{t_1, \ldots, t_{k-1}} |t_j c_{a_1, \ldots, a_k}(t_1, \ldots, t_{k-1})|\, dt_1 \ldots dt_{k-1} < \infty \quad (a_1, \ldots, a_k = 1, 2, \ldots, r;\ k = 1, 2, \ldots). \tag{7·3}$$

8. Proofs

In this section we present proofs of the various theorems of the paper.

Proof of Theorem 3·1. Let us determine

$$\operatorname{cum}\{d_{a_1}^{(T)}(\lambda_1)\, d_{b_1}^{(T)}(-\lambda_1), \ldots, d_{a_k}^{(T)}(\lambda_k)\, d_{b_k}^{(T)}(-\lambda_k)\}$$

as (3·1), (3·2), (3·3) follow directly from it. We use a result of Leonov & Shiryaev (1959) and argue as did Brillinger (1965), Brillinger & Rosenblatt (1967) and Brillinger (1968). The cumulant in question is given by

$$\sum_{\nu} \operatorname{cum}(C_{\nu_1}) \ldots \operatorname{cum}(C_{\nu_p}), \tag{8·1}$$

where $[C_{\nu_1}, \ldots, C_{\nu_p}]$ is an indecomposable partition of the elements of the table

$$\begin{matrix} d_{a_1}^{(T)}(\lambda_1), & d_{b_1}^{(T)}(-\lambda_1) \\ \vdots & \vdots \\ d_{a_k}^{(T)}(\lambda_k), & d_{b_k}^{(T)}(-\lambda_k) \end{matrix} \tag{8·2}$$

and the summation in (8·1) extends over all such indecomposable partitions. We may now use Lemma 2·1 to evaluate cum (C_{ν_j}) and obtain (3·1), (3·2), (3·3) by retaining only the principal terms.

Proof of Corollary. This follows from the fact that $T^{-1}|\Delta^{(T)}(\lambda)| \to 0$ unless $\lambda \equiv 0 \pmod{2\pi}$.

Before turning to a proof of Theorem 3·2, we first note that the characteristic function of a $W_r(\nu, \mathbf{\Sigma})$ variate is given by

$$\det(\mathbf{I} - 2i\mathbf{\Sigma\Theta})^{-\frac{1}{2}\nu} \tag{8·3}$$

(Anderson, 1958), while that of a $W_r^C(\nu, \mathbf{\Sigma})$ variate is given (Goodman, 1963) by

$$\det(\mathbf{I} - i\mathbf{\Sigma\Theta})^{-\nu}. \tag{8·4}$$

These are both analytic in a neighbourhood of the origin so the variates are determined by their moments.

We can now turn to

Proof of Theorem 3·2. The stated asymptotic independence follows from the Corollary of Theorem 3·1.

Suppose $\lambda \not\equiv 0 \pmod{\pi}$, then from (3·4), it follows that

$$\lim_{T\to\infty} \text{cum}\{I_{a_1b_1}^{(T)}(\lambda), \ldots, I_{a_kb_k}^{(T)}(\lambda)\} = \Sigma f_{c_1d_1}(\lambda) \ldots f_{c_kd_k}(\lambda), \tag{8·5}$$

where the summation in (8·5) extends over permutations $(c_1, \ldots, c_k)$ of $(a_1, \ldots, a_k)$, permutations $(d_1, \ldots, d_k)$ of $(b_1, \ldots, b_k)$, no $d_j = b_m$ if $a_j = a_m$. The rules of Leonov & Shiryaev indicate that (8·5) is cum $(W_{a_1b_1}, \ldots, W_{a_kb_k})$, where $\mathbf{W}$ is $W_r^C\{1, \mathbf{f}_{XX}(\lambda)\}$. Because the complex Wishart is determined by its moments, the proof is completed in the case $\lambda \not\equiv 0 \pmod{\pi}$. The case $\lambda \equiv 0 \pmod{\pi}$ follows in a similar manner.

Before proceeding to the proof of Theorem 4·1, we note two properties of the function $\Delta^{(T)}(\alpha)$.

To begin, we have (Edwards, 1967, p. 80)

$$\int_0^\pi |\Delta^{(T)}(\alpha_2 - \alpha_1)| \, d\alpha_2 = O(\log T). \tag{8·6}$$

Also for $0 \leqslant \alpha_1, \lambda_2 \leqslant \pi$

$$\int_0^{\lambda_2} T^{-1}|\Delta^{(T)}(\alpha_1 - \alpha_2)|^2 d\alpha_2 = \begin{Bmatrix} O(T^{-1}) & (\lambda_2 < \alpha_1) \\ 2\pi + O(T^{-1}) & (\lambda_2 > \alpha_1). \end{Bmatrix} \tag{8·7}$$

This last expression follows from the fact (Edwards, 1967, p. 79) that

$$\int_0^{2\pi} T^{-1}|\Delta^{(T)}(\alpha)|^2 d\alpha = 2\pi. \tag{8·8}$$

Proof of Theorem 4·1. We note that

$$F_{a_jb_j}^{(T)}(\lambda_j) = \int_0^{\lambda_j} I_{a_jb_j}^{(T)}(\alpha_j) \, d\alpha_j \quad (j = 1, 2, \ldots, k), \tag{8·9}$$

and so results may be made to follow from corresponding results concerning $I_{a_jb_j}^{(T)}(\alpha_j)$.

Relation (4·3) is seen to follow directly from relations (4·2) and (3·1). Turning to (4·4), from (3·2) one has

$$\begin{aligned}\operatorname{cov}\{F^{(T)}_{a_1b_1}(\lambda_1),\ F^{(T)}_{a_2b_2}(\lambda_2)\} &= \int_0^{\lambda_1} f_{a_1a_2}(\alpha_1)f_{b_1b_2}(-\alpha_1)\left\{\int_0^{\lambda_2} T^{-2}\,|\Delta^{(T)}(\alpha_1-\alpha_2)|^2\,d\alpha_2\right\}d\alpha_1\\ &+\int_0^{\lambda_1} f_{a_1b_2}(\alpha_1)f_{b_1a_2}(-\alpha_1)\left\{\int_0^{\lambda_2} T^{-2}\,|\Delta^{(T)}(\alpha_1+\alpha_2)|^2\,d\alpha_2\right\}d\alpha_1\\ &+2\pi T^{-1}\int_0^{\lambda_1}\int_0^{\lambda_2} f_{a_1b_1a_2b_2}(\alpha_1,\,-\alpha_1,\,-\alpha_2)\,d\alpha_1\,d\alpha_2+O(T^{-2}\log T),\end{aligned} \tag{8·10}$$

and the indicated result follows from the previous discussion.

Expression (4·5) follows from (3·4) directly.

Proof of Corollary. Equation (4·6) states that $\mathbf{F}^{(T)}_{XX}(\lambda)$ tends to $\mathbf{F}_{XX}(\lambda)$ with probability one. Now (4·3), (4·4) and (4·5) indicate that

$$E\{|F^{(T)}_{a_1b_1}(\lambda_1)-F_{a_1b_1}(\lambda_1)|^4\} = O(T^{-2}) \quad (a_1, b_1 = 1, 2, \ldots, r).$$

Equation (4·6) now follows from the convergent series criterion.

Because $\mathbf{F}_{XX}(\lambda)$ is a continuous bounded monotonic function of λ, (4·6) implies (4·7) following a theorem of Polya.

In the proof of Theorem 4·3 we will make use of the identity

$$E(Y_1Y_2\ldots Y_k) = \sum_\nu \operatorname{cum}\{Y_j\ (j\in\nu_1)\}\ldots\operatorname{cum}\{Y_j\ (j\in\nu_p)\}, \tag{8·11}$$

where the summation is over all partitions $(\nu_1, \nu_2, \ldots, \nu_p)$ $(p = 1, 2, \ldots, k)$ of the integers $1, 2, \ldots, k$.

Proof of Theorem 4·3. We note that the various cumulant spectra of $\mathbf{X}(t)$ are bounded following Assumption I. If we note this and use (8·11) above with Theorem 4·3, then for n a positive integer

$$\begin{aligned}|T^nE([I^{(T)}_{ab}(\alpha_1)-E\{I^{(T)}_{ab}(\alpha_1)\}]\ldots[I^{(T)}_{ab}(\alpha_{2n})-E\{I^{(T)}_{ab}(\alpha_{2n})\}])|\\ \leqslant K\Sigma\left\{\frac{|\Delta^{(T)}(\alpha_{\nu_1}\pm\alpha_{\nu_2})|^2}{T}\ldots\frac{|\Delta^{(T)}(\alpha_{\nu_{2n-1}}\pm\alpha_{\nu_{2n}})|^2}{T}\right\}\end{aligned} \tag{8·12}$$

for some $K > 0$, where the summation extends over all permutations $(\nu_1, \nu_2, \ldots, \nu_{2n})$ of $(1, 2, \ldots, 2n)$ and all choices of $\pm$.

Therefore,

$$\begin{aligned}T^nE(|[F^{(T)}_{ab}(\lambda)-E\{F^{(T)}_{ab}(\lambda)\}]-[F^{(T)}_{ab}(\mu)-E\{F^{(T)}_{ab}(\mu)\}]|^{2n})\\ \leqslant K\Sigma\int_\mu^\lambda\ldots\int_\mu^\lambda\frac{|\Delta^{(T)}(\alpha_{\nu_1}\pm\alpha_{\nu_2})|^2}{T}\ldots\frac{|\Delta^{(T)}(\alpha_{\nu_{2n-1}}\pm\alpha_{\nu_{2n}})|^2}{T}\,d\alpha_1\ldots d\alpha_{2n}\\ \leqslant L\,|\lambda-\mu|^n\end{aligned} \tag{8·13}$$

for some $L > 0$ as we may integrate out n of the α's to remove the $\Delta^{(T)}$ functions and then note that the remaining α's range from μ to λ.

It follows from (3·1) that there exists $M > 0$ such that

$$|[E\{F^{(T)}_{ab}(\lambda)\}-F^{(T)}_{ab}(\lambda)]-[E\{F^{(T)}_{ab}(\mu)\}-F^{(T)}_{ab}(\mu)]| \leqslant M\,|\lambda-\mu|. \tag{8·14}$$

The combination of (8·13) and (8·14) gives

$$T^n E\,|\{F^{(T)}_{ab}(\lambda)-F_{ab}(\lambda)\}-\{F^{(T)}_{ab}(\mu)-F_{ab}(\mu)\}|^{2n} \leqslant N\,|\lambda-\mu|^n \tag{8·15}$$

for some $N > 0$.

The indicated theorem now follows from the multivariate extension of the principal theorem of Lamperti (1962).

Proof of Theorem 5·1. We note that

$$m^{(T)}_{a_jb_j}(u_j) = \int_{-\pi}^{\pi} I^{(T)}_{a_jb_j}(\alpha_j)\exp(iu\alpha_j)\,d\alpha_j \quad (j=1,2,\ldots,k), \tag{8·16}$$

and so (5·4), (5·5) and (5·6) follow directly by the arguments used in the proof of Theorem 4·1.

Proof of Corollary. We may write

$$\begin{aligned} m^{(T)}_{a_jb_j}(u_j) &= \int_0^{\pi} \exp(iu\alpha)\,d\{F^{(T)}_{a_jb_j}(\alpha)+F^{(T)}_{b_ja_j}(\alpha)\} \\ &= \{F^{(T)}_{a_jb_j}(\pi)+F^{(T)}_{b_ja_j}(\pi)\} - iu\int_0^{\pi}\{F^{(T)}_{a_jb_j}(\alpha)+F^{(T)}_{b_ja_j}(\alpha)\}\exp(iu\alpha)\,d\alpha \end{aligned} \tag{8·17}$$

if one integrates by parts. Equations (5·7) and (5·8) now follow from (4·6) and (4·7).

Proof of Theorem 5·2. This follows directly as did the proof of Theorem 4·2.

Proof of Theorem 5·3. Because the mapping of $\text{Lip}^{r\times r}_{\alpha}(0,\pi)$ to $\mathscr{A}^{r\times r}_{\alpha}$ indicated by (5·10) is continuous, Theorem 5·3 follows directly from Theorem 5·4 once we note that

$$T^{\frac{1}{2}}\{m^{(T)}_{a_jb_j}(u_j)-m_{a_jb_j}(u_j)\} = \int_0^{\pi}\exp(iu\alpha)\,dT^{\frac{1}{2}}\{F^{(T)}_{a_jb_j}(\alpha)-F_{a_jb_j}(\alpha)+F^{(T)}_{b_ja_j}(\alpha)-F_{b_ja_j}(\alpha)\}. \tag{8·18}$$

Proof of Theorem 6·1. Expression (6·3) follows directly from (6·2), (3·1) and the definition of $H^{(T)}(\alpha)$. Turning to (6·4), from (6·2) we have

$$\operatorname{cov}\{f^{(T)}_{a_1b_1}(\lambda_1), f^{(T)}_{a_2b_2}(\lambda_2)\} = \iint_{-\pi}^{\pi} H^{(T)}(\alpha_1)\,H^{(T)}(\alpha_2)\operatorname{cov}\{I^{(T)}_{a_1b_1}(\lambda_1-\alpha_1), I^{(T)}_{a_2b_2}(\lambda_2-\alpha_2)\}\,d\alpha_1\,d\alpha_2. \tag{8·19}$$

We will substitute into this expression from (3·2). Now in the case that $B_T T\to\infty$ we have

$$\int_{-\pi}^{\pi} H^{(T)}(\alpha)\,T^{-1}\,|\Delta^{(T)}(\gamma-\alpha)|^2\,d\alpha = H^{(T)}(\gamma)+O(B_T^{-2}T^{-1}) \quad (-\pi\leqslant\gamma\leqslant\pi). \tag{8·20}$$

Also

$$\int_{-\pi}^{\pi} |H^{(T)}(\alpha)|\,|\Delta^{(T)}(\gamma-\alpha)|\,d\alpha = O(B_T^{-1}\log T). \tag{8·21}$$

These two indicate that the covariance in question is given by

$$\begin{aligned} T^{-1}\Big\{&\int_{-\pi}^{\pi} H^{(T)}(\alpha)\,H^{(T)}(\lambda_2-\lambda_1-\alpha)f_{a_1a_2}(\lambda_1-\alpha_1)f_{b_1b_2}(\alpha_1-\lambda_1)\,d\alpha \\ &+\int_{-\pi}^{\pi} H^{(T)}(\alpha)\,H^{(T)}(\lambda_2+\lambda_1-\alpha)f_{a_1b_2}(\lambda_1-\alpha)f_{b_1a_2}(\alpha-\lambda_1)\,d\alpha\Big\} \\ &+2\pi T^{-1}\iint_{-\pi}^{\pi} H^{(T)}(\alpha_1)\,H^{(T)}(\alpha_2)f_{a_1b_1a_2b_2}(\lambda_1-\alpha_1,\alpha_1-\lambda_1,\alpha_2-\lambda_2)\,d\alpha_1\,d\alpha_2 \\ &+O(B_T^{-2}T^{-2})+O(B_T^{-1}T^{-3}\log T) \end{aligned} \tag{8·22}$$

from which (6·4) follows.

Expression (6·5) follows by a similar, but cruder, argument using (3·4).

Proof of Theorem 6·2. We note that all cumulants of order greater than two tend to 0 as $T \to \infty$. This gives the result.

Proof of Theorem 6·3. This follows directly from Theorem 3·2.

I would like to thank Professor J. Durbin for a number of helpful comments.

References

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. New York: Wiley.

Anderson, T. W. & Walker, A. M. (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Statist.* **35**, 1296–1303.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of auto-correlated time series. *J. R. Statist. Soc. Suppl.* **8**, 27–41.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* **37**, 1–16.

Bartlett, M. S. (1966). *An Introduction to Stochastic Processes*, 2nd edition. Cambridge University Press.

Blackman, R. B. & Tukey, J. W. (1958). *The Measurement of Power Spectra*. New York: Dover.

Brillinger, D. R. (1965). An introduction to polyspectra. *Ann. Math. Statist.* **36**, 1351–74.

Brillinger, D. R. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros. *J. R. Statist. Soc.* B **30**, 145–59.

Brillinger, D. R. & Rosenblatt, M. (1967). Asymptotic theory of estimates of k-th order spectra. In *Advanced Seminar on Spectral Analysis of Time Series* (ed. B. Harris), pp. 153–88. New York: Wiley.

Edwards, R. E. (1967). *Fourier Series*, vol. 1. New York: Holt, Rinehart and Winston.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.* **34**, 152–77.

Grenander, U. & Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.

Hannan, E. J. (1960). *Time Series Analysis*. London: Methuen.

Hannan, E. J. (1967). The estimation of a lagged regression relation. *Biometrika* **54**, 409–18.

Ibragimov, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8**, 366–401.

Kawata, T. (1959). Some convergence theorems for stationary stochastic processes. *Ann. Math. Statist.* **30**, 1192–214.

Lamperti, J. (1962). On convergence of stochastic processes. *Trans. Amer. Math. Soc.* **104**, 430–5.

Leonov, V. P. & Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory Probab. Appl.* **4**, 319–29.

Lomnicki, Z. A. & Zaremba, S. K. (1957). On some moments and distributions occurring in the theory of linear stochastic processes. Part I. *Mh. Math.* **61**, 318–58.

Lomnicki, Z. A. & Zaremba, S. K. (1959). On some moments and distributions occurring in the theory of linear stochastic processes. Part II. *Mh. Math.* **63**, 128–68.

Malevich, T. L. (1964). The asymptotic behavior of an estimate for the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **9**, 350–3.

Malevich, T. L. (1965). Some properties of the estimators of the spectrum of a stationary process. *Theory Probab. Appl.* **10**, 457–65.

Olshen, R. A. (1967). Asymptotic properties of the periodogram of a discrete stationary process. *J. Appl. Prob.* **4**, 508–28.

Parzen, E. (1957*a*). A central limit theorem for multilinear stochastic processes. *Ann. Math. Statist.* **28**, 252–5.

Parzen, E. (1957*b*). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28**, 329–48.

Parzen, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32**, 951–89.

Rao, S. T. (1967). On the cross periodogram of a stationary Gaussian vector process. *Ann. Math. Statist.* **38**, 593–7.

Rosenblatt, M. (1959). Statistical analysis of stochastic processes with stationary residuals. In *Studies in Probability and Statistics* (ed. U. Grenander), pp. 246–75. New York: Wiley.

ROSENBLATT, M. (1962). Asymptotic behavior of eigenvalues of Toeplitz forms. *J. Math. Mech.* **11**, 941–50.

SLUTSKY, E. E. (1934). Alcune applicazioni dei coefficienti di Fourier all'analisi delle funzioni aleatorie stazionarie. *Gion. Ist. Ital. Attuari* **5**, 1–50.

TUKEY, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In *Advanced Seminar on Spectral Analysis of Time Series* (ed. B. Harris), pp. 25–46. New York: Wiley.

WAHBA, G. (1968). On the distribution of some statistics useful in the analysis of jointly stationary time series. *Ann. Math. Statist.* **39**, 1849–62.

WALKER, A. M. (1954). The asymptotic distribution of serial correlation coefficients for autoregressive processes with dependent residuals. *Proc. Camb. Phil. Soc.* **50**, 60–4.

WALKER, A. M. (1965). Some asymptotic results for the periodogram of a stationary time series. *J. Aust. Math. Soc.* **5**, 107–28.

[*Received December* 1968. *Revised February* 1969]

Fourier Analysis of Stationary Processes

DAVID R. BRILLINGER

Invited Paper

Abstract—This paper begins with a description of some of the important procedures of the Fourier analysis of real-valued stationary discrete time series. These procedures include the estimation of the power spectrum, the fitting of finite parameter models, and the identification of linear time invariant systems. Among the results emphasized is the one that the large sample statistical properties of the Fourier transform are simpler than those of the series itself. The procedures are next generalized to apply to the cases of vector-valued series, multidimensional time series or spatial series, point processes, random measures, and finally to stationary random Schwartz distributions. It is seen that the relevant Fourier transforms are evaluated by different formulas in these further cases, but that the same constructions are carried out after their evaluation and the same statistical results hold. Such generalizations are of interest because of current work in the fields of picture processing and pulse-code modulation.

I. Introduction

THE FOURIER analysis of data has a long history, dating back to Stokes [1] and Schuster [2], for example. It has been done by means of arithmetical formulas (Whittaker and Robinson [3], Cooley and Tukey [4]), by means of a mechanical device (Michelson [5]), and by means of real-time filters (Newton [6], Pupin [7]). It has been carried out on discrete data, such as monthly rainfall in the Ohio valley (Moore [8]), on continuous data, such as radiated light (Michelson [5]), on vector-valued data, such as vertical and horizontal components of wind speed (Panofsky and McCormick [9]), on spatial data, such as satellite photographs (Leese and Epstein [10]), on point processes, such as the times at which vehicles pass a position on a road (Bartlett [11]), and on point processes in space, such as the positions of pine trees in a field (Bartlett [12]). It has even been carried out on the logarithm of a Fourier transform (Oppenheim *et al.* [13]) and on the logarithm of a power spectrum estimate (Bogert *et al.* [14]).

This invited paper is one of a series planned on topics of general interest—The Editor.

Manuscript received June 7, 1974; revised August 13, 1974. This paper was prepared while the author was a Miller Research Professor and was supported by NSF under Grant GP-31411.

The author is with the Department of Statistics, University of California, Berkeley, Calif. 94720.

The summary statistic examined has been: the Fourier transform itself (Stokes [1]), the modulus of the transform (Schuster [2]), the smoothed modulus squared (Bartlett [15]), the smoothed product of two transforms (Jones [16]), and the smoothed product of three transforms (Hasselman *et al.* [17]).

The summary statistics are evaluated in an attempt to measure population parameters of interest. Foremost among these parameters is the power spectrum. This parameter was initially defined for real-valued-time phenomena (Wiener [18]). In recent years it has been defined and shown useful for spatial series, point processes, and random measures as well. Our development in this paper is such that the definitions set down and mathematics employed are virtually the same for all of these cases.

Our method of approach to the topic is to present first an extensive discussion of the Fourier analysis of real-valued discrete-time series emphasizing those aspects that extend directly to the cases of vector-valued series, of continuous spatial series, of point processes, and finally of random distributions. We then present extensions to the processes just indicated. Throughout, we indicate aspects of the analysis that are peculiar to the particular process under consideration. We also mention higher order spectra and nonlinear systems. Wold [19] provides a bibliography of papers on time series analysis written prior to 1960. Brillinger [20] presents a detailed description of the Fourier analysis of vector-valued discrete-time series.

We now indicate several reasons that suggest why Fourier analysis has proved so useful in the analysis of time series.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_13, © Springer Science+Business Media, LLC 2012

II. WHY THE FOURIER TRANSFORM?

Several arguments can be advanced as to why the Fourier transform has proved so useful in the analysis of empirical functions. For one thing, many experiments of interest have the property that their essential character is not changed by moderate translations in time or space. Random functions produced by such experiments are called *stationary*. (A definition of this term is given later.) Let us begin by looking for a class of functions that behave simply under translation. If, for example, we wish

$$f(t+u) = C_u f(t), \qquad t, u = 0, \pm 1, \pm 2, \cdots$$

with $C_1 \neq 0$, then by recursion

$$f(t) = C_1 f(t-1) = C_2 f(t-2) = \cdots = C_1^t f(0)$$

for $t \geqslant 0$ and so $f(t) = f(0) \exp\{\alpha t\}$ for $\alpha = \ln C_1$. If $f(t)$ is to be bounded, then $\alpha = i\lambda$, for $i = \sqrt{-1}$ and λ real. We have been led to the functions $\exp\{i\lambda t\}$. Fourier analysis is concerned with such functions and their linear combinations.

On the other hand, we might note that many of the operations we would like to apply to empirical functions are linear and translation invariant, that is such that; if $X_1(t) \to Y_1(t)$ and $X_2(t) \to Y_2(t)$ then $\alpha_1 X_1(t) + \alpha_2 X_2(t) \to \alpha_1 Y_1(t) + \alpha_2 Y_2(t)$ and if $X(t) \to Y(t)$ then $X(t-u) \to Y(t-u)$. Such operations are called *linear filters*. It follows from these conditions that if $X(t) = \exp\{i\lambda t\} \to Y_\lambda(t)$ then

$$X(t+u) = \exp\{i\lambda u\}\, X(t) \to \exp\{i\lambda t\}\, Y_\lambda(t) = Y(t+u).$$

Setting $u = t$, $t = 0$ gives $Y_\lambda(t) = \exp\{i\lambda t\}\, Y_\lambda(0)$. In summary, $\exp\{i\lambda t\}$ the complex exponential of frequency λ is carried over into a simple multiple of itself by a linear filter. $A(\lambda) = Y_\lambda(0)$ is called the *transfer function* of the filter. If the function $X(t)$ is a Fourier transform, $X(t) = \int \exp\{i\alpha t\}\, x(\alpha)\, d\alpha$, then from the linearity (and some continuity) $X(t) \to \int \exp i\alpha t\, A(\alpha)\, x(\alpha)\, d\alpha$. We see that the effect of a linear filter is easily described for a function that is a Fourier transform.

In the following sections, we will see another reason for dealing with the Fourier transforms of empirical functions, namely, in the case that the functions are realizations of a stationary process, the large sample statistical properties of the transforms are simpler than the properties of the functions themselves.

Finally, we mention that with the discovery of fast Fourier transform algorithms (Cooley and Tukey [4]), the transforms may often be computed exceedingly rapidly.

III. STATIONARY REAL-VALUED DISCRETE-TIME SERIES

Suppose that we are interested in analyzing T real-valued measurements made at the equispaced times $t = 0, \cdots, T-1$. Suppose that we are prepared to model these measurements by the corresponding values of a realization of a stationary discrete-time series $X(t)$, $t = 0, \pm 1, \pm 2, \cdots$. Important parameters of such a series include its *mean*,

$$c_X = EX(t) \tag{1}$$

giving the average level about which the values of the series are distributed and its *autocovariance function*

$$\begin{aligned} c_{XX}(u) &= \operatorname{cov}\{X(t+u), X(t)\} \\ &= E\{[X(t+u) - c_X][X(t) - c_X]\}, \qquad u = 0, \pm 1, \cdots \end{aligned} \tag{2}$$

providing a measure of the degree of dependence of values of the process $|u|$ time units apart. (These parameters do not depend on t because of the assumed stationarity of the series.) In many cases of interest the series is *mixing*, that is, such that values well separated in time are only weakly dependent in a formal statistical sense to be described later. Suppose, in particular, that $c_{XX}(u) \to 0$ sufficiently rapidly as $|u| \to \infty$ for

$$f_{XX}(\lambda) = (2\pi)^{-1} \sum_{u=-\infty}^{\infty} c_{XX}(u) \exp\{-i\lambda u\}, \qquad -\infty < \lambda < \infty \tag{3}$$

to be defined. The parameter $f_{XX}(\lambda)$ is called the *power spectrum* of the series $X(t)$ at frequency λ. It is symmetric about 0 and has period 2π. The definition (3) may be inverted to obtain the representation

$$c_{XX}(u) = \int_{-\pi}^{\pi} \exp\{i\alpha u\}\, f_{XX}(\alpha)\, d\alpha \tag{4}$$

of the autocovariance function in terms of the power spectrum.

If the series $X(t)$ is passed through the linear filter

$$X(t) \to Y(t) = \sum_u a(t-u)\, X(u)$$

with well-defined transfer function

$$A(\lambda) = \sum_u a(u) \exp\{-i\lambda u\}$$

then we can check that

$$c_{YY}(u) = \sum_u \sum_v a(u+v)\, a(w)\, c_{XX}(w-v) \tag{5}$$

and, by taking Fourier transforms, that

$$f_{YY}(\lambda) = |A(\lambda)|^2 f_{XX}(\lambda) \tag{6}$$

under some regularity conditions. Expression (6), the frequency domain description of linear filtering, is seen to be much nicer than (5), the time-domain description.

Expressions (4) and (6) may be combined to obtain an interpretation of the power spectrum at frequency λ. Suppose that we consider a narrow band-pass filter at frequency λ having transfer function

$$A(\alpha) \doteq \begin{cases} 1, & |\alpha \pm \lambda| \leqslant \Delta \\ 0, & \text{otherwise} \end{cases}$$

with Δ small. Then the variance of the output series $Y(t)$, of the filter, is given by

$$\begin{aligned} \operatorname{var} Y(t) &= c_{YY}(0) \\ &= \int f_{YY}(\alpha)\, d\alpha \\ &= \int |A(\alpha)|^2 f_{XX}(\alpha)\, d\alpha \\ &= 4\Delta f_{XX}(\lambda). \end{aligned} \tag{7}$$

In words, the power spectrum of the series $X(t)$ at frequency λ is proportional to the variance of the output of a narrow band-pass filter of frequency λ. In the case that $\lambda \neq 0, \pm 2\pi, \pm 4\pi, \cdots$

the mean of the output series is 0 and the variance of the output series is the same as its mean-squared value. Expression (7) shows incidentally that the power spectrum is nonnegative.

We mention, in connection with the representation (4), that Khintchine [21] shows that for $X(t)$ a stationary discrete time series with finite second order moments, we necessarily have

$$c_{XX}(u) = \int_{-\pi}^{\pi} \exp \{i\alpha u\}\, dF_{XX}(\alpha) \tag{8}$$

where $F_{XX}(\alpha)$ is a monotonic nondecreasing function. $F_{XX}(\lambda)$ is called the *spectral measure*. Its derivative is the power spectrum. Going along with (8), Cramér [22] demonstrated that the series itself has a Fourier representation

$$X(t) = \int_{-\pi}^{\pi} \exp \{i\alpha t\}\, dZ_X(\alpha), \qquad t = 0, \pm 1, \cdots \tag{9}$$

where $Z_X(\lambda)$ is a random function with the properties;

$$E\, dZ_X(\lambda) = \eta(\lambda)\, c_X\, d\lambda \tag{10}$$

$$\operatorname{cov}\{dZ_X(\lambda), dZ_X(\mu)\} = \eta(\lambda - \mu)\, dF_{XX}(\lambda)\, d\mu. \tag{11}$$

(In these last expressions, if $\delta(\lambda)$ is the Dirac delta function then $\eta(\lambda) = \sum \delta(\lambda - 2\pi j)$ is the Kronecker comb.) Also expression (11) concerns the covariance of two complex-varied variates. Such a covariance is defined by $\operatorname{cov}\{X, Y\} = E\{(X - EX)\overline{(Y - EY)}\}$.) Expression (9) writes the series $X(t)$ as a Fourier transform. We can see that if the series $X(t)$ is passed through a linear filter with transfer function $A(\lambda)$, then the output series has Fourier representation

$$\int_{-\pi}^{\pi} \exp \{i\alpha t\}\, A(\alpha)\, dZ_X(\alpha), \qquad t = 0, \pm 1, \cdots.$$

In Section XV, we will see that the first and second-order relations (10), (11) may be extended to kth order relations with the definition of kth order spectra.

IV. The Finite Fourier Transform

Let the values of the series $X(t)$ be available for $t = 0, 1, 2, \cdots, T-1$ where T is an integer. The *finite Fourier transform* of this stretch of series is defined to be

$$d_X^{(T)}(\lambda) = \sum_{t=0}^{T-1} X(t) \exp\{-i\lambda t\}, \qquad -\infty < \lambda < \infty. \tag{12}$$

A number of interpretations may be given for this variate. For example, suppose we take a linear filter with transfer function concentrated at the frequency λ, namely $A(\alpha) = \delta(\alpha - \lambda)$. The corresponding time domain coefficients of this filter are

$$a(u) = (2\pi)^{-1} \int A(\alpha) \exp\{iu\alpha\}\, d\alpha$$
$$= (2\pi)^{-1} \exp\{iu\lambda\}, \qquad u = 0, \pm 1, \cdots.$$

The output of this filter is the series

$$(2\pi)^{-1} \sum_u X(u) \exp\{i\lambda(t-u)\} \doteq (2\pi)^{-1} \exp\{i\lambda t\}\, d_X^{(T)}(\lambda).$$

These remarks show that the finite Fourier transform may be interpreted as, essentially, the result of narrow band-pass filtering the series.

Before presenting a second interpretation, we first remark that the sample covariance of pairs of values $X(t), Y(t), t = 0, 1, \cdots, T-1$ is given by $T^{-1} \sum X(t)\, Y(t)$, when the $Y(t)$ values have 0 mean. This quantity is a measure of the degree of linear relationship of the $X(t)$ and $Y(t)$ values. The finite Fourier transform is essentially, then, the sample covariance between the $X(t)$ values and the complex exponential of frequency λ. It provides some measure of the degree of linear relationship of the series $X(t)$ and phenomena of exact frequency λ.

In the case that $\lambda = 0$, the finite Fourier transform (12) is the sample sum. The central limit theorem indicates conditions under which a sum of random variables is asymptotically normal as the sample size grows to ∞. Likewise, there are theorems indicating that $d_X^{(T)}(\lambda)$ is asymptotically normal as $T \to \infty$. Before indicating some aspects of these theorems we set down a definition. A complex-valued variate w is called *complex normal* with mean 0 and variance σ^2 when its real and imaginary parts are independent normal variates with mean 0 and variance $\sigma^2/2$. The density function of w is proportional to $\exp\{-|w|^2/\sigma^2\}$. The variate $|w|^2$ is exponential with mean σ^2 in this case.

In the case that the series $X(t)$ is stationary, with finite second-order moments, and mixing (that is, well-separated values are only weakly dependent) the finite Fourier transform has the following useful asymptotic properties as $T \to \infty$:

a) $d_X^{(T)}(0) - Tc_X$ is asymptotically normal with mean 0 and variance $2\pi T f_{XX}(0)$;
b) for $\lambda \neq 0, \pm\pi, \pm 2\pi, \cdots, d_X^{(T)}(\lambda)$ is asymptotically complex normal with mean 0 and variance $2\pi T f_{XX}(\lambda)$;
c) for $s^j(T)$, $j = 1, \cdots, J$ integers with $\lambda^j(T) = 2\pi s^j(T)/T \to \lambda \neq 0, \pm\pi, \pm 2\pi, \cdots$ the variates $d_X^{(T)}(\lambda^1(T)), \cdots, d_X^{(T)}(\lambda^J(T))$ are asymptotically independent complex normals with mean 0 and variance $2\pi T f_{XX}(\lambda)$,
d) for $\lambda \neq 0, \pm\pi, \pm 2\pi, \cdots$ and $U = T/J$ and integer, the variates

$$d_X^{(U)}(\lambda, j) = \sum_{u=0}^{U-1} X(u + jU) \exp\{-i\lambda u\}, \qquad j = 0, \cdots, J-1$$

are asymptotically independent complex normals with mean 0 and variance $2\pi U f_{XX}(\lambda)$.

These results are developed in Brillinger [20]. Related results are given in Section XV and proved in the Appendix. Other references include: Leonov and Shiryaev [23], Picinbono [24], Rosenblatt [25], Brillinger [26], Hannan and Thomson [27]. We have seen that $\exp\{i\lambda t\}\, d_X^{(T)}(\lambda)$ may be interpreted as the result of narrow band-pass filtering the series $X(t)$. It follows that the preceding result b) is consistent with the "engineering folk" theorem to the effect that narrow band-pass noise is approximately Gaussian.

Result a) suggests estimating the mean c_X by

$$c_X^{(T)} = T^{-1} \sum_{t=0}^{T-1} X(t)$$

and approximating the distribution of this estimate by a normal distribution with mean 0 and variance $2\pi f_{XX}(0)/T$. Result b) suggests estimating the power spectrum $f_{XX}(\lambda)$ by the *periodogram*

$$I_{XX}^{(T)}(\lambda) = (2\pi T)^{-1} |d_X^{(T)}(\lambda)|^2 \tag{13}$$

in the case $\lambda \neq 0, \pm 2\pi, \cdots$. We will say more about this statistic later. It is interesting to note, from c) and d), that asymptotically independent statistics with mean 0 and variance proportional to the power spectrum at frequency λ may be obtained by either computing the Fourier transform at particular distinct frequencies near λ or by computing them at the frequency λ but based on different time domains. We warn the reader that the results a)–d) are asymptotic. They are to be evaluated in the sense that they might prove reasonable approximations in practice when the domain of observation is large and when values of the series well separated in the domain are only weakly dependent.

On a variety of occasions we will *taper* the data before computing its Fourier transform. This means that we take a data window $\phi^{(T)}(t)$ vanishing for $t < 0, t > T - 1$, and compute the transform

$$d_X^{(T)}(\lambda) = \sum_t \phi^{(T)}(t) \exp\{-i\lambda t\} X(t) \tag{14}$$

for selected values of λ. One intention of tapering is to reduce the interference of neighboring frequency components. If

$$\Phi^{(T)}(\lambda) = \sum_t \phi^{(T)}(t) \exp\{-i\lambda t\}$$

then the Cramér representation (9) shows that (14) may be written

$$d_X^{(T)}(\lambda) = \int \Phi^{(T)}(\lambda - \alpha)\, dZ_X(\alpha). \tag{15}$$

From what we have just said, we will want to choose $\phi^{(T)}(t)$ so that $\Phi^{(T)}(\alpha)$ is concentrated near $\alpha = 0, \pm 2\pi, \cdots$. (One convenient choice of $\phi^{(T)}(t)$ takes the form $\phi(t/T)$ where $\phi(u) = 0$ for $u < 0, u \geq 1$.) The asymptotic effect of tapering may be seen to be to replace the variance in b) by $2\pi \sum \phi^{(T)}(t)^2 f_{XX}(\lambda)$.

Hannan and Thomson [27] investigate the asymptotic distribution of the Fourier transform of tapered data in a case where $f_{XX}(\lambda)$ depends on T in a particular manner. The hope is to obtain better approximations to the distribution.

V. Estimation of the Power Spectrum

In the previous section, we mentioned the periodogram, $I_{XX}^{(T)}(\lambda)$, as a possible estimate of the power spectrum $f_{XX}(\lambda)$ in the case that $\lambda \neq 0, \pm 2\pi, \cdots$. If result b) holds true, then $I_{XX}^{(T)}(\lambda)$, being a continuous function of $d_X^{(T)}(\lambda)$, will be distributed asymptotically as $|w|^2$, where w is a complex normal variate with mean 0 and variance $f_{XX}(\lambda)$. That is $I_{XX}^{(T)}(\lambda)$ will be distributed asymptotically as an exponential variate with mean $f_{XX}(\lambda)$. From the practical standpoint this is interesting, but not satisfactory. It suggests that no matter how large the sample size T is, the variate $I_{XX}^{(T)}(\lambda)$ will tend to be distributed about $f_{XX}(\lambda)$ with an appreciable scatter. Luckily, results c) and d) suggest means around this difficulty. Following c), the variates $I_{XX}^{(T)}(\lambda^j(T)), j = 1, \cdots, J$ are distributed asymptotically as independent exponential variates with mean $f_{XX}(\lambda)$. Their average

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda^j(T)) \tag{16}$$

will be distributed asymptotically as the average of J independent exponential variates having mean $f_{XX}(\lambda)$. That is, it will be distributed as

$$f_{XX}(\lambda)\chi_{2J}^2/2J \tag{17}$$

where χ_{2J}^2 denotes a chi-squared variate with $2J$ degrees of freedom. The variance of the variate (17) is

$$f_{XX}(\lambda)^2/J = f_{XX}(\lambda)^2\, U/T \tag{18}$$

if $U = T/J$. By choice of J the experimenter can seek to obtain an estimate of which the sampling fluctuations are small enough for his needs. From the standpoint of practice, it seems to be useful to compute the estimate (16) for a number of values of J. This allows us to tailor the choice of J to the situation at hand and even to use different values of J for different frequency ranges. Result d) suggests our consideration of the estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=0}^{J-1} (2\pi U)^{-1} |d_X^{(U)}(\lambda, j)|^2. \tag{19}$$

It too will have the asymptotic distribution (17) with variance (18).

We must note that it is not sensible to take J in (16) and (19) arbitrarily large as the preceding arguments might have suggested. It may be seen from (15) that

$$E I_{XX}^{(T)}(\lambda) = \int_{-\pi}^{\pi} F_T(\lambda - \alpha) f_{XX}(\alpha)\, d\alpha + F_T(\lambda)\, c_X^2 \tag{20}$$

where

$$F_T(\lambda) = (2\pi T)^{-1} \left| \frac{\sin \frac{T\lambda}{2}}{\sin \frac{\lambda}{2}} \right|^2$$

is the Fejér kernel. This kernel, or frequency window, is nonnegative, integrates to 1, and has most of its mass in the interval $(-2\pi/T, 2\pi/T)$. The term in c_X^2 may be neglected for $\lambda \neq 0, \pm 2\pi, \cdots$ and T large. From (16) and (20) we now see that

$$E f_{XX}^{(T)}(\lambda) \doteq \int_{-\pi}^{\pi} J^{-1} \sum_{j=1}^{J} F_T(\lambda^j(T) - \alpha) f_{XX}(\alpha)\, d\alpha. \tag{21}$$

If we are averaging J periodogram values at frequencies $2\pi/T$ apart and centered at λ, then the bandwidth of the kernel of (21) will be approximately $4\pi J/T$. If J is large and $f_{XX}(\alpha)$ varies substantially in the interval $-2\pi J/T < \alpha - \lambda < 2\pi J/T$, then the value of (21) can be very far from the desired $f_{XX}(\lambda)$. In practice we will seek to have J large so that the estimate is reasonably stable, but not so large that it has appreciable bias. This same remark applies to the estimate (19). Parzen [28] constructed a class of estimates such that $E f_{XX}^{(T)}(\lambda) \to f_{XX}(\lambda)$ and var $f_{XX}^{(T)}(\lambda) \to 0$. These estimates have an asymptotic distribution that is normal, rather than χ^2, Rosenblatt [29]. Using the notation preceding these estimates correspond to having J depend on T in such a way that $J_T \to \infty$, but $J_T/T \to 0$ as $T \to \infty$.

Estimates of the power spectrum have proved useful; i) as simple descriptive statistics, ii) in informal testing and discrimination, iii) in the estimation of unknown parameters, and iv) in the search for hidden periodicities. As an example of i), we mention their use in the description of the color of an object, Wright [30]. In connection with ii) we mention the estimation of the spectrum of the seismic record of an event in attempt to see if the event was an earthquake or a nuclear explosion, Carpenter [31], Lampert *et al.* [32]. In case iii), we mention that Munk and MacDonald [33] derived estimates of the fundamental parameters of the rotation of the Earth from the periodogram. Turning to iv), we remind the reader that the original problem that led to the definition of the power spectrum, was that of the search for hidden periodicities. As a modern example, we mention the examination of spectral estimates for the periods of the fundamental vibrations of the Earth, MacDonald and Ness [34].

VI. Other Estimates of the Power Spectrum

We begin by mentioning minor modifications that can be made to the estimates of Section V. The periodograms of (16) may be computed at frequencies other than those of the form $2\pi s/T$, s an integer, and they may be weighted unequally. The periodograms of the estimate (19) may be based on overlapping stretches of data. The asymptotic distributions are not so simple when these modifications are made, but the estimate is often improved. The estimate (19) has another interpretation. We saw in Section IV that $\exp\{i\lambda t\}\, d_X^{(U)}(\lambda, j)$ might be interpreted as the output of a narrow band-pass filter centered at λ. This suggests that (19) is essentially the first power spectral estimate widely employed in practice, the average of the squared output of a narrow band-pass filter (Wegel and Moore [35]). We next turn to a discussion of some spectral estimates of quite different character.

We saw in Section III that if the series $X(t)$ was passed through a linear filter with transfer function $A(\lambda)$, then the output series $Y(t)$ had power spectrum given by $f_{YY}(\lambda) = |A(\lambda)|^2 f_{XX}(\lambda)$. In Section V, we saw that the estimates (16), (19) could have substantial bias were there appreciable variation in the value of the population power spectrum. These remarks suggest a means of constructing an improved estimate, namely: we use our knowledge of the situation at hand to devise a filter, with transfer function $A(\lambda)$, such that the output series $Y(t)$ has spectrum nearer to being constant. We then estimate the power spectrum of the filtered series in the manner of Section V and take $|A(\lambda)|^{-2} f_{YY}^{(T)}(\lambda)$ as our estimate of $f_{XX}(\lambda)$. This procedure is called spectral estimation by prewhitening and is due to Tukey (see Panofsky and McCormick [9]). We mention that in many situations we will be content to just examine $f_{YY}^{(T)}(\lambda)$. This would be necessary were $A(\lambda) = 0$.

One useful means of determining an $A(\lambda)$ is to fit an autoregressive scheme to the data by least squares. That is, for some K, choose $\hat{a}(1), \cdots, \hat{a}(K)$ to minimize

$$\sum |X(t) + a(1)\, X(t-1) + \cdots + a(K)\, X(t-K)|^2$$

where the summation extends over the available data. In this case $\hat{A}(\lambda) = 1 + \hat{a}(1) \exp\{-i\lambda\} + \cdots + \hat{a}(K) \exp\{-i\lambda K\}$. An algorithm for efficient computation of the $\hat{a}(u)$ is given in Wiener [36, p. 136]. This procedure should prove especially effective when the series $X(t)$ is near to being an autoregressive scheme of order K. Related procedures are discussed in Grenander and Rosenblatt [37, p. 270], Parzen [38], Lacoss [39], and Burg [40]. Berk [41] discusses the asymptotic distribution of the estimate $|\hat{A}(\lambda)|^{-2}(2\pi T)^{-1} \sum |X(t) + \hat{a}(1)\, X(t-1) + \cdots + \hat{a}(K)\, X(t-K)|^2$. Its asymptotic variance is shown to be (18) with $U = 2K$.

Pisarenko [42] has proposed a broad class of estimates including the high resolution estimate of Capon [43] as a particular case. Suppose $\hat{\Sigma}$ is an estimate of the covariance matrix of the variate

$$\begin{bmatrix} X(1) \\ \vdots \\ X(U) \end{bmatrix}$$

determined from the sample values $X(0), \cdots, X(T-1)$. Suppose $\hat{\mu}_u$, $\hat{\alpha}_u$, $u = 1, \cdots, U$ are the latent roots and vectors of $\hat{\Sigma}$. Suppose $H(\mu)$, $0 < \mu < \infty$, is a strictly monotonic function with inverse $h(\cdot)$. Pisarenko proposed the estimate

$$h\left(\sum_{u=1}^{U} H(\hat{\mu}_u)(2\pi U)^{-1} \left|\sum_{j=1}^{U} \hat{\alpha}_{uj} \exp\{-i\lambda j\}\right|^2\right). \tag{22}$$

He presents an argument indicating that the asymptotic variance of this estimate is also (18). The hope is that it is less biased. Its character is that of a nonlinear average of periodogram values in contrast to the simple average of (16) and (19). The estimates (16) and (19) essentially correspond to the case $H(\mu) = \mu$. The high resolution estimate of Capon [43] corresponds to $H(\mu) = \mu^{-1}$.

The autoregressive estimate, the high-resolution estimate and the Pisarenko estimates are not likely to be better than an ordinary spectral estimate involving steps of prewhitening, tapering, naive spectral estimation and recoloring. They are probably better than a naive spectral estimate for a series that is a sum of sine waves and noise.

VII. Finite Parameter Models

Sometimes a situation arises in which we feel that the form of the power spectrum is known except for the value of a finite dimensional parameter θ. For example existing theory may suggest that the series $X(t)$ is generated by the mixed moving average autoregressive scheme

$$X(t) + a(1)X(t-1) + \cdots + a(K)X(t-K) = \epsilon(t) + b(1)\epsilon(t-1) + \cdots + b(L)\epsilon(t-L) \tag{23}$$

where U, V are nonnegative integers and $\epsilon(t)$ is a series of independent variates with mean 0 and variance σ^2. The power spectrum of this series is

$$f_{XX}(\lambda;\theta) = \frac{\sigma^2}{2\pi} \frac{|1 + b(1)\exp\{-i\lambda\} + \cdots + b(L)\exp\{-i\lambda L\}|^2}{|1 + a(1)\exp\{-i\lambda\} + \cdots + a(K)\exp\{-i\lambda K\}|^2} \tag{24}$$

with $\theta = \sigma^2, a(1), \cdots, a(K), b(1), \cdots, b(L)$. A number of procedures have been suggested for estimating the parameters of the model (23), see Hannan [44] and Anderson [45], for example.

The following procedure is useful in situations more general than the above. It is a slight modification of a procedure of Whittle [46]. Choose as an estimate of θ the value that maximizes

$$\prod_{0<s<T/2} f_{XX}\left(\frac{2\pi s}{T};\theta\right)^{-1} \exp\left\{-I_{XX}^{(T)}\left(\frac{2\pi s}{T}\right) f_{XX}\left(\frac{2\pi s}{T};\theta\right)^{-1}\right\}. \tag{25}$$

Expression (25) is the likelihood corresponding to the assumption that the periodogram values $I_{XX}^{(T)}(2\pi s/T)$, $0 < s < T/2$, are independent exponential variates with means $f_{XX}(2\pi s/T;\theta)$, $0 < s < T/2$, respectively. Under regularity conditions we can show that this estimate, $\hat{\theta}$, is asymptotically normal with mean

θ and covariance matrix $2\pi T^{-1}A^{-1}(A+B)A^{-1}$ where; if $\nabla f_{XX}(\lambda;\theta)$ is the gradient vector with respect to θ and f_{XXXX} the 4th order cumulant spectrum (see Section XV)

$$A=\int_0^\pi \nabla f_{XX}(\alpha;\theta)\cdot\nabla f_{XX}(\alpha;\theta) f_{XX}(\alpha;\theta)^{-2}\,d\alpha$$

$$B=\int_0^\pi\int_0^\pi \nabla f_{XX}(\alpha;\theta)\cdot\nabla f_{XX}(\beta;\theta) f_{XX}(\alpha;\theta)^{-2} f_{XX}(\beta;\theta)^{-2}\cdot f_{XXXX}(\alpha,-\alpha,-\beta)\,d\alpha\,d\beta.$$

We may carry out the maximization of (25) by a number of computer algorithms, see the discussion in Chambers [47]. In [48], we used the method of scoring. Other papers investigating estimates of this type are Whittle [49], Walker [50], and Dzaparidze [51].

The power spectrum itself may now be estimated by $f_{XX}(\lambda;\hat\theta)$. This estimate will be asymptotically normal with mean $f_{XX}(\lambda;\theta)$ and variance $2\pi T^{-1}\nabla f_{XX}(\lambda;\theta)^{\tau}A^{-1}(A+B)\cdot A^{-1}\nabla f_{XX}(\lambda;\theta)$ following the preceding asymptotic normal distribution for θ. In the case that we model the series by an autoregressive scheme and proceed in the same way, the estimate $f_{XX}(\lambda;\hat\theta)$ has the character of the autoregressive estimate of the previous section.

VIII. Linear Models

In some circumstances we may find ourselves considering a linear time invariant model of the form

$$X(t)=\mu+\sum_{u=-\infty}^{\infty} a(t-u)S(u)+\epsilon(t) \tag{26}$$

where the values $X(t)$, $S(t)$, $t=0,1,\cdots,T-1$ are given, $\epsilon(t)$ is an unknown stationary error series with mean 0 and power spectrum $f_{\epsilon\epsilon}(\lambda)$, the $a(u)$ are unknown coefficients, μ is an unknown parameter, and $S(t)$ is a fixed function. For example, we might consider the linear trend model

$$X(t)=\mu+at+\epsilon(t)$$

with μ and a unknown, and be interested in estimating $f_{\epsilon\epsilon}(\lambda)$. Or we might have taken $S(t)$ to be the input series to a linear filter with unknown impulse-response function $a(u)$, $u=0,\pm1,\cdots$ in an attempt to identify the system, that is, to estimate the transfer function $A(\lambda)=\Sigma\, a(u)\exp\{-i\lambda u\}$ and the $a(u)$. The model (26) for the series $X(t)$ differs in an important way from the previous models of this paper. The series $X(t)$ is not generally stationary, because $EX(t)=\mu+\Sigma\, a(t-u)S(u)$.

Estimates of the preceding parameters may be constructed as follows: define

$$d_X^{(T)}(\lambda)=\sum_{t=0}^{T-1} X(t)\exp\{-i\lambda t\}$$

with similar definitions for $d_S^{(T)}(\lambda)$, $d_\epsilon^{(T)}(\lambda)$. Then (26) leads to the approximate relationship

$$d_X^{(T)}(\lambda)\doteq\mu\sum_{t=0}^{T-1}\exp\{-i\lambda t\}+A(\lambda)d_S^{(T)}(\lambda)+d_\epsilon^{(T)}(\lambda). \tag{27}$$

Suppose $\lambda^1(T),\cdots,\lambda^J(T)\doteq\lambda$ are as in Section IV. Then

$$d_X^{(T)}(\lambda^j(T))\doteq A(\lambda)d_S^{(T)}(\lambda^j(T))+d_\epsilon^{(T)}(\lambda^j(T)) \tag{28}$$

for $j=1,\cdots,J$. Following b) of Section IV, the $d_\epsilon^{(T)}(\lambda^j(T))$ are, for large T, approximately independent complex normal variates with mean 0 and variance $2\pi T f_{\epsilon\epsilon}(\lambda)$. The approximate model (28) is seen to take the form of linear regression. The results of linear least-squares theory now suggest our consideration of the estimates,

$$A^{(T)}(\lambda)=f_{XS}^{(T)}(\lambda)f_{SS}^{(T)}(\lambda)^{-1} \tag{29}$$

and

$$f_{\epsilon\epsilon}^{(T)}(\lambda)=f_{XX}^{(T)}(\lambda)-f_{XS}^{(T)}(\lambda)f_{SS}^{(T)}(\lambda)^{-1}f_{SX}^{(T)}(\lambda)$$

where

$$f_{SX}^{(T)}(\lambda)=J^{-1}\sum_{j=1}^{J}(2\pi T)^{-1}d_S^{(T)}(\lambda^j(T))\overline{d_X^{(T)}(\lambda^j(T))}$$

with similar definitions for $f_{XS}^{(T)}$, $f_{XX}^{(T)}$, $f_{SS}^{(T)}$. The impulse response could be estimated by an expression such as

$$a^{(T)}(u)=P^{-1}\sum_{p=0}^{P-1}A^{(T)}\left(\frac{2\pi p}{P}\right)\exp\left\{\frac{-i2\pi pu}{P}\right\}$$

for some integer P. In some circumstances it may be appropriate to taper the data prior to computing the Fourier transform. In others it might make sense to base the Fourier transforms on disjoint stretches of data in the manner of d) of Section IV.

Under regularity conditions the estimate $A^{(T)}(\lambda)$ may be shown to be asymptotically complex normal with mean $A(\lambda)$ and variance $J^{-1}f_{\epsilon\epsilon}(\lambda)f_{SS}^{(T)}(\lambda)^{-1}$ (see [20]). The degree of fit of the model (26) at frequency λ may be measured by the sample coherence function

$$|R_{XS}^{(T)}(\lambda)|^2=|f_{XS}^{(T)}(\lambda)|^2/[f_{SS}^{(T)}(\lambda)f_{XX}^{(T)}(\lambda)]$$

satisfying

$$f_{\epsilon\epsilon}^{(T)}(\lambda)=[1-|R_{XS}^{(T)}(\lambda)|^2]f_{XX}^{(T)}(\lambda).$$

This function provides a time series analog of the squared coefficient of correlation of two variates (see Koopmans [52]).

The procedure of prefiltering is often essential in the estimation of the parameters of the model (26). Consider a common relationship in which the series $X(t)$ is essentially a delayed version of the series $S(t)$, namely

$$X(t)=\alpha S(t-v)+\epsilon(t)$$

for some v. In this case

$$A(\lambda)=\alpha\exp\{-i\lambda v\},$$

$$d_X^{(T)}(\lambda^j(T))=\alpha\exp\{-i\lambda^j(T)v\}d_S^{(T)}(\lambda^j(T))+d_\epsilon^{(T)}(\lambda^j(T))$$

and

$$f_{XS}^{(T)}(\lambda)=\alpha J^{-1}\sum_j\exp\{-i\lambda^j(T)v\}I_{SS}^{(T)}(\lambda^j(T))+f_{\epsilon S}^{(T)}(\lambda). \tag{30}$$

If v is large, the complex exponential fluctuates rapidly about 0 as j changes and the first term on the right-hand side of (30) may be near 0 instead of the desired $\alpha\exp\{-i\lambda v\}f_{SS}^{(T)}(\lambda)$. A useful prefiltering for this situation is to estimate v by $\hat v$, the lag that maximizes the magnitude of the sample cross-covariance function, and then to carry out the spectral computations on the data $X(t)$, $S(t - \hat{v})$, see Akaike and Yamanouchi [53] and Tick [54]. In general, one should prefilter the $X(t)$ series or the $S(t)$ series or both, so that the relationship between the filtered series is as near to being instantaneous as is possible.

The most important use of the calculations we have described is in the identification of linear systems. It used to be the case that the transfer function of a linear system was estimated by probing the system with pure sine waves in a succession of experiments. Expression (29) shows, however, that we can estimate the transfer function, for all λ, by simply employing a single input series $S(t)$ such that $f_{SS}^{(T)}(\lambda) \neq 0$.

In some situations we may have reason to believe that the system (26) is *realizable* that is $a(u) = 0$ for $u < 0$. The factorization techniques of Wiener [36] may be paralleled on the data in order to obtain estimates of $A(\lambda)$, $a(u)$ appropriate to this case, see Bhansali [55]. In Section IX, we will discuss a model like (26), but for the case of stochastic $S(t)$.

Another useful linear model is

$$X(t) = \theta_1\phi_1(t) + \cdots + \theta_K\phi_K(t) + \epsilon(t)$$

with $\phi_1(t), \cdots, \phi_K(t)$ given functions and $\theta_1, \cdots, \theta_K$ unknown. The estimation of these unknowns and $f_{\epsilon\epsilon}(\lambda)$ is considered in Hannan [44] and Anderson [45]. This model allows us to handle trends and seasonal effects.

Yet another useful model is

$$X(t) = \mu + \rho_1 \sin(\theta_1 t + \alpha_1) + \cdots + \rho_K \sin(\theta_K t + \alpha_K) + \epsilon(t)$$

with $\mu, \rho_1, \theta_1, \alpha_1, \cdots, \rho_K, \theta_K, \alpha_K$ unknown. The estimation of these unknowns and $f_{\epsilon\epsilon}(\lambda)$ is considered in Whittle [49]. It allows us to handle hidden periodicities.

IX. Vector-Valued Continuous Spatial Series

In this section we move on from a consideration of real-valued discrete time series to series with a more complicated domain, namely p-dimensional Euclidean space, and with a more complicated range, namely r-dimensional Euclidean space. This step will allow us to consider data such as: that received by an array of antennas or seismometers, picture or TV, holographic, turbulent field.

Provided we set down our notation judiciously, the changes involved are not dramatic. The notation that we shall adopt includes the following: boldface letters such as $\mathbf{X}$, $\mathbf{a}$, $\mathbf{A}$ will denote vectors and matrices. $\mathbf{A}^\tau$ will denote the transpose of a matrix $\mathbf{A}$, tr $\mathbf{A}$ will denote its trace, det $\mathbf{A}$ will denote its determinant. $E\mathbf{X}$ will denote the vector whose entries are the expected values of the corresponding entries of the vector-valued variate $\mathbf{X}$. $\text{cov}\{\mathbf{X}, \mathbf{Y}\} = E\{(\mathbf{X} - E\mathbf{X})\overline{(\mathbf{Y} - E\mathbf{Y})}^\tau\}$ will denote the covariance matrix of the two vector-valued variates $\mathbf{X}$, $\mathbf{Y}$ (that may have complex entries). t, u, λ will lie in p-dimensional Euclidean space, R^p, with

$$t = (t_1, \cdots, t_p) \qquad dt = dt_1 \cdots dt_p$$
$$u = (u_1, \cdots, u_p) \qquad du = du_1 \cdots du_p$$
$$\lambda = (\lambda_1, \cdots, \lambda_p) \qquad d\lambda = d\lambda_1 \cdots d\lambda_p$$
$$\langle \lambda, t \rangle = \lambda_1 t_1 + \cdots + \lambda_p t_p$$
$$\langle \lambda, u \rangle = \lambda_1 u_1 + \cdots + \lambda_p u_p$$
$$|u| = (u_1^2 + \cdots + u_p^2)^{1/2}$$
$$|\lambda| = (\lambda_1^2 + \cdots + \lambda_p^2)^{1/2}.$$

The limits of integrals will be from $-\infty$ to ∞, unless indicated otherwise.

We will proceed by paralleling the development of Sections III and IV. Suppose that we are interested in analyzing measurements made simultaneously on r series of interest at location t, for all locations in some subset of the hypercube $0 < t_1, \cdots, t_p < T$. Suppose that we are prepared to model the measurements by the corresponding values of a realization of an r vector-valued stationary continuous spatial series $\mathbf{X}(t)$, $t \in R^p$. We define the *mean*

$$\mathbf{c}_X = E\mathbf{X}(t)$$

the *autocovariance function*

$$\mathbf{c}_{XX}(u) = \text{cov}\{\mathbf{X}(t + u), \mathbf{X}(t)\}$$

and the *spectral density matrix*

$$\mathbf{f}_{XX}(\lambda) = (2\pi)^{-p} \int \exp\{-i\langle\lambda, u\rangle\} \mathbf{c}_{XX}(u)\, du, \qquad \lambda \in R^p \quad (31)$$

in the case that the integral exists. (The integral will exist when well-separated values of the series are sufficiently weakly dependent.) The inverse of the relationship (31) is

$$\mathbf{c}_{XX}(u) = \int \exp\{i\langle\lambda, \alpha\rangle\} \mathbf{f}_{XX}(\alpha)\, d\alpha. \quad (32)$$

Let

$$\mathbf{X}(t) \rightarrow \mathbf{Y}(t) = \int \mathbf{a}(t - u)\mathbf{X}(u)\, du$$

be a linear filter carrying the r vector-valued series $\mathbf{X}(t)$ into the s vector-valued series $\mathbf{Y}(t)$. Let

$$\mathbf{A}(\lambda) = \int \mathbf{a}(u) \exp\{-i\langle\lambda, u\rangle\}\, du$$

denote the transfer function of this filter. Then the spectral density matrix of the series $\mathbf{Y}(t)$ may be seen to be

$$\mathbf{f}_{YY}(\lambda) = \mathbf{A}(\lambda)\mathbf{f}_{XX}(\lambda)\overline{\mathbf{A}(\lambda)}^\tau. \quad (33)$$

As in Section III, expressions (32) and (33) may be combined to see that the entry in row j, column k of the matrix $\mathbf{f}_{XX}(\lambda)$ may be interpreted as the covariance of the series resulting from passing the jth and kth components of $\mathbf{X}(t)$ through narrow band-pass filters with transfer functions $A(\alpha) = \delta(\alpha - \lambda)$.

The series has a Cramér representation

$$\mathbf{X}(t) = \int \exp\{i\langle\alpha, t\rangle\}\, d\mathbf{Z}_X(\alpha)$$

where $\mathbf{Z}_X(\lambda)$ is an r vector-valued random function with the properties

$$E\, d\mathbf{Z}_X(\lambda) = \delta(\lambda)\mathbf{c}_X\, d\lambda$$
$$\text{cov}\{d\mathbf{Z}_X(\lambda), d\mathbf{Z}_X(\mu)\} = \delta(\lambda - \mu)\mathbf{f}_{XX}(\lambda)\, d\lambda\, d\mu.$$

If $\mathbf{Y}(t)$ is the filtered version of $\mathbf{X}(t)$, then it has Cramér representation

$$\mathbf{Y}(t) = \int \exp\{i\langle\alpha, t\rangle\} \mathbf{A}(\alpha)\, d\mathbf{Z}_X(\alpha).$$

We turn to a discussion of useful computations when values of the series $X(t)$ are available for t in some subset of the hypercube $0 < t_1, \cdots, t_p < T$. Let $\phi^{(T)}(t)$ be a data window whose support (that is the region of locations where $\phi^{(T)}(t) \neq 0$) is the region of observation of $X(t)$. (We might take $\phi^{(T)}(t)$ of the form $\phi(t/T)$ where $\phi(t) = 0$ outside $0 < t_1, \cdots, t_p < 1$.) We consider the Fourier transform

$$d_X^{(T)}(\lambda) = \int X(t)\phi^{(T)}(t) \exp\{-i(\lambda, t)\}\, dt$$

based on the observed sample values.

Before indicating an approximate large sample distribution for $d_X^{(T)}(\lambda)$, we must first define the complex multivariate normal distribution and the complex Wishart distribution. We say that a vector-valued variate X, with complex entries, is *multivariate complex normal* with mean 0 and covariance matrix Σ when it has probability density proportional to $\exp\{-\overline{X}^\tau \Sigma^{-1} X\}$. We shall say that a matrix-valued variate is *complex Wishart* with n degrees of freedom and parameter Σ when it has the form $X_1\overline{X}_1^\tau + \cdots + X_n\overline{X}_n^\tau$, where $X_1, \cdots, X_n$ are independent multivariate complex normal variates with mean 0 and covariance matrix Σ. In the one dimensional case, the complex Wishart with n degrees of freedom is a multiple of a chi-squared variate with $2n$ degrees of freedom.

In the case that well-separated values of the series $X(t)$ are only weakly dependent, the $d_X^{(T)}(\lambda)$ have useful asymptotic properties as $T \to \infty$. These include:

a') $d_X^{(T)}(0)$ is asymptotically multivariate normal with mean $\int \phi^{(T)}(t)\, dt\, c_X$ and covariance matrix $(2\pi)^p \int \phi^{(T)}(t)^2\, dt\, f_{XX}(0)$;

b') for $\lambda \neq 0$, $d_X^{(T)}(\lambda)$ is asymptotically multivariate complex normal with mean 0 and covariance matrix

$$(2\pi)^p \int \phi^{(T)}(t)^2\, dt\, f_{XX}(\lambda);$$

c') for $\lambda^j(T) \to \lambda \neq 0$, with $\lambda^j(T) - \lambda^k(T)$ not tending to 0 too rapidly, $1 \leq j < k \leq J$, the variates $d_X^{(T)}(\lambda^1(T)), \cdots, d_X^{(T)}(\lambda^J(T))$ are asymptotically independent multivariate complex normal with mean 0 and covariance matrix

$$(2\pi)^p \int \phi^{(T)}(t)^2\, dt\, f_{XX}(\lambda);$$

d') if $\phi_j^{(T)}(t)\phi_k^{(T)}(t) = 0$, for all t, $1 \leq j < k \leq J$, and if $\lambda \neq 0$ the variates

$$d_X^{(T)}(\lambda, j) = \int X(t)\phi_j^{(T)}(t) \exp\{-i(\lambda, t)\}\, dt \tag{34}$$

$j = 1, \cdots, J$ are asymptotically independent multivariate complex normal with mean 0 and respective covariance matrices $(2\pi)^p \int \phi^{(T)}(t, j)^2\, dt\, f_{XX}(\lambda)$, $j = 1, \cdots, J$.

Specific conditions under which these results hold are given in Section XV. A proof is given in the Appendix.

Results a'), b') are forms of the central limit theorem. In result d') the Fourier transforms are based on values of $X(t)$ over disjoint domains. It is interesting to note, from c') and d') that asymptotically independent statistics may be obtained by either taking the Fourier transform at distinct frequencies or at the same frequency, but over disjoint domains.

Result a') suggests estimating the mean c_X by

$$c_X^{(T)} = \frac{\int X(t)\phi^{(T)}(t)\, dt}{\int \phi^{(T)}(t)\, dt}. \tag{35}$$

Result b') suggests the consideration of the *periodogram matrix*

$$I_{XX}^{(T)}(\lambda) = (2\pi)^{-p}\left(\int \phi^{(T)}(t)^2\, dt\right)^{-1} d_X^{(T)}(\lambda)\overline{d_X^{(T)}(\lambda)}^\tau \tag{36}$$

as an estimate of $f_{XX}(\lambda)$ when $\lambda \neq 0$. From b') its asymptotic distribution is complex Wishart with 1 degree of freedom and parameter $f_{XX}(\lambda)$. This estimate is often inappropriate because of its instability and singularity. Result c') suggests the consideration of the estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda^j(T)) \tag{37}$$

where J is chosen large enough to obtain acceptable stability, but not so large that the estimate becomes overly biased. From c') the asymptotic distribution of the estimate (37) is complex Wishart with J degrees of freedom and parameter $f_{XX}(\lambda)$. In the case $J = 1$ this asymptotic distribution is that of $f_{XX}(\lambda)\chi_{2J}^2/2J$. Result d') suggests the consideration of the periodogram matrices

$$I_{XX}^{(T)}(\lambda, j) = (2\pi)^{-p}\left(\int \phi_j^{(T)}(t)^2\, dt\right)^{-1} d_X^{(T)}(\lambda, j)\overline{d_X^{(T)}(\lambda, j)}^\tau \tag{38}$$

$j = 1, \cdots, J$ as estimates of $f_{XX}(\lambda)$, $\lambda \neq 0$. The estimate

$$f_{XX}^{(T)}(\lambda) = J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda, j) \tag{39}$$

will have as asymptotic distribution J^{-1} times a complex Wishart with J degrees of freedom and parameter $f_{XX}(\lambda)$ following result d'). We could clearly modify the estimates (37), (39) by using a finer spacing of frequencies and by averaging periodograms based on data over nondisjoint domains. The exact asymptotic distributions will not be so simple in these cases.

The method of fitting finite parameter models, described in Section VII, extends directly to this vector-valued situation. Result b') suggests the replacement of the likelihood function (25) by

$$\prod_{0 < s_j < S_j} \det f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1} \cdot \exp\left\{-\operatorname{tr} I_{XX}^{(T)}\left(\frac{2\pi s}{T}\right) f_{XX}\left(\frac{2\pi s}{T}; \theta\right)^{-1}\right\} \tag{40}$$

in this new case for some large values $S_1, \cdots, S_p$ such that there is little power left beyond the cutoff frequency $(2\pi S_1/T, \cdots, 2\pi S_p/T)$. Suppose that $\hat{\theta}$ is the value of θ leading to the maximum of (40). Under regularity conditions, we can show that $\hat{\theta}$ is asymptotically normal with mean θ and covariance matrix $2\pi T^{-1}A^{-1}(A + B)A^{-1}$ where if A_{jk}, B_{jk} are row j, column k of A, B

$$A_{jk} = \int_0^{2\pi S/T} \operatorname{tr} \frac{\partial f(\alpha)}{\partial \theta_j} f(\alpha)^{-1} \frac{\partial f(\alpha)}{\partial \theta_k} f(\alpha)^{-1} \, d\alpha$$

$$B_{jk} = \int_0^{2\pi S/T} \sum_a \sum_b \sum_c \sum_d C_{abj}(\alpha) C_{cdk}(\beta) f_{abcd}(\alpha, -\alpha, -\beta) \cdot d\alpha \, d\beta$$

with $C_{abj}(\alpha)$ the entry in row a column b of

$$f(\alpha)^{-1} \frac{\partial f(\alpha)}{\partial \theta_j} f(\alpha)^{-1}.$$

In a number of situations we find ourselves led to consider an $(r + s)$ vector-valued series,

$$\begin{bmatrix} S(t) \\ X(t) \end{bmatrix} \tag{41}$$

satisfying a linear model of the form

$$E\{X(t) | S(u), u \in R^p\} = \mu + \int a(t - u) S(u) \, du \tag{42}$$

for some s vector μ and some $s \times r$ matrix-valued function $a(u)$. The model says that the average level of the series $X(t)$ at position t, given the series $S(t)$, is a linear filtered version of the series $S(t)$. If (41) is a stationary series and if $A(\lambda)$ is the transfer function of the filter $a(u)$, then (42) implies

$$c_X = \mu + A(0) c_S \tag{43}$$

$$f_{XS}(\lambda) = A(\lambda) f_{SS}(\lambda). \tag{44}$$

If we define the error series $\epsilon(t)$ by

$$\epsilon(t) = X(t) - \mu - \int a(t - u) S(u) \, du$$

then the degree of fit of the model (42) may be measured by the error spectral density

$$f_{\epsilon\epsilon}(\lambda) = f_{XX}(\lambda) - f_{XS}(\lambda) f_{SS}(\lambda)^{-1} f_{SX}(\lambda). \tag{45}$$

The relationships (43)–(45) suggest the estimates

$$A^{(T)}(\lambda) = f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} \tag{46}$$

$$\mu^{(T)} = c_X^{(T)} - A^{(T)}(0) c_S^{(T)} \tag{47}$$

$$f_{\epsilon\epsilon}^{(T)}(\lambda) = f_{XX}^{(T)}(\lambda) - f_{XS}^{(T)}(\lambda) f_{SS}^{(T)}(\lambda)^{-1} f_{SX}^{(T)}(\lambda) \tag{48}$$

respectively. The asymptotic distributions of these statistics are given in [26].

If there is a possibility that the matrix $f_{SS}^{(T)}(\lambda)$ might become nearly singular, then we would be better off replacing the estimate (46) by a frequency domain analog of the ridge regression estimate (Hoerl and Kennard [56], Hunt [57]), such as

$$f_{XS}^{(T)}(\lambda) [f_{SS}^{(T)}(\lambda) + kI]^{-1} \tag{49}$$

for some $k > 0$ and I the identity matrix. This estimate introduces further bias, over what was already present, but it is hoped that its increased stability more than accounts for this. In some circumstances we might choose k to depend on λ and to be matrix-valued.

X. Additional Results in the Spatial Series Case

The results of the previous section have not taken any essential notice of the fact that the argument t of the random function under consideration is multidimensional. We now indicate some new results pertinent to the multidimensional character.

In some situations, we may be prepared to assume that the series $X(t)$, $t \in R^p$, is *isotropic*, that is the autocovariance function $c_{XX}(u) = \operatorname{cov}\{X(t + u), X(t)\}$ is a function of $|u|$ only. In this case the spectral density matrix $f_{XX}(\lambda)$ is also rotationally symmetric, depending only on $|\lambda|$. In fact (see in Bochner and Chandrasekharan [58, p. 69])

$$f_{XX}(\lambda) = (2\pi)^{-p/2} |\lambda|^{(2-p)/2} \int_0^\infty |u|^{p/2} \cdot J_{(p-2)/2}(|\lambda||u|) c_{XX}(u) \, d|u| \tag{50}$$

where $J_k(t)$ is the Bessel function of the first kind of order k. The relationship (50) may be inverted as follows,

$$c_{XX}(u) = (2\pi)^{p/2} |u|^{(2-p)/2} \int_0^\infty |\lambda|^{p/2} \cdot J_{(p-2)/2}(|\lambda||u|) f_{XX}(\lambda) \, d|\lambda|.$$

The simplified character of $f_{XX}(\lambda)$ in the isotropic case makes its estimation and display much simpler. We can estimate it by an expression such as

$$J^{-1} \sum_{j=1}^{J} I_{XX}^{(T)}(\lambda^j(T)) \tag{51}$$

where the $\lambda^j(T)$ are distinct, but with $|\lambda^j(T)|$ near $|\lambda|$. There are many more $\lambda^j(T)$ with $|\lambda^j(T)|$ near $|\lambda|$ than there are $\lambda^j(T)$ with $\lambda^j(T)$ near λ. It follows that we generally obtain a much better estimate of the spectrum in this case over the estimate in the general case. Also the number of $\lambda^j(T)$ with $|\lambda^j(T)|$ near $|\lambda|$ increases as $|\lambda|$ increases. If follows that the estimate formed will generally be more stable for the frequencies with $|\lambda|$ large. Examples of power spectra estimated in this manner may be found in Mannos [59].

Another different thing that can occur in the general p dimensional case is the definition of marginal processes and marginal spectra. We are presently considering processes $X(t_1, \cdots, t_p)$. Suppose that for some n, $1 \leq n < p$, we are interested in the process with $t_{n+1}, \cdots, t_p$ fixed, say at $0, \cdots, 0$. By inspection we see that the marginal process $X(t_1, \cdots, t_n, 0, \cdots, 0)$ has autocovariance function $c_{XX}(u_1, \cdots, u_n, 0, \cdots, 0)$. The spectral density matrix of the marginal process is, therefore,

$$(2\pi)^{-n} \int \cdots \int c_{XX}(u_1, \cdots, u_n, 0, \cdots, 0) \cdot \exp\{-i(\lambda_1 u_1 + \cdots + \lambda_n u_n)\} \, du_1 \cdots du_n$$

$$= \int \cdots \int f_{XX}(\lambda_1, \cdots, \lambda_n, \lambda_{n+1}, \cdots, \lambda_p) \cdot d\lambda_{n+1} \cdots d\lambda_p.$$

We see that we obtain the spectral density of the marginal process by integrating the complete spectral density. The same remark applies to the Cramér representation for

$$X(t_1, \cdots, t_n, 0, \cdots, 0) = \int \cdots \int \exp\{i(t_1\lambda_1 + \cdots + t_n\lambda_n)\} \cdot \int \cdots \int dZ_X(\lambda_1, \cdots, \lambda_p).$$

Vector-valued series with multidimensional domain are discussed in Hannan [44] and Brillinger [26].

XI. Additional Results in the Vector Case

In the case that the series $X(t)$ is r vector-valued with $r > 1$, we can describe analogs of the classical procedures of multivariate analysis including for example; i) partial correlation, ii) principal component analysis, iii) canonical correlation analysis, iv) cluster analysis, v) discriminant analysis, vi) multivariate analysis of variance, and vii) simultaneous equations. These analogs proceed from c′) or d′) of earlier section. The procedures listed are often developed for samples from multivariate normal distributions. We obtain the time series procedure by identifying the $d_X^{(T)}(\lambda^j(T))$, $j = 1, \cdots, J$ or $d_X^{(T)}(\lambda, j)$, $j = 0, \cdots, J - 1$ with independent multivariate normals having mean **0** and covariance matrix $(2\pi)^p \int \phi^{(T)}(t)^2\, dt\, f_{XX}(\lambda)$ and substituting into the formulas developed for the classical situation. For example, stationary time series analogs of correlation coefficients are provided by the

$$R_{jk}(\lambda) = f_{jk}(\lambda)/\sqrt{f_{jj}(\lambda) f_{kk}(\lambda)}$$

$$\sim \operatorname{cov}\{d_j^{(T)}(\lambda), d_k^{(T)}(\lambda)\} \Big/ \sqrt{\operatorname{var} d_j^{(T)}(\lambda) \operatorname{var} d_k^{(T)}(\lambda)}$$

the *coherency* at frequency λ of the jth component with the kth component of $X(t)$, where $f_{jk}(\lambda)$ is the entry in row j, column k of $f_{XX}(\lambda)$ and $d_j^{(T)}(\lambda)$ is the entry in row j of $d_X^{(T)}(\lambda)$ for $j, k = 1, \cdots, r$. The parameter $R_{jk}(\lambda)$ satisfies $0 \leqslant |R_{jk}(\lambda)| \leqslant 1$ and is seen to provide a measure of the degree of linear relationship of the series $X_j(t)$ with the series $X_k(t)$ at frequency λ. Its modulus squared, $|R_{jk}(\lambda)|^2$, is called the *coherence*. It may be estimated by

$$R_{jk}^{(T)}(\lambda) = f_{jk}^{(T)}(\lambda) \Big/ \sqrt{f_{jj}^{(T)}(\lambda) f_{kk}^{(T)}(\lambda)}$$

where $f_{jk}^{(T)}(\lambda)$ is an estimate of $f_{jk}(\lambda)$.

As time series papers on corresponding multivariate topics, we mention in case i) Tick [60], Granger [61], Goodman [62], Bendat and Piersol [63], Groves and Hannan [64], and Gersch [65]; in case ii) Goodman [66], Brillinger [67], [20], and Priestley *et al.* [68]; in case iii) Brillinger [67], [20], Miyata [69], and Priestley *et al.* [68]; in case iv) Ligett [70]; in case v) Brillinger [20]; in case vi) Brillinger [71]; in case vii) Brillinger and Hatanaka [72], and Hannan and Terrell [73].

Instead of reviewing each of the time series analogs we content ourselves by indicating a form of discriminant analysis that can be carried out in the time series situation. Suppose that a segment of the r vector-valued series $X(t)$ is available and that its spectral density matrix may be any one of $f_i(\lambda)$, $i = 1, \cdots, I$. Suppose that we wish to construct a rule for assigning $X(t)$ to one of the $f_i(\lambda)$.

In the case of a variate U coming from one of I multivariate normal populations with mean **0** and covariance matrix Σ_i, $i = 1, \cdots, I$, a common discrimination procedure is to define a discriminant score

$$-\tfrac{1}{2} \log \det \Sigma_i - \tfrac{1}{2} U^\tau \Sigma_i^{-1} U$$

for the ith population and then to assign the observation U to the population for which the discriminant score has the highest value (see Rao [74, p. 488]). The discriminant score is essentially the logarithm of the probability density of the ith population.

Result 2) suggests a time series analog for this procedure. If the spectral density of the series $X(t)$ is $f_i(\lambda)$, the log density of $d_X^{(T)}(\lambda)$ is essentially

$$-\log \det f_i(\lambda) - \operatorname{tr} I_{XX}^{(T)}(\lambda) f_i(\lambda)^{-1}. \tag{52}$$

This provides a discriminant score for each frequency λ. A more stable score would be provided by the smoothed version

$$-J^{-1} \log \det f_i(\lambda) - \operatorname{tr} f_{XX}^{(T)}(\lambda) f_i(\lambda)^{-1}$$

with $f_{XX}^{(T)}(\lambda)$ given by (37) or (39). These scores could be plotted against λ for $i = 1, \cdots, I$ in order to carry out the required discrimination. In the case that the $f_i(\lambda)$ are unknown, their values could be replaced by estimates in (52).

XII. Additional Results in the Continuous Case

In Section IX, we changed to a continuous domain in contrast to the discrete domain we began with in Section III. In many problems, we must deal with both sorts of domains, because while the phenomenon of interest may correspond to a continuous domain, observational and computational considerations may force us to deal with the values of the process for a discrete domain. This occurrence gives rise to the complication of *aliasing*. Let Z denote the set of integers, $Z = 0, \pm 1, \cdots$. Suppose $X(t)$, $t \in R^p$, is a stationary continuous spatial series with spectral density matrix $f_{XX}(\lambda)$ and Cramér representation

$$X(t) = \int \exp\{i\langle\alpha, t\rangle\}\, dZ_X(\alpha).$$

Suppose $X(t)$ is observable only for $t \in Z^p$. For these values of t

$$X(t) = \int_{(-\pi, \pi)^p} \exp\{i\langle\alpha, t\rangle\} \sum_{j \in Z^p} dZ_X(\alpha + 2\pi j).$$

This is the Cramér representation of a discrete series with spectral density matrix

$$\sum_{j \in Z^p} f_{XX}(\lambda + 2\pi j).$$

We see that if the series $X(t)$ is observable only for $t \in Z^p$, then there is no way of untangling the frequencies

$$\lambda + 2\pi j, \quad j \in Z^p.$$

These frequencies are called the aliases of the fundamental frequency λ.

XIII. Stationary Point Processes

A variety of problems, such as those of traffic systems, queues, nerve pulses, shot noise, impulse noise, and microscopic theory of gases lead us to data that has the character of times or positions in space at which certain events have occurred. We turn now to indicating how the formulas we have presented so far in this paper must be modified to apply to data of this new character.

Suppose that we are recording the positions in p-dimensional Euclidean space at which events of r distinct types occur. For $j = 1, \cdots, r$ let $X_j(t) = X_j(t_1, \cdots, t_p)$ denote the number of events of the jth type that occur in the hypercube $(0, t_1] \times \cdots \times (0, t_p]$. Let $dX_j(t)$ denote the number that occur in the small hypercube $(t_1, t_1 + dt_1] \times \cdots \times (t_p, t_p + dt_p]$. Suppose that joint distributions of variates such as $dX(t^1), \cdots, dX(t^k)$ are unaffected by simple translation of $t^1, \cdots t^k$, we then say that $X(t)$ is a stationary point process.

Stationary point process analogs of definitions set down previously include

$$c_X \, dt = E \, dX(t) \tag{53}$$

c_X is called the *mean intensity* of the process,

$$dC_{XX}(u) \, dt = \operatorname{cov} \{dX(t + u), dX(t)\} \tag{54}$$

$$f_{XX}(\lambda) = (2\pi)^{-p} \int \exp \{-i\langle\lambda, u\rangle\} \, dC_{XX}(u) \tag{55}$$

$$X(t) = \int \cdots \int \left[\frac{\exp \{i\lambda_1 t_1\} - 1}{i\lambda_1}\right] \cdots \left[\frac{\exp \{i\lambda_p t_p\} - 1}{i\lambda_p}\right] \cdot dZ_X(\lambda_1, \cdots, \lambda_p) \tag{56}$$

$$dX(t) = \int \exp \{i\langle\lambda, t\rangle\} \, dZ_X(\lambda) \, dt \tag{57}$$

$$E\{dX(t)|S(u), u \in R^p\} = \left[\mu + \int a(t - u) \, dS(u)\right] dt. \tag{58}$$

This last refers to an $(r + s)$ vector-valued point process. It says that the instantaneous intensity of the series $X(t)$ at position t, given the location of all the points of the process $S(u)$, is a linear translation invariant function of the process $S(u)$. The locations of the points of $X(t)$ are affected by where the points of $S(u)$ are located. We may define here a stationary random measure $d\epsilon(t)$ by

$$d\epsilon(t) = dX(t) - \left[\mu + \int a(t - u) \, dS(u)\right] dt. \tag{59}$$

We next indicate some statistics that it is useful to calculate when the process $X(t)$ has been observed over some region. The Fourier transform is now

$$d_X^{(T)}(\lambda) = \int \phi^{(T)}(t) \exp \{-i\langle\lambda, t\rangle\} \, dX(t) \tag{60}$$

for the data window $\phi^{(T)}(t)$ whose support corresponds to the domain of observation. If $r = 1$ and points occur at the positions $\tau_1, \tau_2, \cdots$, then this last has the form

$$\phi^{(T)}(\tau_1) \exp \{-i\langle\lambda, \tau_1\rangle\} + \phi^{(T)}(\tau_2) \exp \{-i\langle\lambda, \tau_2\rangle\} + \cdots.$$

We may compute Fourier transforms for different domains in which case we define

$$d_X^{(T)}(\lambda, j) = \int \phi_j^{(T)}(t) \exp \{-i\langle\lambda, t\rangle\} \, dX(t). \tag{61}$$

The change in going from the case of spatial series to the case of point processes is seen to be the replacement of $X(t) \, dt$ by $dX(t)$. In the case that well-separated increments of the process are only weakly dependent, the results a')–d') of Section IX hold without further redefinition.

References to the theory of stationary point processes include: Cox and Lewis [75], Brillinger [76], Daley and Vere-Jones [77], and Fisher [78]. We remark that the material of this section applies equally to the case in which $dX(t)$ is a general stationary random measure, for example with $p, r = 1$, we might take $dX(t)$ to be the amount of energy released by earthquakes in the time interval $(t, t + dt)$. In the next section we indicate some results that do take note of the specific character of a point process.

XIV. New Things in the Point Process Case

In the case of a point process, the parameters c_X, $C_{XX}(u)$ have interpretations further to their definitions (53), (54). Suppose that the process is *orderly*, that is the probability that a small region contains more than one point is very small. Then, for small dt

$$c_j \, dt = E \, dX_j(t) \doteq \Pr[\text{there is an event of type } j \text{ in } (t, t + dt]].$$

It follows that c_j may be interpreted as the intensity with which points of type j are occurring. Likewise, for $u \neq 0$

$$dC_{jk}(u) \, dt = \operatorname{cov} \{dX_j(t + u), dX_k(t)\}$$
$$\doteq \Pr[\text{there is an event of type } j \text{ in } (t + u, t + u + du] \text{ and an event of type } k \text{ in } (t, t + dt]] - c_j c_k \, dt \, du.$$

It follows that

$$\frac{dC_{jk}(u) + c_j c_k \, du}{c_k} = \Pr[\text{event of sort } j \text{ in } (t + u, t + u + du] \text{ given an event of sort } k \text{ in } (t, t + dt]]. \tag{62}$$

In the case that the processes $X_j(t)$ and $X_k(t)$ are independent, expression (62) is equal to $c_j du$.

If the derivative $c_{jk}(u) = dC_{jk}(u)/du$ exists for $u \neq 0$ it is called the *cross-covariance density* of the two processes in the case $j \neq k$ and the *autocovariance density* in the case $j = k$. For many processes

$$dC_{jj}(u) = c_j \delta(u) \, du + c_{jj}(u) \, du$$

and so the power spectrum of the process $X_j(t)$ is given by

$$f_{jj}(\lambda) = (2\pi)^{-p} \left[c_j + \int \exp \{-i\langle\lambda, u\rangle\} \, c_{jj}(u) \, du\right].$$

For a Poisson process $c_{jj}(u) = 0$ and so $f_{XX}(\lambda) = (2\pi)^{-p} c_X$.

The parameter $(2\pi)^p f_{XX}(0)/c_X$ is useful in the classification of real-valued point processes. From 1)

$$\operatorname{var} X(T, \cdots, T) \sim (2\pi)^p T^p f_{XX}(0).$$

It follows that, for large T, $(2\pi)^p f_{XX}(0)/c_X$ is the ratio of the variance of the number of points in the hypercube $(0, T]^p$ for the process $X(t)$ to the variance of the number of points in the same hypercube for a Poisson process with the same intensity c_X. For this reason we say that the process $X(t)$ is *underdispersed* or *clustered* if the ratio is greater than 1 and *overdispersed* if the ratio is less than 1.

The estimation procedure described in Section XI for models with a finite number of parameters is especially useful in the point process case as, typically, convenient time domain estimation procedures do not exist at all. Results of applying such a procedure are indicated in [79].

XV. Stationary Random Schwartz Distributions

In this section, we present the theory of Schwartz distributions (or generalized functions) needed to develop properties of the Fourier transforms of random Schwartz distributions. These last are important as they contain the processes discussed so far in this paper as particular cases. In addition they contain other interesting processes as particular cases, such as processes whose components are a combination of the processes discussed so far and such as the processes with stationary increments that are useful in the study of turbulence, see Yaglom [80]. A further advantage of this abstract approach is that the assumptions needed to develop results are cut back to essentials. References to the theory of Schwartz distributions include Schwartz [81] and Papoulis [82].

Let $\mathfrak{D}$ denote the space of infinitely differentiable functions on R^p with compact support. Let $\mathcal{S}$ denote the space of infinitely differentiable functions on R^p with rapid decrease, that is such that if $\phi^{(q)}(t)$ denotes a derivative of order q then

$$\lim_{|t|\to\infty} (1+|t|)^n \phi^{(q)}(t) \to 0 \quad \text{for all } n, q.$$

A continuous linear functional on $\mathfrak{D}$ is called a *Schwartz distribution* or *generalized function*. The Dirac delta function that we have been using throughout the paper is an example. A continuous linear functional on $\mathfrak{D}$ is called a *tempered distribution*.

Suppose now that a random experiment is being carried out, the possible results of which are continuous linear maps X from $\mathfrak{D}$ to $L^2(P)$, the space of square integrable functions for a probability measure P. Suppose that r of these maps are collected into an r vector, $X(\phi)$. We call $X(\phi)$ an r vector-valued *random Schwartz distribution*. It is possible to talk about things such as $E\,X(\phi)$, $\operatorname{cov}\{X(\phi), X(\psi)\}$ in this case. An important family of transformations on $\mathfrak{D}$ consists of the shifts S^u defined by $S^u\phi(t) = \phi(t+u)$, $t, u \in R^p$. The random Schwartz distribution is called *wide-sense stationary* when

$$E\,X(S^u\phi) = E\,X(\phi)$$

$$\operatorname{cov}\{X(S^u\phi), X(S^u\psi)\} = \operatorname{cov}\{X(\phi), X(\psi)\}$$

for all $u \in R^p$ and $\phi, \psi \in \mathfrak{D}$. It is called *strictly stationary* when all the distributions of finite numbers of values are invariant under the shifts.

Let us denote the convolution of two functions $\phi, \psi \in \mathfrak{D}$ by

$$\phi * \psi(t) = \int \phi(t-u)\,\overline{\psi(u)}\,du$$

and the Fourier transform of a function in $\mathcal{S}$ by the corresponding capital letter

$$\Phi(\lambda) = \int \phi(u) \exp\{-i\langle\lambda, u\rangle\}\,du$$

then we can set down the following Theorem.

Theorem 1: (Ito [83], Yaglom [80].) If $X(\phi)$, $\phi \in \mathfrak{D}$ is a wide-sense stationary random Schwartz distribution, then

$$E\,X(\phi) = c_X \int \phi(t)\,dt \tag{63}$$

$$\operatorname{cov}\{X(\phi), X(\psi)\} = c_{XX}(\phi * \psi^{\sim}) \tag{64}$$

$$= \int \Phi(-\alpha)\,\overline{\Psi(-\alpha)}\,dF_{XX}(\alpha) \tag{65}$$

and

$$X(\phi) = \int \Phi(-\alpha)\,dZ_X(\alpha) \tag{66}$$

where c_X is an r vector, $c_{XX}(\cdot)$ is an $r \times r$ matrix of tempered distributions, $F_{XX}(\lambda)$ is a nonnegative matrix-valued measure satisfying

$$\int (1+|\alpha|)^{-k}\,dF_{XX}(\alpha) < \infty \tag{67}$$

for some nonnegative integer k, and finally $Z_X(\lambda)$ is a random function satisfying

$$E\,dZ_X(\lambda) = \delta(\lambda)\,c_X\,d\lambda \tag{68}$$

$$\operatorname{cov}\{dZ_X(\lambda), dZ_X(\mu)\} = \delta(\lambda-\mu)\,dF_{XX}(\lambda)\,d\mu. \tag{69}$$

The spatial series of Section IX is a random Schwartz distribution corresponding to the functional

$$X(\phi) = \int X(t)\phi(t)\,dt$$

for $\phi \in \mathfrak{D}$. The representations indicated in that section may be deduced from the results of Theorem 1. It may be shown that k of (67) may be taken to be 0 for this case.

The stationary point process of Section XII is likewise a random Schwartz distribution corresponding to the functional

$$X(\phi) = \int \phi(t)\,dX(t)$$

for $\phi \in \mathfrak{D}$. The representations of Section XII may be deduced from Theorem 1. It may be shown that k of (67) may be taken to be 2 for this case.

Gelfand and Vilenkin [84] is a general reference to the theory of random Schwartz distributions. Theorem 1 is proved there.

A linear model that extends those of (42) and (58) to the present situation is one in which the $(r+s)$ vector-valued stationary random Schwartz distribution

$$\begin{bmatrix} S(\phi) \\ X(\phi) \end{bmatrix}$$

satisfies

$$E\{X(\phi) \mid S(\psi), \psi \in \mathfrak{D}\} = \mu \int \phi(t)\,dt + S(\phi * a)$$

$$= \mu\Phi(0) + \int \Phi(-\alpha)A(\alpha)\,dZ_S(\alpha). \tag{70}$$

In the case that the spectral measure is differentiable this last implies that

$$f_{XS}(\lambda) = A(\lambda) f_{SS}(\lambda) \tag{71}$$

suggesting that the system may be identified if the spectral density may be estimated. We next set down a mixing assumption, before constructing such an estimate and determining its asymptotic properties.

Given k variates $X_1, \cdots, X_k$ let cum $\{X_1, \cdots, X_k\}$ denote their joint cumulant or semi-invariant. Cumulants are defined and discussed in Kendall and Stuart [85] and Brillinger [20]. They are the elementary functions of the moments of the variates that vanish when the variates are independent. As such they provide measures of the degree of dependence of variates. We will make use of

Assumption 1. $X(\phi)$ is a stationary random Schwartz distribution with the property that for $\phi_1, \cdots, \phi_k \in \mathcal{S}$ and $a_1, \cdots, a_k = 1, \cdots, r$; $k = 2, 3, \cdots$,

$$\operatorname{cum}\{X_{a_1}(\phi_1), \cdots, X_{a_k}(\phi_k)\} = \int \cdots \int \Phi_1(-\alpha^1) \cdots \cdot \Phi_{k-1}(-\alpha^{k-1}) \Phi_k(\alpha^1 + \cdots + \alpha^{k-1}) \cdot f_{a_1 \cdots a_k}(\alpha^1, \cdots, \alpha^{k-1})\, d\alpha^1 \cdots d\alpha^{k-1} \tag{72}$$

with

$$(1 + |\alpha^1|)^{-m_1} \cdots (1 + |\alpha^{k-1}|)^{-m_{k-1}} |f_{a_1 \cdots a_k} \cdot (\alpha^1, \cdots, \alpha^{k-1})| < L_k$$

for some finite $m_1, \cdots, m_{k-1}, L_k$.

In the case that the spectral measure $F_{XX}(\lambda)$ is differentiable, relation (65) corresponds to the case $k = 2$ of (72). The character of Assumption 1 is one of limiting the size of the cumulants of the functionals of the process $X(\phi)$. It will be shown that it is a form of weak dependence requirement, for functionals of the process that are far apart in t, in the Appendix. The function $f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})$ appearing in (72) is called a *cumulant spectrum* of order k, see Brillinger [86] and the references therein. From (66) we see that it is also given by

$$\operatorname{cum}\{dZ_{a_1}(\lambda^1), \cdots, dZ_{a_k}(\lambda^k)\} = \delta(\lambda^1 + \cdots + \lambda^k) \cdot f_{a_1 \cdots a_k}(\lambda^1, \cdots, \lambda^{k-1})\, d\lambda^1 \cdots d\lambda^k. \tag{73}$$

The fact that it only depends on $k - 1$ arguments results from the assumed stationarity of the process.

Let $\phi^{(T)}(t) = \phi(t/T)$ with $\phi \in \mathfrak{D}$. As an analog of the Fourier transforms of Sections IX and XII we now define

$$d_X^{(T)}(\lambda) = X(\exp\{-i\langle\lambda, \cdot\rangle\}\, \phi^{(T)}) \tag{74}$$

for the stationary random Schwartz distribution $X(\phi)$. We can now state the following theorem.

Theorem 2: If Assumption 1 is satisfied, if $d_X^{(T)}(\lambda)$ is given by (74) and if $T|\lambda^j(T) - \lambda^k(T)| \to \infty$, $1 \leq j < k \leq J$, then 1)–4) of Section IX hold.

This theorem is proved in the Appendix. It provides a justification for the estimation procedures suggested in the paper and for the large sample approximations suggested for the distributions of the estimates.

We end this section by mentioning that a point process with events at positions τ_k, $k = 1, \cdots$ may be represented by the generalized function

$$\sum_k \delta(t - \tau_k)$$

the sampled function of Section III may be represented by the generalized function

$$\sum_{j=-\infty}^{\infty} X(j)\delta(t - j)$$

and that a point process with associated variate S may be represented by

$$\sum_k S_k \delta(t - \tau_k)$$

see Beutler and Leneman [87]. Mathéron [92] discusses the use of random Schwartz distributions in the smoothing of maps.

XVI. Higher Order Spectra and Nonlinear Systems

In the previous section we have introduced the higher order cumulant spectra of stationary random Schwartz distributions. In this section we will briefly discuss the use of such spectra and how they may be estimated.

In the case that the process under consideration is Gaussian, the cumulant spectra of order greater than two are identically 0. In the non-Gaussian case, the higher order spectra provide us with important information concerning the distribution of the process. For example were the process real-valued Poisson on the line with intensity c_N, then the cumulant spectrum of order k would be constant equal to $c_N(2\pi)^{1-k}$. Were the process the result of passing a series of independent identically distributed variates through a filter with transfer function $A(\lambda)$, then the cumulant spectrum of order k would be proportional to

$$A(\lambda^1) \cdots A(\lambda^{k-1}) A(-\lambda^1 - \cdots - \lambda^{k-1}).$$

Such hypotheses might be checked by estimating higher cumulant spectra.

An important use of higher order spectra is in the identification of polynomial systems such as those discussed in Wiener [88] and Brillinger [86] and Halme [89]. Tick [90] shows that if $S(t)$ is a stationary real-valued Gaussian series, if $\epsilon(t)$ is an independent stationary series and if the series $X(t)$ is given by

$$X(t) = \mu + \int a(t - u) S(u)\, du + \iint b(t - u, t - v) S(u) S(v)\, du\, dv + \epsilon(t) \tag{75}$$

then

$$f_{SX}(\lambda) = A(-\lambda) f_{SS}(\lambda)$$

$$f_{SSX}(\lambda, \mu) = 2B(-\lambda, -\mu) f_{SS}(\lambda) f_{SS}(\mu)$$

where

$$A(\lambda) = \int a(u) \exp\{-i\lambda u\}\, du$$

$$B(\lambda, \mu) = \iint b(u, v) \exp\{-i(\lambda u + \mu v)\}\, du\, dv$$

and $f_{SSX}(\lambda, \mu)$ is a third-order cumulant spectrum. It follows that both the linear transfer function $A(\lambda)$ and the bitransfer function $B(\lambda, \mu)$ of the system may be estimated, from estimates of second- and third-order spectra, following the probing of the system by a single Gaussian series. References to the identification of systems of order greater than 2, and to the case of non-Gaussian $S(t)$ are given in [86].

We turn to the problem of constructing an estimate of a kth order cumulant spectrum. In the course of the proof of

Theorem 2 given in the Appendix, we will see that

$$\text{cum}\,\{d_{a_1}^{(T)}(\lambda^1),\cdots,d_{a_k}^{(T)}(\lambda^k)\} \sim \begin{cases}(2\pi)^{p(k-1)}\int \phi^{(T)}(t)^k\,dt\,f_{a_1\cdots a_k}(\lambda^1,\cdots,\lambda^{k-1}), & \text{if } \lambda^1+\cdots+\lambda^{k-1}=0\\ 0, & \text{if } \lambda^1+\cdots+\lambda^{k-1}\neq 0.\end{cases}$$

Suppose that no proper subset of $\lambda^1,\cdots,\lambda^k$ sums to 0. It then follows from the principal relation connecting moments and cumulants that

$$E\{d_{a_1}^{(T)}(\lambda^1)\cdots d_{a_k}^{(T)}(\lambda^k)\} \sim (2\pi)^{p(k-1)}\int \phi^{(T)}(t)^k \cdot dt\, f_{a_1\cdots a_k}(\lambda^1,\cdots,\lambda^{k-1})$$

provided $\lambda^1+\cdots+\lambda^k=0$. This last one suggests the use of kth order periodogram

$$I_{a_1\cdots a_k}^{(T)}(\lambda^1,\cdots,\lambda^{k-1}) = (2\pi)^{-p(k-1)}\left(\int \phi^{(T)}(t)^k dt\right)^{-1} \cdot d_{a_1}^{(T)}(\lambda^1)\cdots d_{a_{k-1}}^{(T)}(\lambda^{k-1}) \times d_{a_k}^{(T)}(-\lambda^1-\cdots-\lambda^{k-1}) \qquad (76)$$

as a naive estimate of the spectrum $f_{a_1\cdots a_k}(\lambda^1,\cdots,\lambda^{k-1})$ provided that no proper subset of $\lambda^1,\cdots,\lambda^{k-1}$ sums to 0. From what we have seen in the case $k=2$ this estimate will be unstable. It follows that we should in fact construct an estimate by smoothing the periodogram (76) over $(k-1)$-tuples of frequencies in the neighborhood of $\lambda^1,\cdots,\lambda^{k-1}$, but such that no proper subset of the $(k-1)$-tuple sums to 0. Details of this construction are given in Brillinger and Rosenblatt [91] for the discrete time case. We could equally well have constructed an estimate using the Fourier transforms $d_X^{(T)}(\lambda, j)$ based on disjoint domains.

Appendix

We begin by providing a motivation for Assumption 1 of Section XIV. Suppose that

$$\text{cum}\,\{X_{a_1}(\phi_1),\cdots,X_{a_k}(\phi_k)\}, \qquad \phi_1,\cdots,\phi_k \in \mathfrak{D}$$

is continuous in each of its arguments. Being a continuous multilinear functional it can be written

$$c_{a_1\cdots a_k}(\phi_1\otimes\phi_2\otimes\cdots\otimes\phi_k)$$

where $c_{a_1\cdots a_k}$ is a Schwartz distribution on $\mathfrak{D}(R^{pk})$, from the Schwartz nuclear theorem. If the process is stationary this distribution satisfies

$$c_{a_1\cdots a_k}(S^u\phi_1\otimes S^u\phi_2\otimes\cdots\otimes S^u\phi_k) = c_{a_1\cdots a_k}\cdot(\phi_1\otimes\phi_2\otimes\cdots\otimes\phi_k).$$

It follows that it has the form

$$\mathcal{C}\left(\int \phi(t+u^1,\cdots,t+u^{k-1},t)\,dt\right)$$

for $\phi\in\mathfrak{D}(R^{pk})$ where $\mathcal{C}$ is a distribution on $\mathfrak{D}(R^{p(k-1)})$.

Now consider the case in which the process $X(\phi)$ has the property that

$$\text{cum}\,\{X_{a_1}(\phi_1),\cdots,X_{a_k}(\phi_k)\} = 0$$

when the supports of $\phi_1,\cdots,\phi_{k-1}$ are farther away from that of ϕ_k than some number ρ. This means that the distribution $\mathcal{C}$ has compact support. By the Schwartz-Paley-Wiener theorem, $\mathcal{C}$ is, therefore, the Fourier transform of a function of slow growth, say $f_{a_1\cdots a_k}(\lambda^1,\cdots,\lambda^{k-1})$ and we may write the relation (72). In the case that values of the process $X(\phi)$ at a distance from each other are only weakly dependent, we can expect the cumulant to be small and for the representation (72) to hold with (73) satisfied.

Proof of Theorem 2: We see from (66) and (73)

$$\begin{aligned}
&\text{cum}\,\{d_{a_1}^{(T)}(\lambda^1),\cdots,d_{a_k}^{(T)}(\lambda^k)\}\\
&= \int\cdots\int \Phi_1^{(T)}(\alpha^1-\lambda^1)\cdots\Phi_{k-1}^{(T)}(\alpha^{k-1}-\lambda^{k-1})\,\Phi_k^{(T)}\\
&\quad\cdot(-\alpha^1-\cdots-\alpha^{k-1}-\lambda^k)\,f_{a_1\cdots a_k}(\alpha^1,\cdots,\alpha^{k-1})\\
&\quad\cdot d\alpha^1\cdots d\alpha^{k-1}\\
&= T^p\int\cdots\int \Phi_1(\beta^1)\cdots\Phi_{k-1}(\beta^{k-1})\,\Phi_k\\
&\quad\cdot(-\beta^1-\cdots-\beta^{k-1}-T(\lambda^1+\cdots+\lambda^k))\,f_{a_1\cdots a_k}\\
&\quad\cdot(\lambda^1+T^{-1}\beta^1,\cdots,\lambda^{k-1}+T^{-1}\beta^{k-1})\,d\beta^1\cdots d\beta^{k-1}\\
&\sim T^p\int\cdots\int \Phi_1(\beta^1)\cdots\Phi_{k-1}(\beta^{k-1})\,\Phi_k(-\beta^1-\cdots-\beta^{k-1})\\
&\quad\cdot d\beta^1\cdots d\beta^{k-1} f_{a_1\cdots a_k}(\lambda^1,\cdots,\lambda^{k-1}),\\
&\qquad\qquad \text{for } \lambda^1+\cdots+\lambda^k=0\\
&= 0(T^p), \qquad\qquad \text{for } \lambda^1+\cdots+\lambda^k\neq 0.
\end{aligned}$$

It follows from this last that the standardized joint cumulants of order greater than 2 tend to 0 and so the Fourier transforms are asymptotically normal.

References

[1] G. G. Stokes, "Note on searching for periodicities," *Proc. Roy. Soc.*, vol. 29, p. 122, 1879.

[2] A. Schuster, "The periodogram of magnetic declination," *Cambridge Phil. Soc.*, vol. 18, p. 18, 1899.

[3] E. T. Whittaker and A. Robinson, *The Calculus of Observations.* Cambridge, England: Cambridge Univ. Press, 1944.

[4] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297-301, 1965.

[5] A. A. Michelson, *Light Waves and Their Uses.* Chicago, Ill.: Univ. Chicago Press, 1907.

[6] I. Newton, *Opticks.* London, England: W. Innys, 1730.

[7] M. I. Pupin, "Resonance analysis of alternating and polyphase currents," *Trans. AIEE*, vol. 9, p. 523, 1894.

[8] H. L. Moore, *Economic Cycles Their Law and Cause.* New York: Macmillan, 1914.

[9] H. A. Panofsky and R. A. McCormick, "Properties of spectra of atmospheric turbulence at 100 metres," *Quart. J. Roy. Meteorol. Soc.*, vol. 80, pp. 546-564, 1954.

[10] J. A. Leese and E. S. Epstein, "Application of two-dimensional spectral analysis to the quantification of satelite cloud photographs," *J. Appl. Meteorol.*, vol. 2, pp. 629-644, 1963.

[11] M. S. Bartlett, "The spectral analysis of point processes," *J. Roy. Stat. Soc.*, vol. B 25, pp. 264-296, 1963.

[12] ——, "The spectral analysis of two dimensional point processes," *Biometrika*, vol. 51, pp. 299–311, 1964.
[13] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, pp. 1264–1291, 1968.
[14] B. P. Bogert, M. J. Healey, and J. W. Tukey, "The quefrency alanysis of time series for echoes: cepstrum, pseudo-covariance, cross-cepstrum and saphe cracking," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 209–243, 1963.
[15] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, pp. 1–16, 1950.
[16] R. H. Jones, "A reappraisal of the periodogram in spectral analysis," *Technometrics*, vol. 7, pp. 531–542, 1965.
[17] K. Hasselman, W. Munk, and G. J. F. MacDonald, "Bispectra of ocean waves," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 125–139, 1963.
[18] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258, 1930.
[19] H. Wold, *Bibliography on Time Series and Stochastic Processes*. London, England: Oliver and Boyd, 1965.
[20] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston, 1974.
[21] A. Ya. Khintchine, "Korrelationstheories der stationären stochastischen Prozesse," *Math. Ann.*, vol. 109, pp. 604–615, 1934.
[22] H. Cramér, "On harmonic analysis in certain functional spaces," *Ark. Mat. Astron. Fys.*, vol. 28B, pp. 1–7, 1942.
[23] V. P. Leonov and A. N. Shiryaev, "Some problems in the spectral theory of higher moments, II," *Theory Prob. Appl. (USSR)*, vol. 5, pp. 460–464, 1960.
[24] B. Picinbono, "Tendence vers le caractère Gaussien par filtrage selectif," *C. R. Acad. Sci. Paris*, vol. 248, p. 2280, 1959.
[25] M. Rosenblatt, "Some comments on narrow band-pass filters," *Quart. Appl. Math.*, vol. 18, pp. 387–393, 1961.
[26] D. R. Brillinger, "The frequency analysis of relations between stationary spatial series," in *Proc. 12th Biennial Seminar of the Canadian Math. Congress*, R. Pyke, Ed. Montreal, P.Q., Canada: Can. Math. Congr., pp. 39–81, 1970.
[27] E. J. Hannan and P. J. Thomson, "Spectral inference over narrow bands," *J. Appl. Prob.*, vol. 8, pp. 157–169, 1971.
[28] E. Parzen, "On consistent estimates of the spectrum of stationary time series," *Ann. Math. Statist.*, vol. 28, pp. 329–348, 1957.
[29] M. Rosenblatt, "Statistical analysis of stochastic processes with stationary residuals," in *Probability and Statistics*, U. Grenander, Ed. New York: Wiley, pp. 246–275, 1959.
[30] W. D. Wright, *The Measurement of Color*. New York: Macmillan. 1958.
[31] E. W. Carpenter, "Explosions seismology," *Science*, vol. 147, pp. 363–373, 1967.
[32] D. G. Lambert, E. A. Flinn, and C. B. Archambeau, "A comparative study of the elastic wave radiation from earthquakes and underground explosions," *Geophys. J. Roy. Astron. Soc.*, vol. 29, pp. 403–432, 1972.
[33] W. H. Munk and G. J. F. MacDonald, *Rotation of the Earth*. Cambridge, England: Cambridge Univ. Press, 1960.
[34] G. J. F. MacDonald and N. Ness, "A study of the free oscillations of the Earth," *J. Geophys. Res.*, vol. 66, pp. 1865–1911, 1961.
[35] R. L. Wegel and C. R. Moore, "An electrical frequency analyzer," *Bell Syst. Tech. J.*, vol. 3, pp. 299–323, 1924.
[36] N. Wiener, *Time Series*. Cambridge, Mass.: M.I.T. Press, 1964.
[37] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.
[38] E. Parzen, "An approach to empirical time series analysis," *Radio Sci.*, vol. 68 D, pp. 551–565, 1964.
[39] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, pp. 661–675, 1971.
[40] J. P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375–376, 1972.
[41] K. N. Berk, "Consistent autoregressive spectral estimates," *Ann. Stat.*, vol. 2, pp. 489–502, 1974.
[42] V. E. Pisarenko, "On the estimation of spectra by means of nonlinear functions of the covariance matrix," *Geophys. J. Roy. Astron. Soc.*, vol. 28, pp. 511–531, 1972.
[43] J. Capon, "Investigation of long-period noise at the large aperture seismic array," *J. Geophys. Res.*, vol. 74, pp. 3182–3194, 1969.
[44] E. J. Hannan, *Multiple Time Series*. New York: Wiley, 1970.
[45] T. W. Anderson, *The Statistical Analysis of Time Series*. New York: Wiley, 1971.
[46] P. Whittle, "Estimation and information in stationary time series," *Ark. Mat. Astron. Fys.*, vol. 2, pp. 423–434, 1953.
[47] J. M. Chambers, "Fitting nonlinear models: numerical techniques," *Biometrika*, vol. 60, pp. 1–14, 1973.
[48] D. R. Brillinger, "An empirical investigation of the Chandler wobble and two proposed excitation processes," *Bull. Int. Stat. Inst.*, vol. 39, pp. 413–434, 1973.
[49] P. Whittle, "Gaussian estimation in stationary time series," *Bull. Int. Stat. Inst.*, vol. 33, pp. 105–130, 1961.
[50] A. M. Walker, "Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary nondeterministic time-series," *J. Australian Math. Soc.*, vol. 4, pp. 363–384, 1964.
[51] K. O. Dzaparidze, "A new method in estimating spectrum parameters of a stationary regular time series," *Teor. Veroyat. Ee Primen.*, vol. 19, p. 130, 1974.
[52] L. H. Koopmans, "On the coefficient of coherence for weakly stationary stochastic processes," *Ann. Math. Stat.*, vol. 35, pp. 532–549, 1964.
[53] H. Akaike and Y. Yamanouchi, "On the statistical estimation of frequency response function," *Ann. Inst. Stat. Math.*, vol. 14, pp. 23–56, 1962.
[54] L. J. Tick, "Estimation of coherency," in *Advanced Seminar on Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967, pp. 133–152.
[55] R. J. Bhansali, "Estimation of the Wiener filter," in *Contributed Papers 39th Session Int. Stat. Inst.*, vol. 1, pp. 82–88, 1973.
[56] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.
[57] B. R. Hunt, "Biased estimation for nonparametric identification of linear systems," *Math. Biosci.*, vol. 10, pp. 215–237, 1971.
[58] S. Bochner and K. Chandrasekharan, *Fourier Transforms*. Princeton, N.J.: Princeton Univ. Press, 1949.
[59] J. Mannos, "A class of fidelity criteria for the encoding of visual images," Ph.D. dissertation, Univ. California, Berkeley, 1972.
[60] L. J. Tick, "Conditional spectra, linear systems and coherency," in *Time Series Analysis*, M. Rosenblatt, Ed. New York: Wiley, pp. 197–203, 1963.
[61] C. W. J. Granger, *Spectral Analysis of Economic Time Series*. Princeton, N.J.: Princeton Univ. Press, 1964.
[62] N. R. Goodman, "Measurement of matrix frequency response functions and multiple coherence functions," Air Force Dynamics Lab., Wright Patterson AFB, Ohio, Tech. Rep. AFFDL-TR-65-56, 1965.
[63] J. S. Bendat and A. Piersol, *Measurement and Analysis of Random Data*. New York: Wiley, 1966.
[64] G. W. Groves and E. J. Hannan, "Time series regression of sea level on weather," *Rev. Geophys.*, vol. 6, pp. 129–174, 1968.
[65] W. Gersch, "Causality or driving in electrophysiological signal analysis," *J. Math. Biosci.*, vol. 14, pp. 177–196, 1972.
[66] N. R. Goodman, "Eigenvalues and eigenvectors of spectral density matrices," Tech. Rep. 179, Seismic Data Lab., Teledyne, Inc., 1967.
[67] D. R. Brillinger, "The canonical analysis of time series," in *Multivariate Analysis–II*, P. R. Krishnaiah, Ed. New York: Academic, pp. 331–350, 1970.
[68] M. B. Priestley, T. Subba Rao, and H. Tong, "Identification of the structure of multivariable stochastic systems," in *Multivariate Analysis–III*, P. R. Krishnaiah, Ed. New York: Academic, pp. 351–368, 1973.
[69] M. Miyata, "Complex generalization of canonical correlation and its application to a sea-level study," *J. Marine Res.*, vol. 28, pp. 202–214, 1970.
[70] W. S. Ligett, Jr., "Passive sonar: Fitting models to multiple time series," paper presented at NATO Advanced Study Institute on Signal Processing, Loughborough, U. K., 1972.
[71] D. R. Brillinger, "The analysis of time series collected in an experimental design," *Multivariate Analysis–III*, P. R. Krishnaiah, Ed. New York: Academic, pp. 241–256, 1973.
[72] D. R. Brillinger and M. Hatanaka, "An harmonic analysis of nonstationary multivariate economic processes," *Econometrica*, vol. 35, pp. 131–141, 1969.
[73] E. J. Hannan and R. D. Terrell, "Multiple equation systems with stationary errors," *Econometrica*, vol. 41, pp. 299–320, 1973.
[74] C. R. Rao, *Linear Statistical Inference and its Applications*. New York: Wiley, 1965.
[75] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*. London, England: Methuen, 1966.
[76] D. R. Brillinger, "The spectral analysis of stationary interval functions," *Proc. 6th Berkeley Symp. Math. Stat. Prob. Vol. 1*, L. M. Le Cam, J. Neyman, and E. L. Scott, Eds. Berkeley, Calif.: Univ. California Press, pp. 483–513, 1972.
[77] D. J. Daley and D. Vere-Jones, "A summary of the theory of point processes," in *Stochastic Point Processes*, P. A. W. Lewis, Ed. New York: Wiley, pp. 299–383, 1972.
[78] L. Fisher, "A survey of the mathematical theory of multidimensional point processes," in *Stochastic Point Processes*, P. A. W. Lewis, Ed. New York: Wiley, pp. 468–513, 1972.
[79] A. G. Hawkes and L. Adamopoulos, "Cluster models for earthquakes—regional comparisons," *Bull. Int. Stat. Inst.*, vol. 39, pp. 454–460, 1973.
[80] A. M. Yaglom, "Some classes of random fields in n-dimensional space related to stationary random processes," *Theory Prob. Appl. (USSR)*, vol. 2, pp. 273–322, 1959.
[81] L. Schwartz, *Théorie des Distributions, Vols. 1, 2.* Paris, France: Hermann, 1957.

Solid-State Control of Electric Drives

ROBERT G. SCHIEMAN, MEMBER, IEEE, EDWARD A. WILKES, AND
HOWARD E. JORDAN, SENIOR MEMBER, IEEE

Invited Paper

***Abstract*—A tutorial review of the dc and ac electric-drive field is presented. The goal is to present fundamental concepts, principle concerns, and key developments in electric-drive technology. Principles of ac and dc power converters and ac and dc motors are presented. Then the combination of the converter and motor to provide a complete drive system is discussed along with drive-system characteristics and methods for analyzing performance. Finally, some application guidelines for both ac and dc systems are given.**

I. Introduction

THE GROWTH of electric drives has closely paralleled the growth of automation in industry. Electric-drive systems provide a convenient means for controlling the operation of industrial machinery. The high reliability and great versatility of electric drives has resulted in their widespread application. In size, electric drives range all the way from fractions of one horsepower up to thousands of horsepower. Speeds range from stalled positioning systems up to 15 000 rev/min and higher.

Historically, the first electric-drive system to gain real prominence was the Ward Leonard System, patented by H. Ward Leonard in the 1890's. The history of dc electric drives proceeded from the basic Ward Leonard principle to various modifications thereof, in approximately the following steps:

1) rheostat control of generator field;
2) tandem field rheostat control of generator field and motor field;
3) thyratron control of generator and motor fields and later thyraton control of the armature voltage of small dc motors;
4) ignitron and mercury pool control of the armature voltage of dc machines too large for thyratrons;
5) magnetic amplifier control of generator field and motor armature voltage; and
6) thyristor control of generator and motor fields and later thyristor control of armature voltage.

During the latter part of the era of the thyratron, the transistor started to replace vacuum tubes in drive regulators. Now solid-state electronic circuits are used to implement special compensating circuits that significantly improve feedback control system response. Microelectronic circuits, particularly operational amplifiers, are used extensively in drive systems today. The operational amplifier circuits are the key to drive-system response, stability, and regulation.

The ac motor variable speed drive development is very similar to the dc.

Initially, the motor alternator set with field rheostats was used to control the ac motor speed. Then other methods of ac motor control were developed. They are as follows:

1) wound rotor resistance control to vary speed with torque load;
2) methods of replacing the resistor in the rotor with other rotating machinery or rectifiers to pump the power back into the ac line;
3) ac motor stator voltage control by the use of resistors, reactors, magnetic amplifiers, thyratrons, ignitrons, mercury-pool tubes or thyristors; and
4) replacement of the motor-alternator set for varying voltage and frequency to the motor with static devices.

This invited paper is one of a series planned on topics of general interest—The Editor.

Manuscript received March 16, 1973; revised April 26, 1974 and August 6, 1974.

The authors are with the Reliance Electric Company, Cleveland, Ohio 44117.

The Canadian Journal of Statistics
Vol. 21, No. 1, 1993, Pages 1–19
La Revue Canadienne de Statistique

The digital rainbow: Some history and applications of numerical spectrum analysis*

David R. BRILLINGER

The University of California, Berkeley

Key words and phrases: Fourier analysis, inverse problem, nuclear magnetic resonance, periodogram, seismology, spectrum, time series.
AMS 1985 subject classifications: Primary 01A99,42-03, 60G12, 62M15.

ABSTRACT

Statistical concepts and techniques are basic to scientific investigation. One concept that enjoys both a theoretical and a physical existence is the spectrum. A spectrum may be described as a display of the intensity or variability of a phenomenon versus period or frequency. Spectra are particularly useful in the study of systems subject to resonance, but have many other uses. This paper begins with some of the historical development of the field, describing a sequence of contributions by Michelson, Schuster, Einstein, Fisher, Bartlett, Tukey, and Whittle. The paper next presents collaborative applications to the study of the free oscillations of the earth, to the dispersion of seismic surface waves and to nuclear-magnetic-resonance spectroscopy. Finally, there is mention of open problems and opinions on future directions.

RÉSUMÉ

Les concepts et les techniques statistiques sont à la base de toute étude scientifique. Le concept de spectre existe tant dans un cadre théorique que dans un cadre physique. Un spectre peut être décrit comme étant la représentation de l'intensité ou de la variabilité d'un phénomème en fonction de la période ou de la fréquence. Les spectres sont particulièrement utiles lors de l'étude de systèmes soumis à une résonance, mais ont également bien d'autres emplois. Cet article débute par un historique des développements dans le domaine, décrivant les contributions de Michelson, Schuster, Einstein, Fisher, Bartlett, Tukey et Whittle. Cet historique est suivi d'applications des spectres à l'étude des oscillations libres de la Terre, à la dispersion des ondes sismiques de surface et à la spectroscopie de résonance magnétique nucléaire. Enfin, questions et opinions pour des recherches futures sont mentionnées.

1. INTRODUCTION

Gerhard Herzberg's research field is the analysis of the spectra of molecules in order to determine their structure. He did this experimentally by passing light through a prism, as Newton had so many years earlier. On the other hand, statisticians have been concerned primarily with spectra as theoretical parameters. It is noteworthy for a concept to have these distinct existences.

To begin, this paper presents a historical development. A recent highlight of the history of spectra is a just-noticed paper by Albert Einstein. This paper, written in 1914, laid out

*This paper is based on the Herzberg Lecture presented at Carleton University 14 February 1992. The lecture is meant as a vehicle for scientists speaking on topics of general interest to academics and to the public at large. This research was supported in part by the National Science Foundation Grant DMS-8900613.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_14, © Springer Science+Business Media, LLC 2012

a practical definition of the power spectrum and a corresponding estimation procedure; see Yaglom (1987a, b). An estimate of the spectrum of sunspot numbers, computed as Einstein might have, is provided here. The work of A.A. Michelson, A. Schuster, R.A. Fisher, M.S. Bartlett, J.W. Tukey, and P. Whittle is highlighted. The second part of the paper provides applications of spectral techniques to the phenomena of free oscillations of the earth, of seismic surface waves, and of nuclear-magnetic resonance spectroscopy. The applications each show a progression from a spectrum analysis to a conceptual-model-based analysis.

By presenting a combination of history and examples it is hoped to interest other statisticians in the topic and to display how statistical concepts can interact with scientific ones.

2. HISTORICAL DEVELOPMENT

Spectrum analysis has its roots in physical science. In 1666, when Newton employed a prism to cast the rainbow on the wall, he coined the term spectrum and began the formal study of the subject. His work was part quantitative, with the counting of the number of colors and the measurement of the widths of the bands; see Topper (1990). In 1801 J.F.W. Herschel measured the temperature at various positions along the image; see Sobel (1989). Herschel's work is more in line with the idea of the spectrum as measuring intensity. A related important step was provided by Gouy (1886), who proposed the representation of light by a Fourier expansion.

2.1. Early Numerical Work.

Michelson (1892) was concerned with finding a length standard. To this end he caused particular substances to emit light. Then, via a mirror, he superposed that light on itself with a time delay. When the superposed light was viewed appropriately, fringes could be seen. The phenomenon is referred to as interference and is understood if one views the original signal as $\cos \lambda t$ and hence after superposition as $\cos \lambda t + \cos \lambda(t+u)$, t denoting time, u delay and λ frequency. As Michelson changed u the clearness of the fringes varied and was recorded. This gave a function $V(u)$, the visibility curve. A variety of examples of the visibility curves he found are presented in Michelson (1892, 1902).

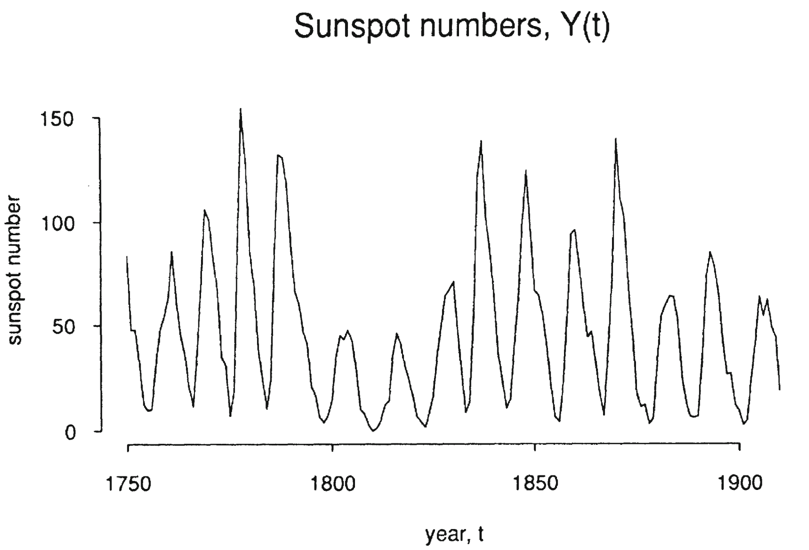
Supposing $f(\lambda) = g(\lambda - \lambda_0)$ to denote the spectrum (for the moment undefined) of the light source at frequency λ to be narrow and centered at λ_0, Michelson (1891) argued that the visibility curve is given by

$$V(u) = \sqrt{\{\int \cos u\lambda\, g(\lambda)\, d\lambda\}^2 + \{\int \sin u\lambda\, g(\lambda)\, d\lambda\}^2}. \tag{1}$$

Michelson called the problem of determining $g(\cdot)$ from $V(\cdot)$ an *inverse problem* and obtained answers by guessing. One of his most important inferences was that the red hydrogen line was a doublet. This inference of splitting led ultimately to important developments in quantum mechanics. Rayleigh (1892) however pointed out that the inverse problem of (1) did not necessarily have a unique solution.

2.2. The Periodogram.

By 1898 Michelson had developed a "harmonic analyser" (Michelson and Stratton 1898). In Michelson (1913) it was employed to compute Schuster's periodogram for



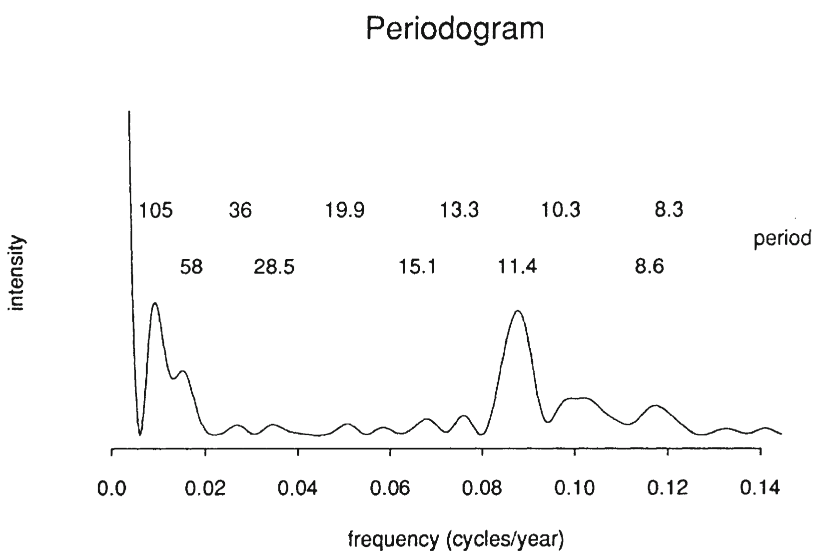


FIGURE 1: The top graph shows monthly relative sunspot numbers for 1750 to 1910. The bottom graph is the periodogram computed for those data. The periods listed are those that Michelson (1913) indicated.

some sunspot data studied by Kimura (1913). For data $Y(t)$, $t = 0, \ldots, T-1$, and period τ, the Schuster periodogram is given by

$$P^T(\tau) = \left(\sum_{t=0}^{T-1} Y(t) \cos \frac{2\pi t}{\tau}\right)^2 + \left(\sum_{t=0}^{T-1} Y(t) \sin \frac{2\pi t}{\tau}\right)^2 . \tag{2}$$

The square root of this quantity was introduced in Schuster (1898). The basic components, in (2), are seen to be correlations of the time-series data with cosine and sine functions respectively. In consequence the periodogram may be expected to highlight periodicities in a series $Y(t)$.

Figure 1 presents the stretch of data analyzed by Michelson, and an attempt at reproducing the periodogram he computed. There were 161 observations in the series. The periodogram of Figure 1 was derived numerically, while Michelson employed the harmonic analyzer. Because Michelson's analyzer could handle only 80 observations and there are 161 here, the reproduction is necessarily approximate. The numbers in the figure are the periods that Michelson mentions, and he seems to mention one for each bump in his estimate. The main hump, in the center of the picture, is near the traditional period of 11 years. The large value near frequency 0 occurs because Michelson did not remove the mean of the data prior to the Fourier analysis.

2.3. The Dark Ages.

Michelson (1913) listed 11 periods for the sunspot series. In a like manner Beveridge (1922) lists 19 periods for a wheat price index. Figure 2 provides Beveridge's data and its periodogram. Many peaks are apparent. Another early researcher eager to ascribe peaks in a periodogram to periodicities was Brownlee (1917), who listed 7 periods for measles data for 1838 to 1913 and in fact remarked:

> It might also be said, in the language of VOLTAIRE, that, if these periods were not found, they would require to be invented.

Frightening words to a statistician. Of such period-chasing, Tukey (1980) remarks:

> More lives have been lost looking at the raw periodogram than by any other action involving time series!

Something was clearly amiss with the naive use of the periodogram.

2.4. Progress to Understanding.

In fact Schuster (1898) had been concerned with whether peaks in the periodogram might simply be due to chance. For an individual τ he assessed their significance (prob-value) via the result

$$\text{Prob}\left\{\frac{P^T(\tau)}{\text{ave}(P^T)} > x\right\} \simeq e^{-x}, \tag{3}$$

where $\text{ave}(P^T)$ refers to the average of all the periodogram values.

Later Fisher (1929) recognized that the periodogram was being examined not just at a single period, but as a function of τ. He derived the more pertinent expression for

$$\text{Prob}\left\{\max_{\tau} \frac{P^T(\tau)}{\text{ave}(P^T)} > x\right\}. \tag{4}$$

Both of the results (3) and (4) were derived for the case of Gaussian white noise. It was for Whittle (1952) to derive the needed result for a stationary time series. He found that in Fisher's result one replaces $P^T(\tau)$ by $P^T(\tau)/f(2\pi/\tau)$, with $f(\cdot)$ being the true spectrum of the series.

2.5. Definition and Estimation of the Spectrum.

In a remarkable paper Einstein (1914), mentioning sunspots as a motivating example, may be seen to define the spectrum and to provide an estimate. He sets down a "mean value"

$$c(u) = \underset{t}{\text{ave}}\,\{Y(t+u)Y(u)\} \tag{5}$$

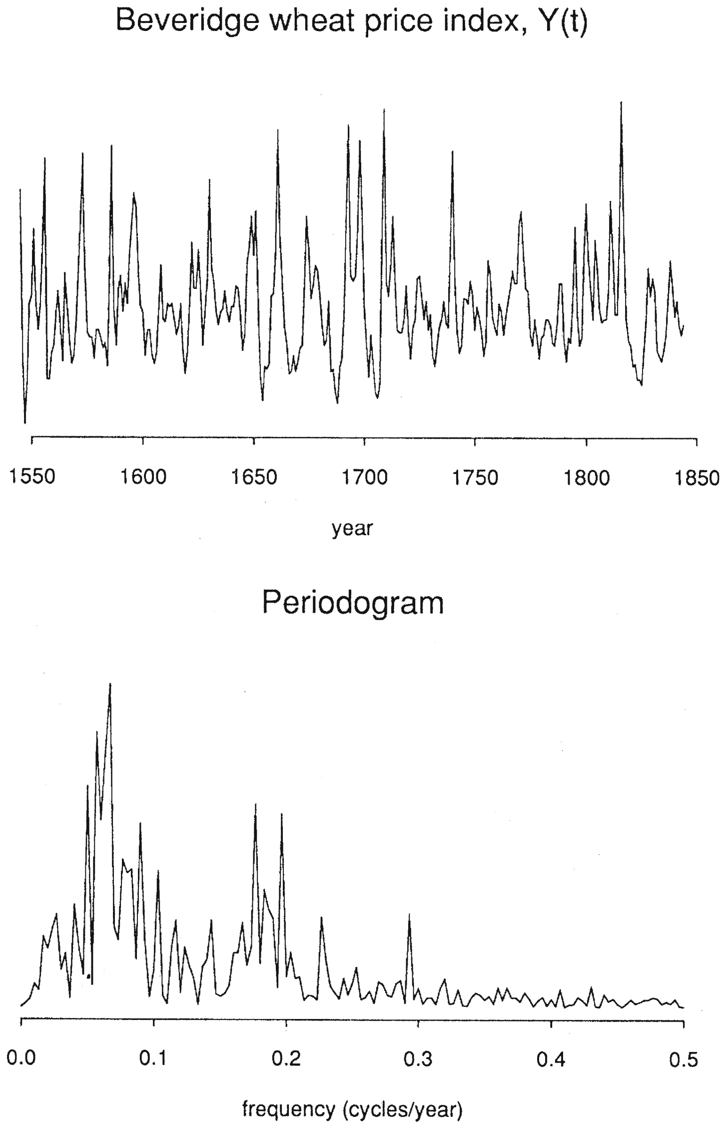


FIGURE 2: The top graph shows the wheat price index data of Beveridge (1922). The bottom graph is the corresponding periodogram, on a linear scale.

and then considers its Fourier transform

$$f(\lambda) = \int \cos \lambda u \; c(u) \; du, \tag{6}$$

referring to the intensity of $Y(\cdot)$.

To estimate $f(\cdot)$, Einstein proposed taking the "mean value" near n of the A_n^2 of the development

$$Y(t) = \sum_n A_n \cos\left(\frac{\pi n t}{T}\right), \qquad 0 < t < T. \tag{7}$$

These A_n are given by

$$\frac{1}{T} \int_0^T \cos\left(\frac{n\pi t}{T}\right) \; Y(t) \; dt \tag{8}$$

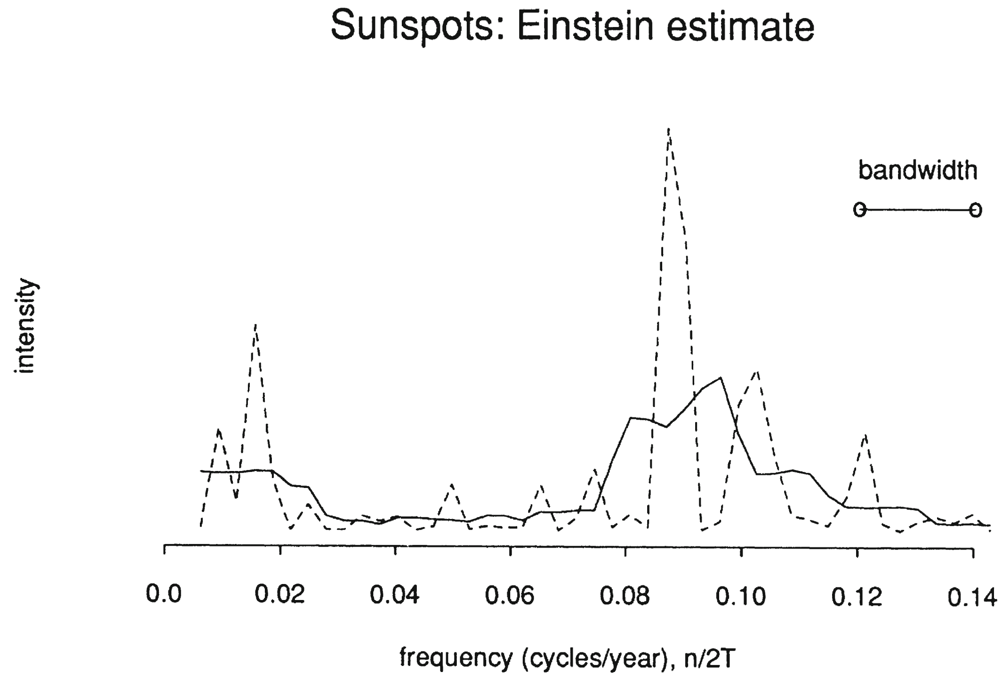


FIGURE 3: An estimate of the spectrum of the sunspot data, smoothing the A_n^2 of (7) over the indicated bandwidth.

Here A_n corresponds to fluctuations at frequency $\lambda = n\pi/T$. For the sunspot data considered above, the results of computing the A_n^2, employing a discrete approximation to (8), are given in Figure 3 as the dashed line. The solid line corresponds to taking a "mean value" over the interval indicated by "bandwidth" on the figure, which produces a much less pronounced 11-year period (0.09 cycles/year).

As provisos, one has to say that Einstein does not make it clear what the "mean value" referred to is, nor whether he had stochastic functions in mind. The notation employed here differs from his.

2.6. The Modern Era.

The modern era of spectrum analysis may be said to begin in the research of Maurice Bartlett and John Tukey in the late 1940s. In particular one may point to the references Tukey and Hamming (1949), Bartlett (1950), and Tukey (1950). The work of these individuals provided an effective estimate of the power spectrum of a stationary time series, given sufficient data. Their research further determined useful approximations to the sampling fluctuations of the estimates. Figure 4 provides an estimate for the sunspot data analyzed earlier. The dashed lines provide approximate 95% confidence limits for each frequency about a heavily smoothed version of the periodogram. From the figure it is apparent that sunspots are far from having a precise period of 11 years. Examination of the series itself shows the lengths of the cycles to vary from 7 to 14 years and for the cycles to be of different shapes.

2.7. Discussion.

It now seems that (6), the formal definition of the power spectrum, is due to Einstein,

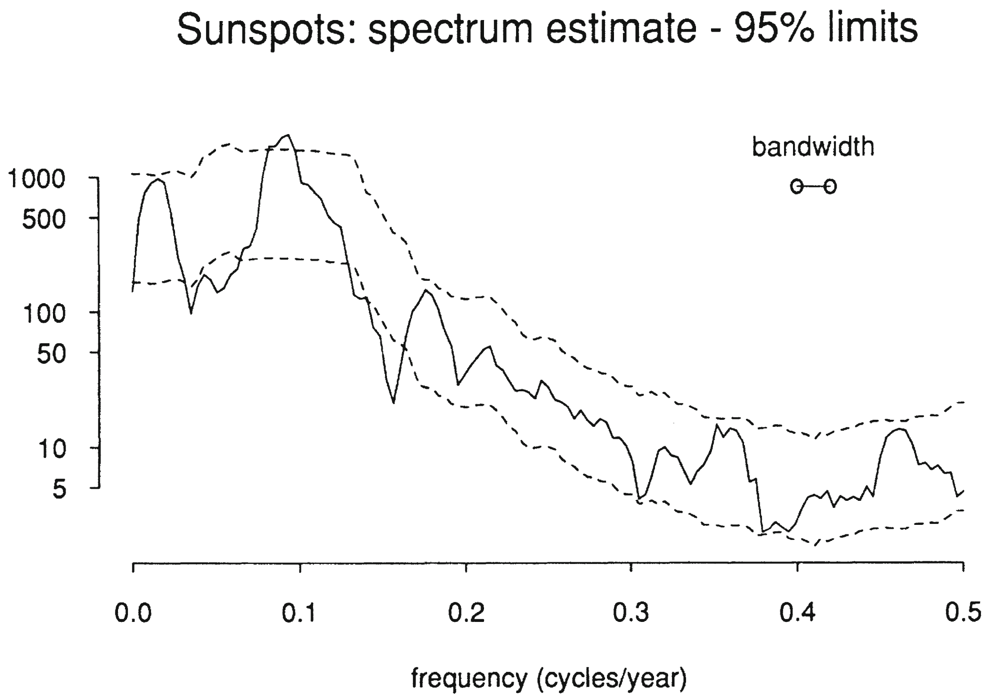


FIGURE 4: An estimate of the power spectrum of the sunspot series with (marginal) confidence bounds indicated.

not Wiener (1930) as had been thought. It also appears that the first effective estimate is due to Einstein, not Wiener (1930) or Daniell (1946) as might be claimed. However, it was Wiener's work that had the influence on the development of the subject. Also some (Masani 1986) doubt Einstein's priority.

It needs to be mentioned that various researchers have made notable contributions to the statistical study of power spectra. One reference to the history and details of others' work is Yaglom (1987b). Another reference on the general history of spectrum analysis is Robinson (1982).

3. SOME APPLICATIONS

Three analyses of scientific data are now presented. This work required close collaboration with substantive scientists. The examples have in common: that frequency analysis elucidates the situation, that a physically based model later provides insight, that a numerical approach is highly flexible, and that a statistical approach handles error and uncertainty.

3.1. Free Oscillations of the Earth.

The empirical spectral analysis of the earth's motion following the great Chilean earthquake of 1960 is viewed by many as *the* success story of numerical spectrum analysis of the sixties. See Tukey (1966), Bath (1974), for example.

After a great earthquake, the Earth "rings" for days at various resonance frequencies (Press 1965). The particular resonance frequencies depend on the structure of the Earth. It is this structure which has long intrigued geophysicists and seismologists. In particular

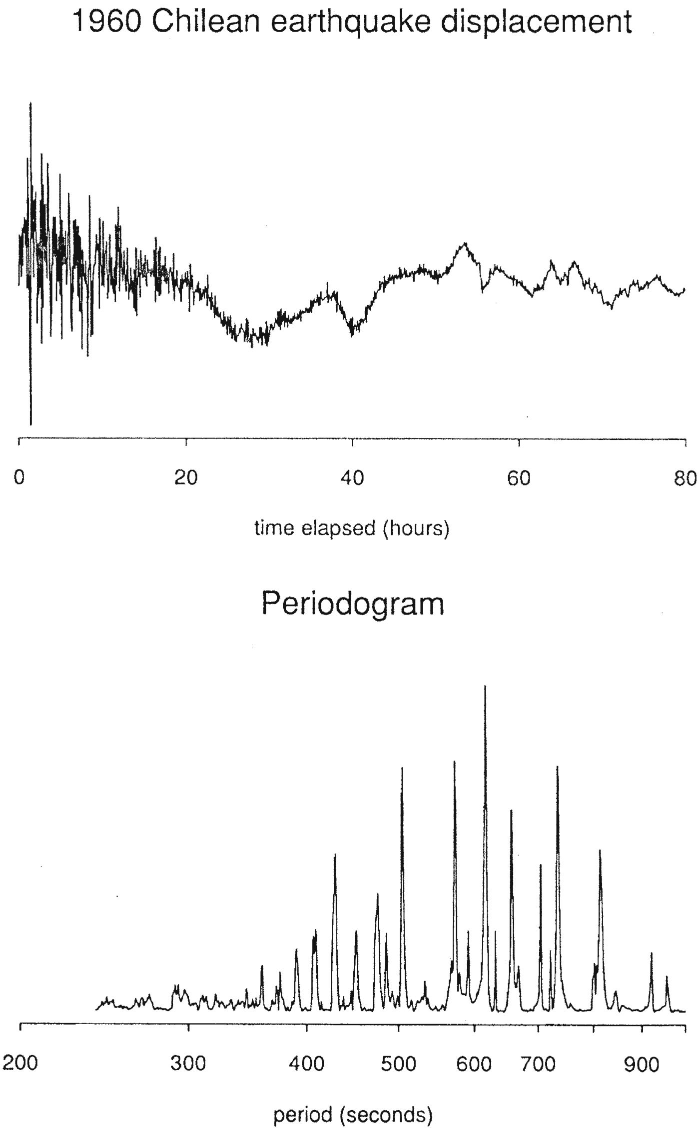


FIGURE 5: The trace of the great Chilean earthquake, corrected for tides, and the corresponding periodogram, on a linear scale.

they have studied the inverse problem of inferring the Earth's structure from the resonance frequencies.

Figure 5 presents a record of the Chilean event recorded at Trieste. The data is described in Bolt and Marusi (1962). The lower figure is the periodogram of that record, graphed against period. (This is the usual seismologist's display.) A variety of peaks are apparent. Values for the periods may be read off, and it seems that the values have varying uncertainty. The periodogram is clearly an important display for assessing the situation, but the trace is far from stationary, and more is needed. In an attempt to estimate the uncertainty (and this is why the seismologist turned to a statistician), one is led to the following steps.

The equations of motion of the earth are approximately linear with constant coefficients

— see Aki and Richards (1980). Such equations have solutions of the form

$$\sum_k \alpha_k e^{-\beta_k t} \cos(\gamma_k t + \delta_k), \qquad t > 0; \tag{9}$$

see Hochstadt (1975). One is thus lead to choose estimates of the α, β, γ, δ so that (9) is near $Y(t)$ for the given data.

A likelihood-motivated analysis of the Chilean data is carried out in Bolt and Brillinger (1979), separately by frequency band. The approximate standard deviations of $\hat{\alpha}_k$ and $\hat{\beta}_k/T$, when derived, are found to have an interesting and intuitive form. Namely they are both proportional to

$$\sqrt{f(\gamma_k)/\alpha_k^2 T^3},$$

where $f(\cdot)$ is the noise spectrum. The estimate of γ_k is more precise for small $f(\lambda_k)$, for large α_k, and for large T.

The derived frequency estimates and associated standard errors can now be taken as input data to the inverse problem of determining the structure of the earth.

3.2. Seismic Surface Waves.

Seismic surface waves are earthquake waves whose energy is trapped near the Earth's surface. They have the property that their velocity of transmission depends on frequency. Figure 6 presents an example of the velocity-frequency relationship for one earth model and Rayleigh waves.

The relationship may be validated by the computation of an empirical dynamic spectrum. This is a display of the estimated intensity of the phenomenon as a function of both time and frequency. A naive way to compute such a spectrum is as

$$\left(\sum_u Y(t-u)\cos\lambda u\right)^2 + \left(\sum_u Y(t-u)\sin\lambda u\right)^2, \tag{10}$$

where u sums over a restricted time interval. Ridges will appear if the different frequency components travel with different velocities. Figure 7 presents, at the top, the vertical displacement trace of the 7 December 1988 magnitude-7.0 Armenian earthquake, as recorded at Berkeley. Below, on the same time scale, is the corresponding dynamic spectrum. One sees the lower-frequency components arriving first, about 1000 seconds after the record starts. In this computation, first a running autoregression was fitted, then (10) was computed based on residuals. The statistic (10) was then corrected for the autoregressive fit. This prewhitening and recoloring is done in order to reduce the bias.

The geophysicist and seimologist hope to learn about the structure of the earth from such data. Figure 8 provides an example of a simple earth model, one of a homogeneous surface layer above an infinite underlayer. For given values of depth h, compressional velocities α_1, α_2, shear velocities β_1, β_2, and densities ρ_1, ρ_2 one can compute curves such as the one of Figure 6; see Bolt and Butcher (1960). This suggests that one should be able to put together the arrival times of the various frequency components with the theoretical arrival times for a given earth model parameter θ, and thereby estimate θ.

In particular one can proceed as follows: determine $\hat{t}(\lambda)$, the time at which the frequency λ component arrives. One knows the distance x from the observatory to the earthquake location. Thus velocities $\hat{U}(\lambda) = x/\hat{t}(\lambda)$ may be computed. Next, suppose one measures the nearness of model parameter θ to the true value by the sum of squares

$$\sum_\lambda \{\hat{U}(\lambda) - U(\lambda|\theta)\}^2,$$

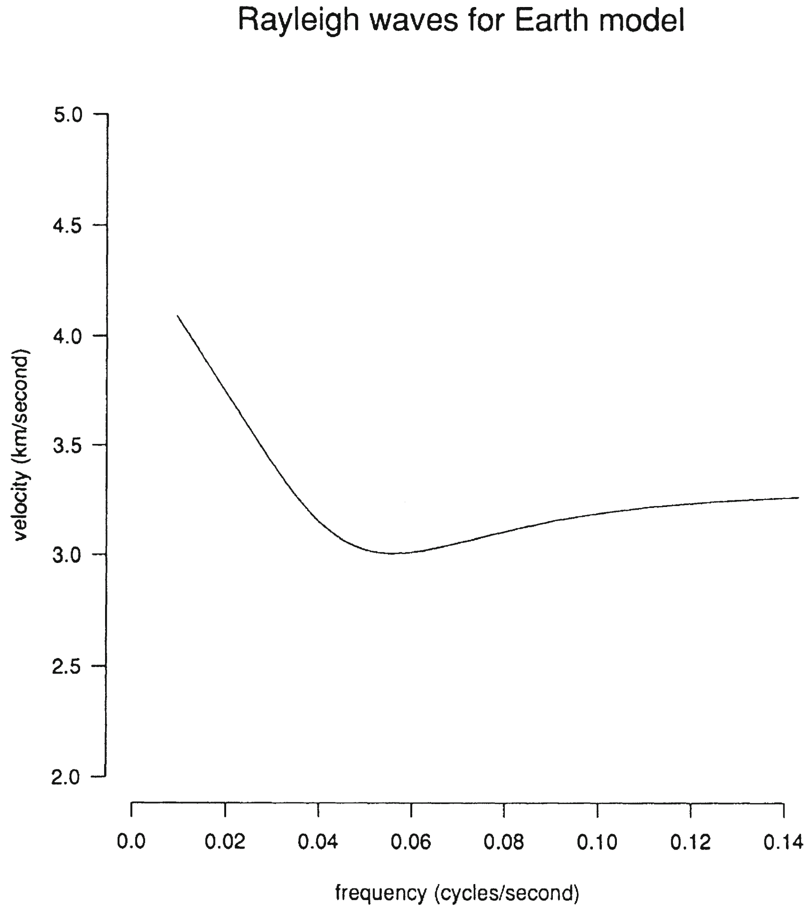


FIGURE 6: Velocity as a function of frequency for an earth model of the form of Figure 8.

summing over the various λ, and then estimates θ by the minimizing value. Figure 9 gives the $\hat{U}(\lambda)$ and the fitted curve $U(\lambda|\theta)$ for the Berkeley station. The fit appears reasonable, particularly at the lower frequencies where the signal-to-noise ratio is greatest. One can aggregate the data for several stations to obtain a group solution and further obtain estimated standard errors for the parameters. Details are provided in Brillinger (1993). A full likelihood procedure for the semiparametric modelling of the seismogram itself is currently under development with B.A. Bolt.

In this example the idea of spectrum has been central to the development of an estimation procedure for parameters which have direct physical interpretation.

3.3. NMR Spectroscopy.

Nuclear magnetic resonance (NMR) is a quantum-mechanical phenomenon. A resonance effect occurs in particular substances when an oscillation frequency of a surrounding radio frequency field coincides with a nuclear precession frequency of the substance. The purpose of the spectroscopy is to infer the structure of the substance. The data consist of the fluctuating voltage response $Y(t)$ to an applied magnetic field $X(t)$. Various inputs $X(\cdot)$ are employed, e.g., a pulse, sequences of fluctuating pulses, and sine waves. Becker and Farrar (1972) is one review paper.

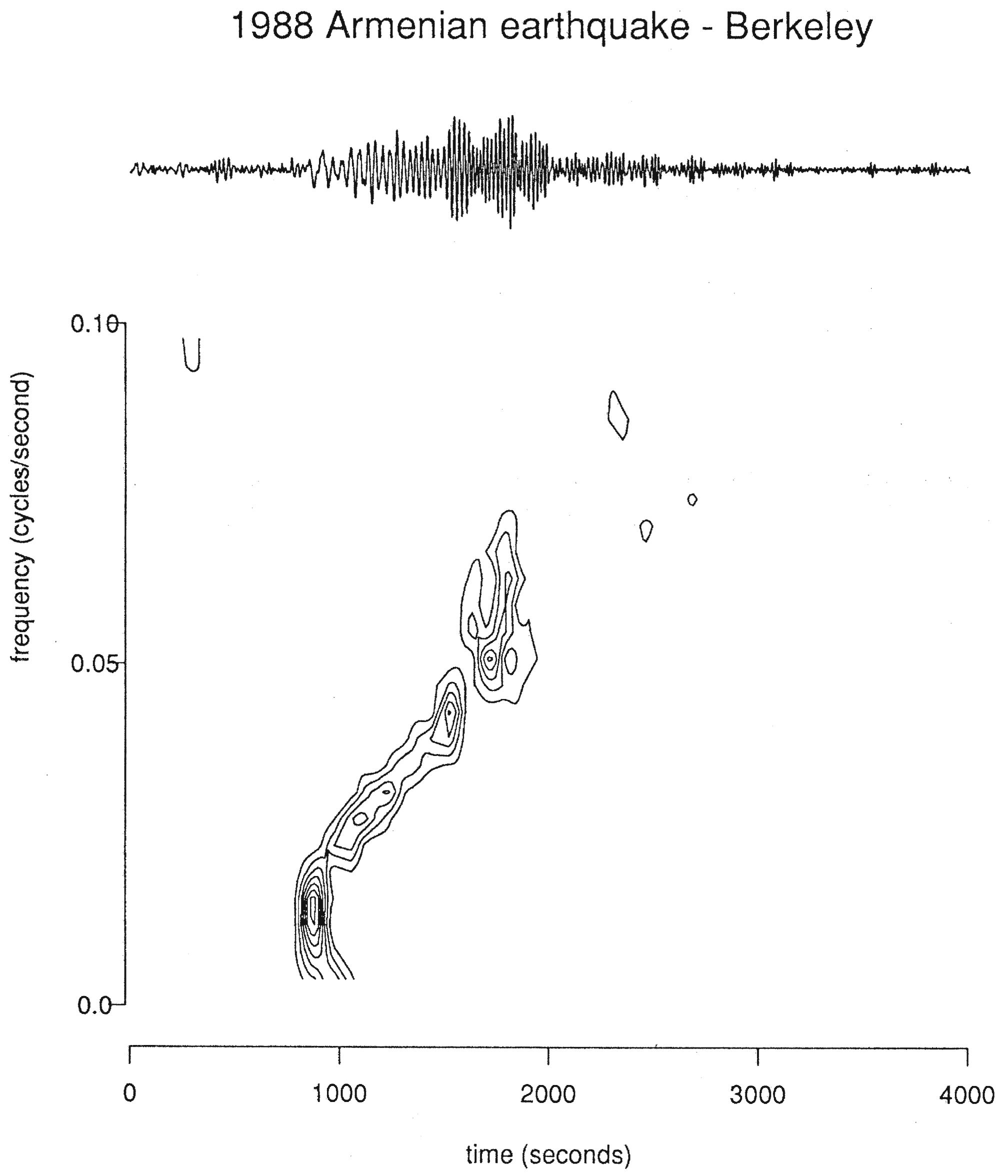


FIGURE 7: The top graph gives the trace of the 7 December 1988 earthquake in northern Armenia. Below, on the same time scale, is a dynamical spectrum expressed as a contour plot.

Ernst and Anderson (1966) proposed the computation of the periodogram

$$\left(\sum_t Y(t)\cos\lambda t\right)^2 + \left(\sum_t Y(t)\sin\lambda t\right)^2,$$

$0 \le \lambda \le \pi$, for such data following a pulse input. An example is given in Figure 10 in the case of a simulated model for the substance 2,3-dibromothiophene (2,3-DBT). The top graph is the response itself, the bottom the absolute value of the Fourier transform of the response. There are two doublets. These result from two hydrogen atoms of the 2,3-DBT. It is much easier to read the lower graph than to puzzle out the structure from the upper one.

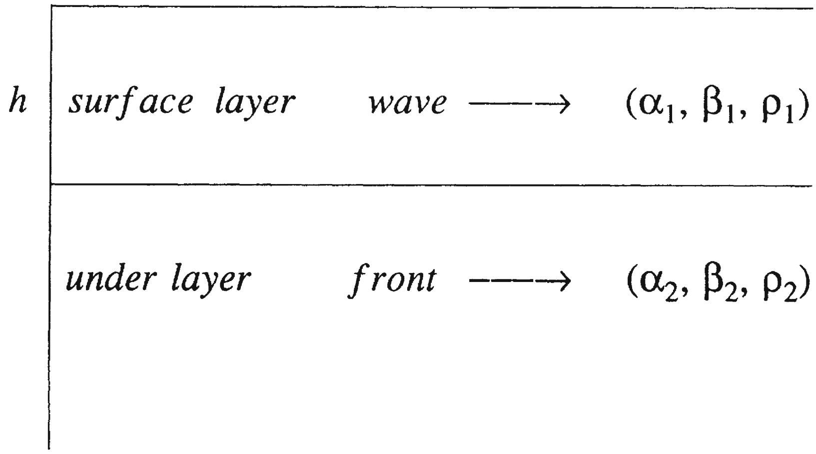


FIGURE 8: Schematic of an earth model of a surface layer of depth h over a half space. The wave propagates from left to right. The parameters are represented by α, β, ρ.

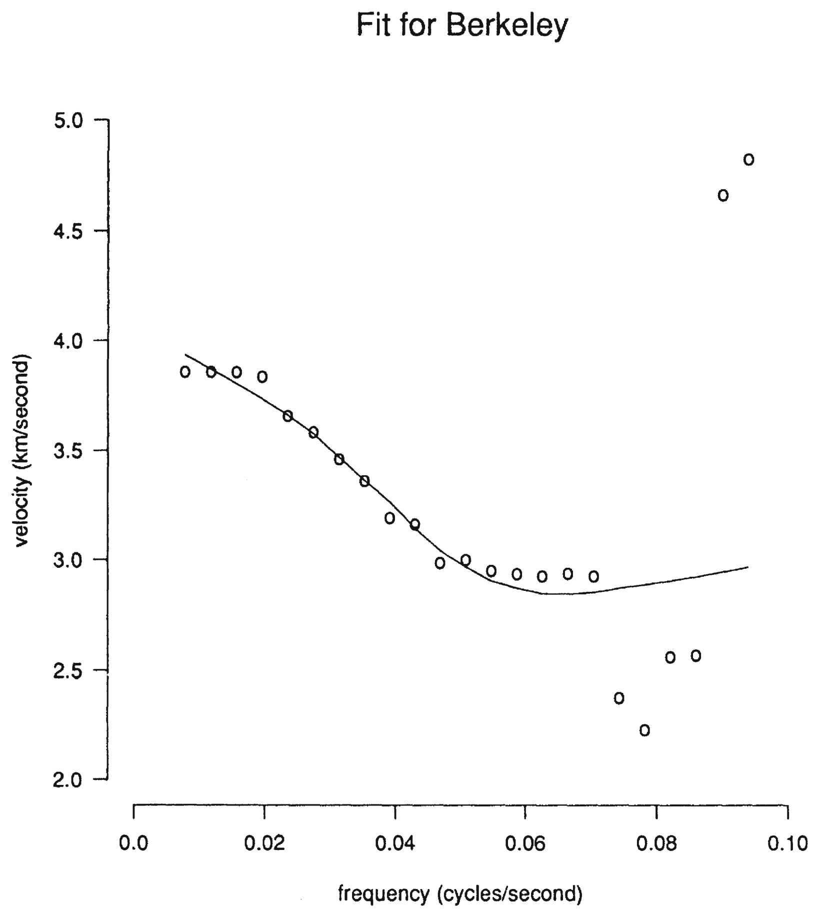
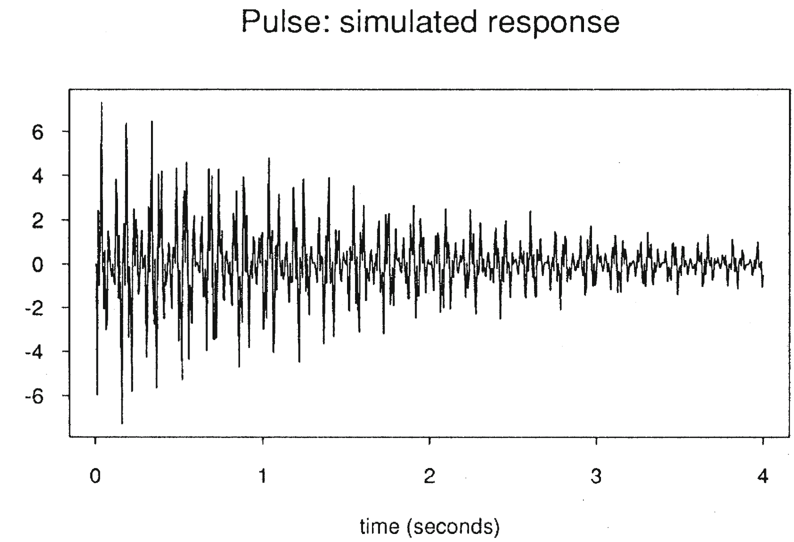


FIGURE 9: For each frequency, the points give the velocity at which the dynamic spectrum is largest. The line is the fitted curve.

Later Ernst (1970) and Kaiser (1970) proposed taking the input $X(\cdot)$ to be random or pseudorandom and employing a form of cross-correlation analysis. Figure 11 shows 4 seconds of actual response of an experimental sample of 2,3-DBT to binary-noise input.



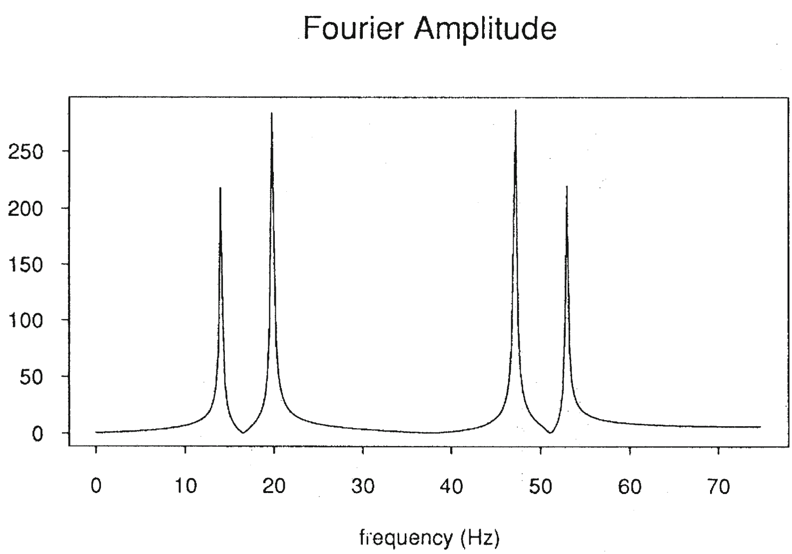


FIGURE 10: Top: plot of response of a simulated 2,3-DBT molecule to a pulse. Bottom: the absolute value Fourier transform of the response.

The lower part of the figure provides the square root of the periodogram. As in Figure 10, two doublets are apparent, but reading their locations is not easy. Since the input $X(\cdot)$ is available, cross-spectral analysis is a more appropriate tool. Figure 12 provides an estimate of the modulus of the transfer function of the system,

$$\frac{|\hat{f}_{YX}(\lambda)|}{\hat{f}_{XX}(\lambda)}, \tag{11}$$

where $\hat{f}_{YX}(\lambda)$ is obtained by smoothing the cross-periodogram

$$(2\pi T)^{-1}\left(\sum_t Y(t)e^{-i\lambda t}\right)\left(\sum_t X(t)e^{i\lambda t}\right)$$

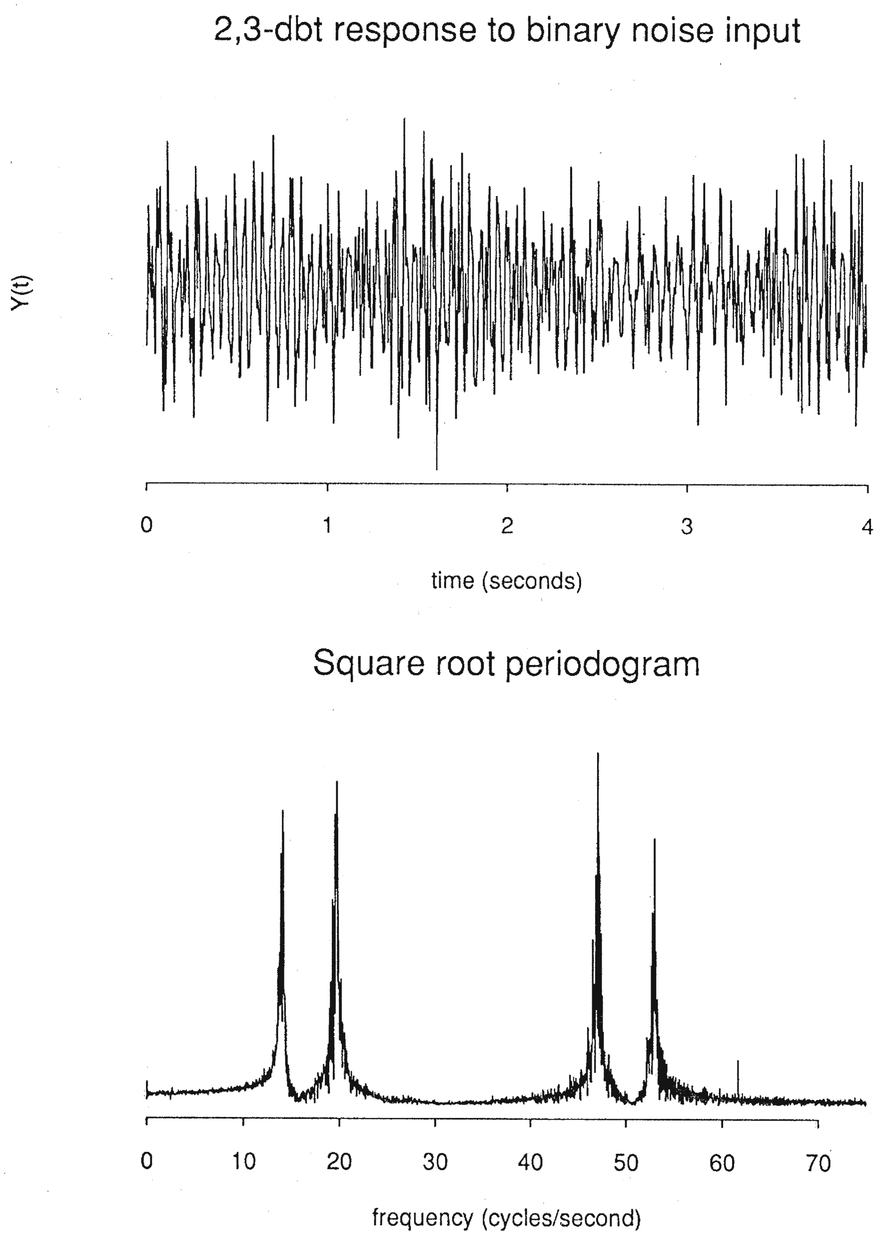


FIGURE 11: Top graph: response of 2,3-DBT to white-noise input. Bottom graph: absolute value of the Fourier transform of the response.

with $\hat{f}_{XX}(\cdot)$ a similar smoothed periodogram of $X(\cdot)$. This figure is much nearer to the ideal of Figure 10.

In fact there exists substantial theory concerning the NMR phenomenon. In particular it may be described by the Bloch equations, which take the form

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{a} + \mathbf{A}\mathbf{S}(t) + \mathbf{B}\mathbf{S}(t)X(t), \tag{12}$$

$$Y(t) = \mathrm{Re}\{\mathbf{c}^{\mathrm{T}}\mathbf{S}(t)\} \tag{13}$$

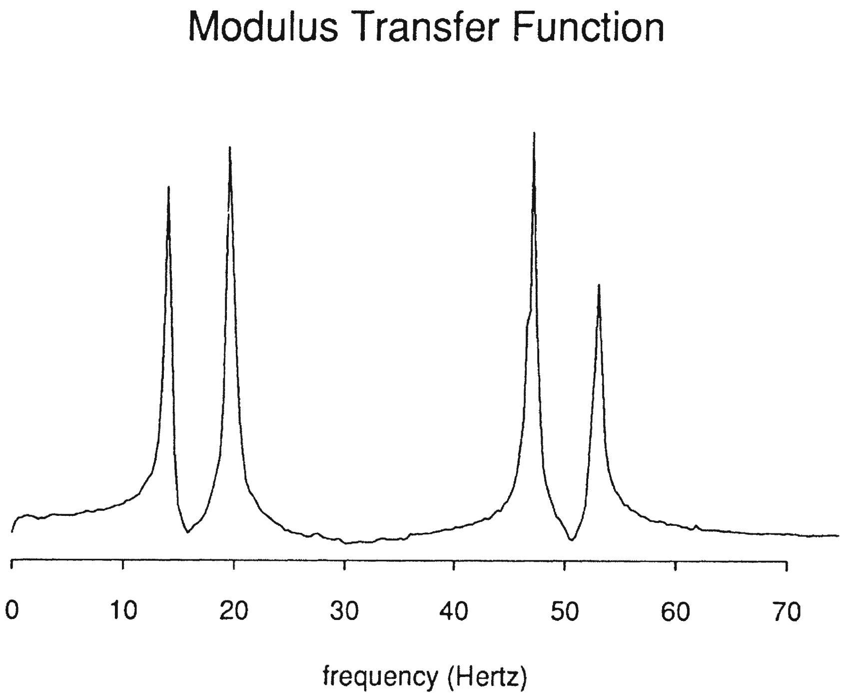


FIGURE 12: Graph of the statistic (11).

for $\mathbf{S}(\cdot)$ a 16-dimensional state vector with complex entries, and $X(\cdot)$, $Y(\cdot)$ corresponding input and output of the system. The parameters of interest are entries of $\mathbf{A}$, $\mathbf{B}$. The vector $\mathbf{c}$ is given by the experimental setup. The parametrization of $\mathbf{A}$, $\mathbf{B}$ is provided by the chemist.

The estimation problem may now be approached by nonlinear regression. For given $\mathbf{A}$, $\mathbf{B}$ and initial state $\mathbf{S}(0)$, the solution of (12) may be determined numerically in the case that the input $X(\cdot)$ is piecewise constant, as it is here. Then the parameters may be estimated by putting the data $Y(\cdot)$, $t = 0, \ldots, T-1$, up against the value determined from (13) and the numerical solution. In this way one obtains estimates and associated standard errors. Further details on this work may be found in Brillinger and Kaiser (1992).

3.4. Discussion.

These examples each involve a progression from a naive (spectral) analysis to a likelihood analysis founded on substantive subject matter. This last leads to efficiency, uncertainty estimation, and the ability to make general statistical inferences.

Because of the complication of the basic circumstances, it seems necessary to collaborate with substantive scientists on these problems. In these cases it was not the data analysis that suggested the model, rather it came from the scientific background.

4. FUTURE PROSPECTS

So what is ahead for spectrum analysis in the pot of gold at the end of the rainbow? Spectrum analysis has already found many uses, and there are quite a number of questions it can address. John Tukey, for example, has emphasized its role in the discovery of phenomena; see Tukey (1966, 1980). Tukey (1980) contains other speculations on the future of time series. Brillinger (1987) mentions a variety of specific problems related to Fourier inference. They are repeated below and some additional ones added.

Turning to future possibilities, there are near-immediate generalizations of the vast majority of time-series techniques to other types of "function" data, such as images, point processes, moving surfaces, and tesselations.

One can speculate on various aspects of contemporary and future research on the spectrum analysis of time series. Many times it has been said that statistical spectrum analysis is part art and part science. The need to be an artist is likely to diminish as automatic bandwidth selection algorithms are developed, analagous to those coming into use in density and nonparametric regression estimation. For making inferences it is necessary to have some indication of the sampling variability of the statistics computed. Various jackknife, perturbation and bootstrap techniques are under development. It will be some time before that research is finished, for procedures will be required for infinite-dimensional parameters and nonstationary series. In connection with nonstationarity, it may be remarked that effective procedures for estimating spectra from short stretches of data are needed, because many series appear to be at most locally stationary. Perhaps those estimates will come from semiparametric approaches involing both finite- and infinite-dimensional parameters. An examination of the contemporary literature shows much work being carried out on non-Gaussian series, on long memory series, on nonlinear models, and on self-similar processes. Finally, new scientific devices, such as lasers and nonlinear crystals, are leading to high-quality data sets with unusual inference problems.

Some particular research problems related to the topics of the paper are:

1. Diagnostics, influence, robust/resistant procedures.
2. Missing values, quantization, jitter.
3. Estimation of dimension, e.g., by AIC.
4. Inverse-problem formulations, e.g., ridge regression.
5. Local asymptotic normality, contiguity.
6. Adaptive procedures.
7. The absorption model.
8. Signal-dependent noise.
9. Law of the iterated logarithm, large deviations, rates of convergence for the estimates.
10. Random-effects models.
11. Vector-valued cases.
12. Partially parametric formulations, e.g., the periodic case.
13. Models for point-process and telegraph-signal cases.
14. Expansions for distributions.
15. Distributions of test statistics, e.g., of

$$\sup_{\rho,\lambda} \frac{\left|\sum_{t=0}^{T-1} \rho^t e^{-i\lambda t} Y(t)\right|}{\sum_{t=0}^{T-1} \rho^{2t}}$$

or of

$$\sup_{\lambda,\mu} \min \{I^T(\lambda),\ I^T(\mu),\ I^T(\lambda+\mu)\}.$$

16. Properties of the estimates when the model is untrue.
17. The broadband-signal case.
18. Parametric analysis of the frequency case.
19. Approximate distribution of the likelihood-ratio test statistic for the presence of a plane wave crossing an array.
20. Sampling properties of the NMR estimates.
21. Techniques for handling small amounts of nonstationary data and massive amounts of regular data.
22. Uncertainty evaluation for the non-i.i.d. case.
23. Errors in variables.
24. Irregularly observed values.
25. Parametric models for nonstandard processes.
26. Asymptotics when the parameter dimension tends to infinity.
27. Characterization of covariance and spectrum functions for 0-1-valued series.
28. Characterization of the spectrum of a stationary point process.
29. Properties of i.i.d.-motivated techniques in dependent cases.
30. Aliasing regions for higher-order spectra of stationary spatial processes.
31. Simulation of processes with given higher-order spectra.
32. Combination of groups of experiments.
33. Populations that are mixtures.
34. Design, e.g., of NMR input.
35. Study of causal networks of general processes.
36. Less-biased estimates of coherence.
37. Detection of long lags between series.
38. Properties of empirical Fourier transforms of unusual processes.
39. Stable algorithms for fitting ARMAs.
40. Employment of contemporary optimization methods.
41. Special models for abstract-valued processes.
42. Automatic determination of bandwidth parameters.
43. Modelling of wildly nonstationary values, e.g., TV signals or computer tasks.
44. Analysis of qualitative-valued processes.
45. How to define "trend" and "period".
46. Fast computation of Fourier transforms for irregularly placed data.
47. Assessing goodness of fit for process models.
48. Measuring association of abstract-valued processes.
49. Strong approximations for statistics based on process values.
50. Detecting change in, or the presence of, a signal.
51. Incorporation of symmetries and invariance.
52. Analysis of censored data.
53. Improved estimates.
54. Relating processes observed at different locations.
55. Study of multivalued random functions.
56. Statistical aspects of determining $g(\cdot)$ from $V(\cdot)$ in (1), including the case when $V(0)$ is unestimated.

5. CONCLUSIONS

Statistics has long had an intimate connection with the physical sciences. Some statistical concepts, like the spectrum, have been motivated by physical considerations; others have arisen abstractly and gone on to influence practice. (The cross-spectrum and bispectrum might be mentioned as examples of the latter.) Statistics researchers benefit from looking towards science for suggestions of new concepts and techniques. There is often a further bonus: the scientific circumstance suggests a path towards a solution. Students are often surprised that statistical concepts have direct physical interpretations.

This paper is part historical and part contemporary collaborative research material. One learns from history. In the case of the present topic one sees the important ideas of spectrum, direct spectrum estimate, and FFT missed for a number of years. [For a discussion of the FFT case see Heideman *et al.* (1984).] It is important to know the past; otherwise gems are missed. Further, history can be a stimulating pedagogic tool for introducing the important concepts and ideas, and often the independent reader will find the original writers much clearer than the later ones.

The applications had in common that a spectral analysis started the investigation, then substantive subject matter was employed to develop a conceptual model and a likelihood analysis carried out to complete the work.

ACKNOWLEDGEMENTS

The author is grateful to Bruce Bolt for various discussions on the material concerning free oscillations and surface waves. For similar help with the material on nuclear magnetic resonance, he is grateful to Reinhold Kaiser. Bob Shumway helped the author in obtaining an old reference, Alan Izenman helped find an important picture of a Spectrum, and Stu McCormick provided some pertinent material.

REFERENCES

Aki, K., and Richards, P.G. (1980). *Quantitative Seismology. Volumes I and II.* Freeman, San Francisco.

Bartlett, M.S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, 37, 1–16.

Bath, M. (1974). *Spectral Analysis in Geophysics*. Elsevier, Amsterdam.

Becker, E.D., and Farrar, T.C. (1972). Fourier transform spectroscopy. *Science*, 178, 361–368.

Beveridge, W.H. (1922). Wheat prices and rainfall in western Europe. *J. Royal Statist. Soc.*, 85, 412–459.

Bolt, B.A., and Butcher, J. (1960). Rayleigh wave dispersion for a single layer on an elastic half space. *Australian J. Physics*, 13, 498–504.

Bolt, B.A., and Brillinger, D.R. (1979). Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series. *Geophys. J. Roy. Astronom. Soc.*, 59, 593–603.

Bolt, B.A., and Marusi, A. (1962). Eigenvibrations of the earth observed at Trieste. *Geophys. J. Roy. Astronom. Soc.*, 6, 299–311.

Brillinger, D.R. (1987). Fitting cosines: Some procedures and some physical examples. *Applied Probability, Stochastic Processes and Sampling Theory* (I.B. MacNeill and G.J. Umphrey, *eds.*), Reidel, Boston, 75–100.

Brillinger, D.R. (1993). An application of statistics to seismology: dispersion and modes. Developments in Time Series Analysis. Chapman and Hall, London.

Brillinger, D.R., and Kaiser, R. (1992). Fourier and likelihood analysis in NMR spectroscopy. *New Directions in Time Series Analysis, Part I* (D.R. Brillinger, E. Parzen, and M. Rosenblatt, *eds.*), Springer, New York, 41–64.

Brownlee, J. (1917). An investigation into the periodicity of measles epidemics in London from 1703 to the present day by the method of the periodogram. *Philos. Trans. Roy. Soc. London Ser. B*, 208, 225–250.

Daniell, P.J. (1946). Discussion of paper by M.S. Bartlett. *J. Roy. Statist. Soc. Suppl.*, 8, 88–90.

Dziewonski, A.M., and Hales, A.L. (1972). Numerical analysis of dispersed seismic waves. *Seismology: Surface Waves and Earth Oscillations* (B.A. Bolt, *ed.*), Academic Press, New York, 39–85.

Einstein, A. (1914). Méthode pour la détermination des valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Arch. Sci. Phys. et Naturelles Ser. 4*, 37, 254–255.

Ernst, R.R. (1970). Magnetic resonance with stochastic excitation. *J. Magnetic Resonance*, 3, 10–25.

Ernst, R.R., and Anderson, W.A. (1966). Application of Fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instrum.*, 37, 93–102.

Fisher, R.A. (1929). Tests of significance in harmonic analysis. *Proc. Royal Soc. London Ser. A*, 125, 54–59.

Gouy, G.L. (1886). Sur le mouvement lumineux. *J. Physique*, 5, 354–362.

Heideman, M.T., Johnson, D.H., and Burrus, C.S. (1984). Gauss and the history of the fast Fourier transform. *IEEE Acoust. Speech Signal Process Magazine.*, 1, 14–21.

Hochstadt, H. (1975). *Differential Equations*. Dover, New York.

Kaiser, R. (1970). Coherent spectrometry with noise signals. *J. Magnetic Resonance*, 3, 28–34.

Kimura, H. (1913). On the harmonic analysis of sun-spot relative numbers. *Monthly Notices Roy. Astronom. Soc.*, 73, 543–548.

Masani, P.R. (1986). Einstein's contribution to generalized harmonic analysis and his intellectual kinship with Norbert Wiener. *Jahrb. Überblicke Math.*, 191–209.

Michelson, A.A. (1891). On the application of interference methods to spectroscopic measurements — I. *Philos. Mag.*, 33, 338–346.

Michelson, A.A. (1892). On the application of interference models to spectroscopic measurements — II. *Philos. Mag.*, 34, 280–299.

Michelson, A.A. (1902). *Light Waves and Their Uses*. Univ. Of Chicago Press, Chicago.

Michelson, A.A. (1913). Determination of periodicities by the harmonic analyzer with an application to the sun-spot cycle. *Astrophys. J.*, 38, 268–274.

Michelson, A.A., and Stratton, S.W. (1898). A new harmonic analyzer. *Amer. J. Sci.*, 5, 1–13.

Press, F. (1965). Resonant vibrations of the earth. *Sci. Amer.*, 213, 28–38.

Robinson, E.A. (1982). A historical perspective of spectrum estimation. *Proc. IEEE*, 70, 885–907.

Rayleigh, L. (1892). On the interference bands of approximately homogeneous light. *Philos. Mag.*, 34, 407–411.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Mag.*, 3, 13–41.

Sobel, M.I. (1989). *Light*. Univ. of Chicago Press, Chicago.

Topper, D. (1990). Newton on the number of colours in the spectrum. *Stud. Hist. Philo. Sci.*, 21, 269–279.

Tukey, J.W. (1950). The sampling theory of power spectrum estimates. *Symposium on Applications of Autocorrelation Analysis to Physical Problems*, Office of Naval Research, Washington, 47–67. *The Collected Works of John W. Tukey. Volume 1: Time Series, 1949–1964* (D.R. Brillinger, *ed.*), Wadsworth, Belmont, Calif., 129–160.

Tukey, J.W. (1966). Uses of numerical spectrum analysis in geophysics. *Bull. Internat. Statist. Inst.*, 41, 267–307.

Tukey, J.W. (1980). Can we predict where "time series" should go next? *Directions in Times Series* (D.R. Brillinger and G.C. Tiao, *eds.*), IMS, Hayward, Calif., 1–31.

Tukey, J.W., and Hamming, R.W. (1949). Measuring noise color. Bell Telephone Laboratories Technical Memorandom. *The Collected Works of John W. Tukey, Volume 1: Time Series, 1949–1964* (D.R. Brillinger, *ed.*), Wadsworth, Belmont, Calif., 1–127.

Wiener, N. (1930). Generalized harmonic analysis. *Acta Math.*, 55, 117–258.

Whittle, P. (1952). The simultaneous estimation of a time series harmonic components and covariance structure. *Trab. Estad.*, 3, 43–57.

Yaglom, A.M. (1987a). Einstein's 1914 paper on the theory of irregularly fluctuating series of observations. *IEEE Acoust. Speech Signal Process Magazine*, 4, 7–11.

Yaglom, A.M. (1987b). *Correlation Theory of Stationary and Related Random Functions. Volumes I, II.* Springer-Verlag, New York.

Received 21 May 1992
Revised 17 July 1992
Accepted 21 August 1992

Department of Statistics
University of California
Berkeley, California
U.S.A. 94720

ELSEVIER

Signal Processing 65 (1998) 161–179

SIGNAL PROCESSING

An investigation of the second- and higher-order spectra of music

D.R. Brillinger*, R.A. Irizarry

Statistics Department, University of California, Berkeley, CA 94720, USA

Received 15 July 1996; accepted 15 May 1997

Abstract

For a variety of musical pieces the following questions are addressed: Are the power spectra of $1/f$ form? Are the processes Gaussian? Are the higher-order spectra of $1/f$ form? Are the processes linear? Is long-range dependence present? Both score and acoustical signal representations of music are discussed and considered. Parametric forms are fit to sample spectra. Approximate distributions of the quantities computed are basic to drawing inferences. In summary, $1/f$ seems to be a reasonable approximation to the overall spectra of a number of pieces selected to be representative of a broad population. The checks for Gaussianity, really for bispectrum 0, in each case reject that hypothesis. The checks for linearity, really for constant bicoherence, reject that hypothesis in the case of the instantaneous power of the acoustical signal but not for the zero crossings of the signal or the score representation. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Für eine Anzahl von Musikstücken werden die folgenden Fragestellungen behandelt: Gehorchen die Leistungsdichtespektren einem $1/f$-Gesetz? Sind die Prozesse gaußverteilt? Besitzen die Spektren höherer Ordnung $1/f$-Form? Sind die Prozesse linear? Sind langfristige Abhängigkeiten vorhanden? Sowohl Partituren als auch akustische Signaldarstellungen von Musik werden betrachtet und diskutiert. Parametrische Darstellungen werden an Spektren von Musterfunktionen angepaßt. Näherungsweise Verteilungen der berechneten Größen sind grundlegend für statistische Rückschlüsse. Zusammenfassend scheint eine $1/f$-Form eine sinnvolle Näherung für die Spektren einer Anzahl von Stücken zu sein, die als repräsentativ für einen großen Bestand ausgewählt wurden. Die Tests bezüglich Gaußverteilung (eigentlich bezüglich verschwindendem Bispektrum) weisen eine solche Hypothese immer zurück. Die Tests bezüglich Linearität (eigentlich bezüglich konstanter Bikohärenz) weisen diese Hypothese im Fall der Momentanleistung des akustischen Signals zurück, nicht aber für die Nulldurchgänge des Signals oder für die Partitur. © 1998 Elsevier Science B.V. All rights reserved.

Résumé

Les questions suivantes sont posées pour une variété de morceaux de musique: Les spectres de puissance sont-ils de la forme $1/f$? Les processus sont-ils gaussiens? Les spectres d'ordre supérieur sont-ils de la forme $1/f$? Les processus sont-ils linéaires? La dépendance à long terme est-elle présente? Les représentations de musique sous forme de partition et de signal acoustique sont toutes deux discutées et examinées. Des formes paramétriques sont ajustées aux spectres

* Corresponding author. E-mail: brill@ stat.berkeley.edu.

0165-1684/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0165-1684(97)00217-X

expérimentaux. Les distributions approximatives des quantités calculées sont essentielles pour tirer des conclusions. En résumé, la forme en $1/f$ semble être une approximation raisonnable des spectres globaux d'un certain nombre de morceaux sélectionnés comme représentatifs d'une population étendue. Les vérifications de gaussianité, en vérité de bispectre nul, rejettent dans chaque cas cette hypothèse. Les vérifications de linéarité, en vérité de bicohérence constante, rejettent cette hypothèse dans le cas de la puissance instantanée du signal acoustique mais pas pour passages par zéro du signal ou de la représentation par partition.

Keywords: Bicoherence; Bispectrum; Linear process; Music; Parametric model; Spectral analysis

1. Introduction

What is music? Probably nobody will ever give a final answer to this question, but something inside of us tells us when a sound we hear is music and when it is not. Most people hear the sound of cars passing by on a road and do not think it is music, but it only takes them a moment after a radio is turned on to identify the sound coming out as music. Certain sounds we classify as music others we do not. In this paper we examine some statistical properties of two different numerical representations of music to see if we can shine some light on the property that makes music, music.

We are able to process music in a data analytic fashion because the time is at hand when music can be treated directly as data to be analyzed by contemporary statistical procedures and packages.

The paper begins with a description of the two basic representations of music, moves on to some review of previous investigations, then presents the results of modeling the second-order spectra and finally employs higher-order spectra to assess Gaussianity and linearity.

The pieces investigated included: Baroque, Classical, Romantic, Atonal, Spanish Guitar, Jazz, Latin, Rock & Roll and Hip Hop.

2. Representations of music

Certainly music is sound. Every sound we hear is the consequence of pressure fluctuations traveling through the air and hitting our ear drums. The signal representation takes this property of sound to represent music as a continuous function.

For years composers have transcribed the music they hear in their heads using what is known as common practice notation (CPN). We use such 'numerical' representations of music for our analyses.

2.1. Signal representation

The function describing the audible pressure fluctuations of air is called a 'sound wave'. The energy transmitted by this 'sound wave' can be transformed into a voltage $Y(t)$, which will be a continuous function in time. Compact Disks are proof of how effective quantized samples of this function are. This time series $Y(t)$, $0 < t < T$, will be called the *signal representation* of music. Throughout this paper we will be using a discrete version of the function, Y_t, $t = 0,1,\ldots$

When such fluctuations of air are approximately periodic we hear a sound with a definite musical pitch. Instruments play different pitches by changing the fundamental frequency of the 'sound wave' they are creating [15]. Some cultures, e.g. Western cultures, have quantized these pitches and created 'notes'. This has permitted composers to write with a notation that an instrumentalist can then turn into sounds. This notation provides the other representation of music, the *score representation.*

2.2. Score representation

Most instruments known to us have the capability to play different 'notes'. In all 'melodic' instruments, for example violins, pianos, trumpets, sitars, etc., as mentioned above different notes correspond to different fundamental frequencies or pitches. The pitch corresponding to 440 Hz has been called A (concert pitch A). Any frequency that holds a 2^n:1 relation with A is also called A, but in another octave. Western music uses the 12 tone equal-tempered scale in which the frequencies between, say 440 Hz (concert pitch A) and 880 Hz (an octave above concert pitch A) have been divided into 12 notes corresponding to frequencies with the same ratio between them. These 12 notes are $A, A\sharp$ (A sharp), $B, C, C\sharp, D, D\sharp, E, F, F\sharp, G, G\sharp$ and that brings us back to A (an octave above). If you look at a piano the black keys correspond to the sharps and you will see a twelve white-black keys pattern repeating 7 times. Adjacent notes are said to be a half-step apart or a semi-tone away, see [15].

The human audible range can hear about 4 octaves below concert pitch A and about 6 octaves above (this is for the keenest of ears). This means that there are about 100 notes that we can hear. Western composers have found a universal way of representing these notes, namely, what is known as common practice notation (CPN). Probably most if not all sheet music you have seen uses this notation. With this notation a composer tells a performer what pitch his instrument should play. Representing notes as numbers is now straightforward. The MIDI standard (see more detail below) assigns to concert pitch A the number 69 and for every adjacent note adds or subtracts one.

To transcribe a melody we also need the rhythm. CPN also provides symbols to denote how long each note is going to be played and also for how long nothing will be played (rests). In Western music the time domain is divided into measures and beats and into sub-beats. For any given song one could find the smallest subdivision of the beat. We will call this a *tatum* (as defined by Bilmes in his master's thesis [1]), such that any distance between any two notes can be represented as k tatums, k an integer. This is usually easy to do by looking at the score. As an example consider a song that has 5 measures. Each measure is divided into 3 beats. Say that a tatum is equivalent to a quarter of a beat. Then each beat is 4 tatums long, each measure is 4×3 tatums and the song is $3 \times 4 \times 5$ tatums long.

Even with this representation we will not have a one-to-one correspondence with sounds. Each note can have millions of different sounds (timbre): loudness, tremolo, staccato, varying with different instruments, who's playing (for some it is not hard to distinguish the timbre of two different players), etc. The same occurs for the rhythm: decrescendos, accelerandos, rubato, swing, etc. Even though a one-to-one correspondence does not exist we can make a good approximation using the MIDI standard.

2.2.1. *The MIDI standard*

The MIDI (Musical Instrument Digital Interface) standard is a hardware specification and communications protocol that allows computers, controllers, and synthesis gear to pass information amongst themselves, see [13]. MIDI uses representations based on the concept of notes by defining a pitch and a velocity (volume) that go on and off. MIDI is mostly controlled by keyboard instruments which can be represented by a series of switches. Each separate key is treated as a switch. When a key is depressed, a *Note On* message is sent out, indicating the note associated with that key and with what velocity it was struck. When the key is released, a *Note Off* message is transmitted with the key number and velocity 0. In a similar way MIDI can be used to go from a score representation of sound to an acoustic signal. The way MIDI, together with sound synthesis techniques, converts scores to music is rather complicated. In the following section we present a method of converting a series of notes represented in a MIDI score to an acoustical signal representation of a sine tone instrument (i.e. an instrument with no harmonics).

2.2.2. *Time series representation*

The time series representation X_j, where j is the tatum number, is defined by $X_j =$ note at tatum j. This representation does not characterize the score exactly, since it makes no distinction between two contiguous identical notes with durations d_1 and d_2 and that note with duration $d_1 + d_2$.

As a numerical representation of a note we could use the MIDI-Note number. In this case an increase of a step would represent a jump to the note a semi-tone away. This presents a problem when dealing with rests. Rests do not have Midi-Note numbers. We could not just assign 0 to rests because then this would be representing a note corresponding to MIDI-number 0. Even though this note is below the audible range it does not correspond to 0 frequency thus its choice is quite arbitrary since notes with MIDI-note numbers smaller than 16 correspond to notes below the audible range and using Eq. (1) below we see that the MIDI-note number corresponding to 0 frequency is $-\infty$. One way to get around this is to prolong the duration of notes preceding rests. In a song with few rests of short duration this would not make much of a difference.

An alternative numerical representation, that is more in accordance with the signal representation, is using the fundamental frequency of the pitch determined by the notes. For example a note of MIDI-number X would be represented by frequency

$$440 \times 2^{(X-69)/12} = 8.175799 \times 2^{X/12}\,\text{Hz}, \tag{1}$$

see [15]. In this case frequencies related to rests could be set to 0 since a sound wave with 0 frequency has no fluctuations and thus is silent. It would be interesting to note how robust our analysis is to this arbitrary assignment.

2.2.3. Marked point process representation

Suppose we have a series of triplets (note, duration, volume), then we can construct an acoustical signal representation via the following definitions:

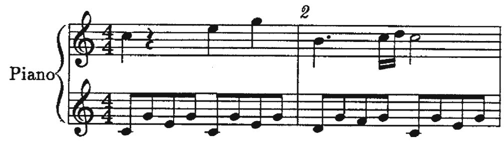
$$Y(t) = \sum_j V_j h\left(\frac{t-\tau_j}{\sigma_j}\right)\cos \lambda_j(t-\tau_j),$$
$$h(\cdot) = \text{a taper function}, \tag{2}$$

where τ_j is the time of commencement of the jth note, λ_j is the frequency of the jth note, V_j is the volume of the jth note and σ_j is the duration of the jth note. Here $\{\tau_j\}$ will be a point process corresponding to times of jumps between notes. For time t near τ_j the signal will look like a cosine wave of frequency λ_j and amplitude V_j. The units t here could be seconds as well as tatums in which case we could represent changes in tempo by using time maps that assign a duration in seconds to each tatum, see [21].

To compute the frequency λ_j from midi-number X_j we use Eq. (1). (We used this conversion method to check for mistakes in the data entry. By converting the entered data and forming the signal produced by relation Eq. (2) we then played the signal through the speakers of a Sparc work-station using the Matlab command `sound`. See Appendix A for some details.)

One reason the taper function is introduced in Eq. (2) is to avoid hearing clicks at instantaneous changes of pitch. It also expresses the restricted duration of a particular note.

2.2.4. An example

The following is the common practice notation (CPN) for the first two bars of Mozart's Sonata in C-major, K545:[1]

The melody in these two bars is played by the right hand (shown in the upper clef). In this case the tatum would correspond to a *sixteenth* note, or a quarter of a beat. If the song were played at an Allegro tempo (about 144 quarter notes per minute) then a tatum would have a duration of 60 (s)/ 144(quarter notes per beat) $\times$ 1/4(tatums per beat) ≈ 0.10 seconds. The note and duration in tatum pairs are the following: (C,4), (rest,4), (E,4), (G,4), (B,6), (C,1), (D,1), (C,8).

The time series representation using MIDI-numbers would be:

72, 72, 72, 72, NA, NA, NA, NA,
76, 76, 76, 76, 79, 79, 79, 79,
71, 71, 71, 71, 71, 71,
72, 74, 72, 72, 72, 72, 72, 72, 72, 72.

Here the NAs represent rests.

The time series representation using frequencies of the pitches would be:

523, 523, 523, 523, 0, 0, 0, 0,
659, 659, 659, 659, 783, 783, 783, 783,
493, 493, 493, 493, 493, 493, 523,
587, 523, 523, 523, 523, 523, 523, 523, 523.

[1] The actual song starts with a half note C and no rest. We put in the rest for illustrative purposes.

A marked point process representation with time measured in tatums characterizes the score. For the sonata we have $\{\tau_j, (V_j, \lambda_j, \sigma_j)\}$: $\{0, (1, 523, 4)\}$, $\{8, (1, 659, 4)\}$, $\{12, (1, 783, 4)\}$, $\{16, (1, 493, 6)\}$, $\{22, (1, 523, 1)\}$, $\{23, (1, 587, 1)\}$, $\{24, (1, 523, 8)\}$. (We have set the volume to 1, this choice is completely arbitrary. This part of the score does not ask for certain notes to be played louder than others. In practice accents are always present.)

3. Some previous work

Electronic musicians have used random processes to create melodies. Completely uncorrelated processes, with constant spectra, seem to create 'melodies' with no structure. 'Melodies' produced with random walks, i.e. spectrum $1/f^2$, seem to be too predictable. In between these two processes is so called $1/f$ noise.

Voss studied the possibility of music having a $1/f$ spectrum [22–24]. He took the signal representations $Y(t)$ of a variety of songs and obtained the 'instantaneous' audio power of music. In order to measure it, the audio signal $Y(t)$ was passed through a bandpass filter in the range 100 Hz to 10 kHz. The output voltage was squared, and filtered with a 20 Hz low-pass filter. Voss remarked that correlations of the resulting process represented correlations of the audio power of successive notes. For a discussion of some properties of this filtering technique see Appendix A.

Another quantity Voss examined was the 'instantaneous' frequency. He measured this by the rate, $Z(t)$, of zero crossings of the audio signal. He remarked that in the case of music, correlations of $Z(t)$ represented correlations in the frequencies of successive notes. This is reasonable because if say frequency λ dominates at time t, the signal will be approximately $\rho\cos(\lambda t + \phi)$ and the rate of zero crossings (or cycles) is $\lambda/2\pi$ per unit time. Of course problems may arise when more than one stream of notes is played at the same time, for example in Mozart's Sonata above you have the right hand playing a stream of notes corresponding to the melody and the left hand playing a stream of notes corresponding to the accompaniment.

These two methods, of seeking information of the melody from the audio signal, work well when the melody is being played by one instrument with no harmonics, as we see in Fig. 1. In a case where the audio signal contains more than one instrument and the sound produced by these instruments contains many harmonics, these methods do not work as well. In Fig. 1 we see the 4 time series plots. The first is the score representation, using frequencies, of the first 10 measures of the melody line of Mozart's *Eine Kleine Nachtmusik*, the second is the smoothed zero crossings of the signal created using Eq. (2) on the score representation, the third is the smoothed zero crossing of the audio signal of an actual orchestra playing the song and finally the fourth plot is the 'instantaneous' power of the audio signal. Notice how well the zero crossings method works when the sound signal contains only one instrument with no harmonics. In the third plot we see that the method does not work well when there is more than one instrument playing. Notice also that at the beginning of the song, when all the instruments are playing the same notes (first 4 measures), the method works better than when there is more than one stream present. (See Appendix A for the procedure used to obtain these figures).

In another formal study of music Hsu and Hsu [9] study the fractal nature of the intervals between successive notes. This corresponds to the intervals in the score representation using the MIDI-note numbers. If in Eq. (2) we used the MIDI-note numbers M_j instead of the frequencies λ_j, then these intervals would be defined by $I_j = M_{j+1} - M_j$ for $j = 1, \ldots, N-1$, where N is the number of notes in the whole piece.

4. Second-order spectra

The second-order spectrum or power spectrum of a stationary process $Y(t)$, $-\infty < t < \infty$, is given by

$$\operatorname{cov}\{Y(t+u), Y(t)\} = \int_{-\infty}^{\infty} \cos(\lambda u) f_2(\lambda)\,\mathrm{d}\lambda, \tag{3}$$

with u the lag. The physical meaning of the spectrum is that $f_2(\lambda)\,\mathrm{d}\lambda$ represents the contribution to

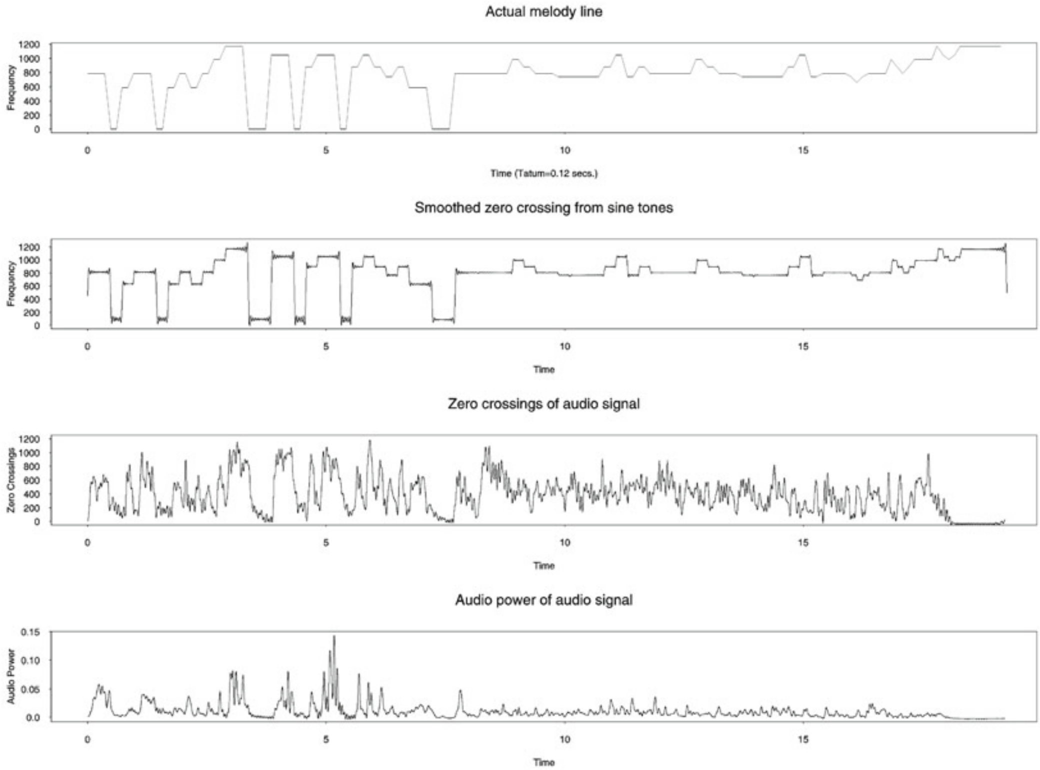


Fig. 1. Score, smoothed zero crossings and instantaneous audio power of Eine Kleine Nachtmusik.

the variance or power of $Y(t)$ components with frequencies in the ranges $(\lambda, \lambda + \mathrm{d}\lambda)$ and $(-\lambda, -\lambda + \mathrm{d}\lambda)$.

These definitions extend directly to the case of a locally stationary process. Crudely the (overall) spectrum of the process Eq. (2) will be proportional to

$$\sum_j V_j^2 \sigma_j \delta(\lambda - \lambda_j), \tag{4}$$

where $\delta(\cdot)$ is the Dirac delta function and the process will have $1/f$ spectrum to the extent that $V_j^2 \sigma_j$ falls of as $1/\lambda_j$. The process itself will be locally stationary with instantaneous frequency λ_j for t near τ_j.

4.1. Estimates

In the case that the stationary process Y_t has mean 0 a naive estimate of the spectrum is provided by the periodogram,

$$I_2^T(\lambda) = \frac{1}{2\pi T} |d^T(\lambda)|^2, \tag{5}$$

where

$$d^T(\lambda) = \sum_{t=1}^{T} \exp\{-\mathrm{i}\lambda t\} Y_t. \tag{6}$$

The periodogram is an asymptotically unbiased but inconsistent estimate (unless $f_2(\lambda) = 0$) since $\mathrm{Var}[I_2^T(\lambda)] \approx f_2(\lambda)^2$ as $T \to \infty$.

If the series Y_t is mixing (see e.g. conditions in [3]), the variates

$$I_2^T(\lambda_t)/f_2(\lambda_t), \quad \lambda_t = 2\pi t/T \text{ for } t = 1, 2, \ldots \qquad (7)$$

are approximately independent exponentials with mean 1.

4.2. Parametric modeling

Voss proposed that the spectrum of music has a $1/f$ (or $1/\lambda$) parametric form. Consider the problem of fitting parametric models to spectra. We can find estimates by maximizing the approximate log likelihood

$$L_T(\theta) = -\sum_{t=1}^{T-1}\left[\log(f_2(\lambda_t;\theta)) + \frac{I_2^T(\lambda_t)}{f_2(\lambda_t;\theta)}\right],$$

$$\lambda_t = \frac{2\pi t}{T}, \qquad (8)$$

see [6]. With θ estimated by $\hat{\theta} = \arg\max_\theta L_T(\theta)$, under certain conditions (including that the trispectrum is 0), $\hat{\theta}$ is consistent and asymptotically normal

$$\sqrt{T}(\hat{\theta}_T - \theta) \to N(0, \Gamma_\theta^{-1}) \qquad (9)$$

as $T \to \infty$, where

$$\Gamma_\theta[k,l] = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k}\log f_2(\lambda;\theta)\frac{\partial}{\partial\theta_l}\log f_2(\lambda;\theta)\,\mathrm{d}\lambda. \qquad (10)$$

The estimate is asymptotically efficient in the Gaussian case.

The goodness of fit of a particular parametric model may be assessed by graphing the estimate, $I_2^T(\lambda)$, as well as the parametric estimate $f_2(\lambda, \hat{\theta})$ surrounded by confidence bounds for the former. This will be done in the examples that follow.

4.2.1. Models for spectra

We consider the following models for the overall power spectrum of music:

1. $f_2(\lambda; \alpha, \beta) = \alpha/\lambda^\beta$,
2. $f_2(\lambda; \alpha, \beta) = \alpha/(1 + \lambda^\beta)$,
3. $f_2(\lambda; a, b, c, \alpha) = \lambda^a/(1 + \alpha\lambda^b)^c$,
4. $f_2(\lambda; a, b, \ldots, \alpha, \beta, \ldots)$
   $= (\lambda^a/(1 + \alpha\lambda^b)^c)(1/(1 + \beta\lambda^d)^e)$.

By choice of functional form and parameter values, these models are able to describe a fairly broad range of behavior.

In all the models $0 \leqslant \lambda \leqslant \pi$ with f_2 symmetric and of period 2π. Notice that the first model is the '$1/f$' noise model.

4.3. Signal representation

Songs from a variety of musical styles were chosen in our study of the power spectrum of the signal representation. These songs included:

1. Baroque: J.S. Bach, Cantata No. 211 (Coffee Cantata) BWV 211, *Recitativo*: *Wenn Du mir nicht den Coffee* and Cantata burlesque (Peasant Cantata) BWV 212, *Aria*: *Heute noch, lieber Vater, tut es doch.* Performed by baritone Kevin McMillan, soprano Dorothea Röschmann and *Le Violins du Roy* chamber orchestra.
2. Classical: J.F. Haydn, Sonata in D-mayor Hob. XVI/37, *Finale* and Sonata in F-mayor Hob. XVI/23, *Finale*. Both performed on Piano by Dominique Cornil.
3. Romantic: C. Debussy, Suite bergamasque L. 75, *Passepeid* and Images L. 87, *Lent*. Both performed on Piano by Zóltan Kocsis.
4. Atonal: A. Schoenberg, Orchesterstücke op. 16, *Vorgefühle* and Orchesterstücke op. 16, *Peripetie*. Both performed by Berlin Philarmonic.
5. Spanish Guitar: L. Milán, *Pavan No.* 6 and *Pavan No.* 5. Both performed on Guitar by Andrés Segovia.
6. Jazz: Wayne Shorter, *Footprints* and Miles Davis, *Four*. *Footprints* performed by Miles Davis. *Four* performed by Sonny Rollins. In both cases we recorded just the head (In most Jazz tunes a song starts off with a fixed melody, called the head, and then improvisations are played).

7. Afro-Cuban: Juan Mesa, *Amalia* and Florencia Calle, *Baba Cuello Mao*. Both performed by *Los Muñequitos de Matanza.*
8. Rock and Roll: Chuck Berry, *Let it Rock* and Chuck Berry, *Bye Bye Johnny*. Both performed by Chuck Berry.
9. Hip-Hop/Rap: R. Stewart, E. Wilcox, R. Jackson, T. Hardson, R. Robinson and J. Martínez, *It's Jiggaboo Time* and *If I Were President*. Both Performed by *The Pharcyde.*

We sampled the audio signal of the mentioned songs at 8000 samples per second. The sampled signal was then filtered using the two methods of Voss described above. It is important to note that the units of the signal Y_t are arbitrary. (See Appendix A for the details of these computations.) First we determined Y_t to be the smoothed zero crossings of the signal. Then we calculated the periodogram of Y_t and minimized the negative of the approximate log likelihood given in Eq. (8) restricting λ to $(0,20)$ Hz since frequencies over 20 Hz where filtered out. Using Powell's algorithm, see [16], the four models were fitted. The results of these fits in the case of Bach's *Coffee Cantata* can be seen in Fig. 2. The goodness of fit may be assessed by the approximate 95% confidence intervals which are given as the dashed lines. Model 1 seems to fit well here. The same was done for the 'instantaneous' audio power of the signals, the four models were fitted. The results of the fits for the *Coffee Cantata* can be seen in Fig. 3. Again Model 1 seems to fit well.

We fitted the $1/f$ model for the smoothed zero crossings obtained from the signal representations of the 18 pieces listed above. A fit for each style can be seen in Fig. 4. Approximate standard errors are calculated using Eq. (10). The $1/f$ model appears to be performing well.

The values obtained for $\hat{\beta}$ are given in Table 1. The fraction of points outside the (approximate) 95% intervals ranges from 4.6% to 6.3%.

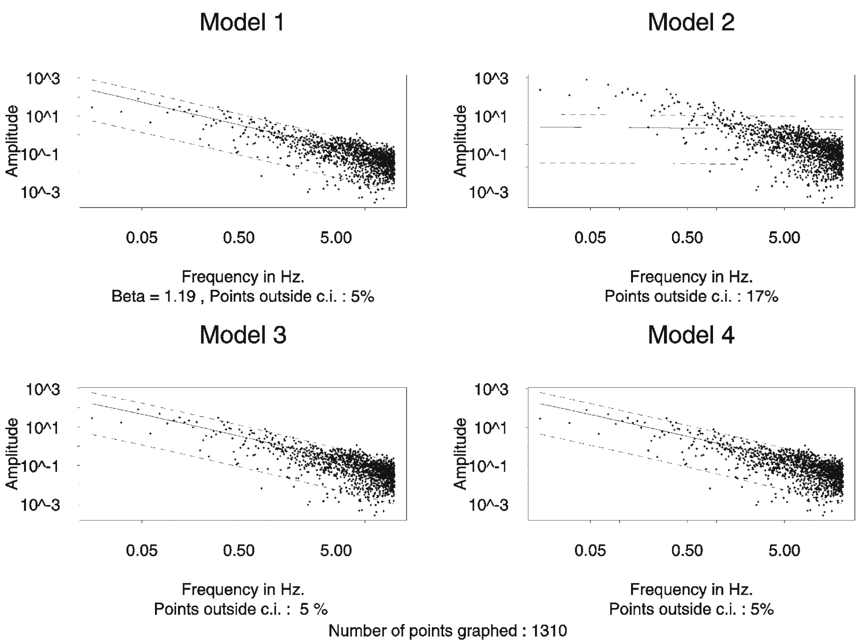


Fig. 2. Fitted models for the smoothed zero crossings obtained from Bach's *Coffee Cantata* signal.

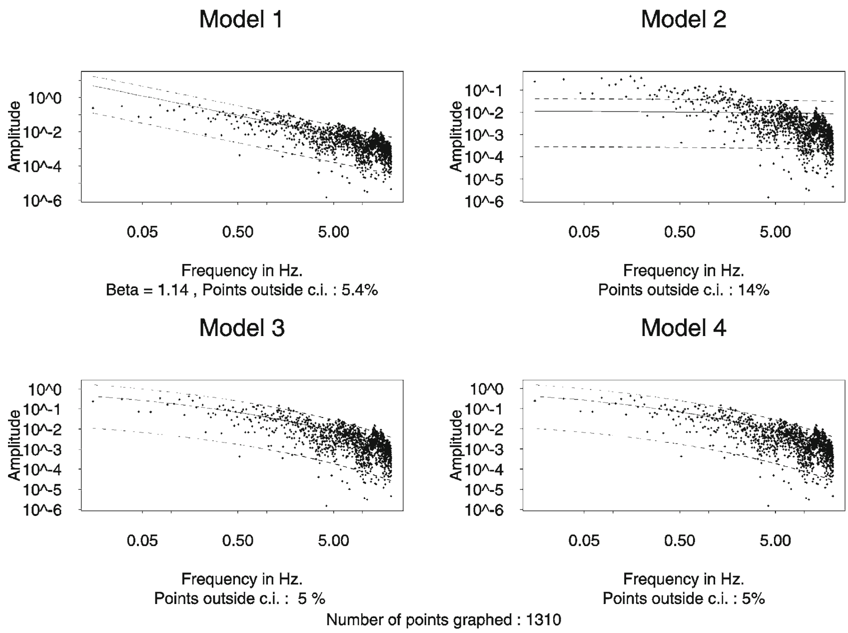


Fig. 3. Fitted models for the 'instantaneous' power obtained for Bach's *Coffee Cantata* signal.

4.4. Score representation

We next looked at the time series expressions of various songs representative of several styles of music. In the following list we give the composer, title of composition, title of the specific part (when applicable), the key, time signature and tempo given in the score, and what type of note corresponds to a tatum.

1. Baroque
   - 1.1. J.S. Bach Cantata No 211 (Coffee Cantata), Be Silent All, Recitativo: Wenn du mir nicht den Coffee, D-major, 4/4, Tempo = 70, Tatum = 1/16.
   - 1.2. J.S. Bach, French Suites, Suite II, Courante, C-minor, 3/4, Tempo = 144, Tatum = 1/8.
2. Classical
   - 2.1. F.J. Haydn, La Roxelane: Air and Variations, Theme, C-minor, 2/4, Tempo = 150 (Allegretto), Tatum = 1/16.
   - 2.2. F.J. Haydn, La Roxelane: Air and Variations, Var I, C-major, 2/4, Tempo = 150 (Allegretto), Tatum = 1/16.
   - 2.3. F.J. Haydn, La Roxelane: Air and Variations, Var II, C-minor, 2/4, Tempo = 150 (Allegretto), Tatum = 1/16.
3. Romantic
   - 3.1. Claude Debussy, Suite Bergamasque, Passepeid, F-minor, 4/4, Tempo = 150 (Allegretto ma non troppo), Tatum = 1/8, (Note: this is an approximation to the melody. Triplets were ignored and replaced by the first note.)
4. Spanish Guitar
   - 4.1. Luis de Milán, Pavan no. 5, in Tone VIII, "La bella Francesca" (Fol. [G vi^{r}]), G-minor, Complex meter varies between 2/4 3/4, Tempo = 120 (Allegro Moderato), Tatum = 1/16.
   - 4.2. Luis de Milán, Pavan no. 6, in Tone VIII, (Fol. [G vi^{r}]), G-minor, Complex meter

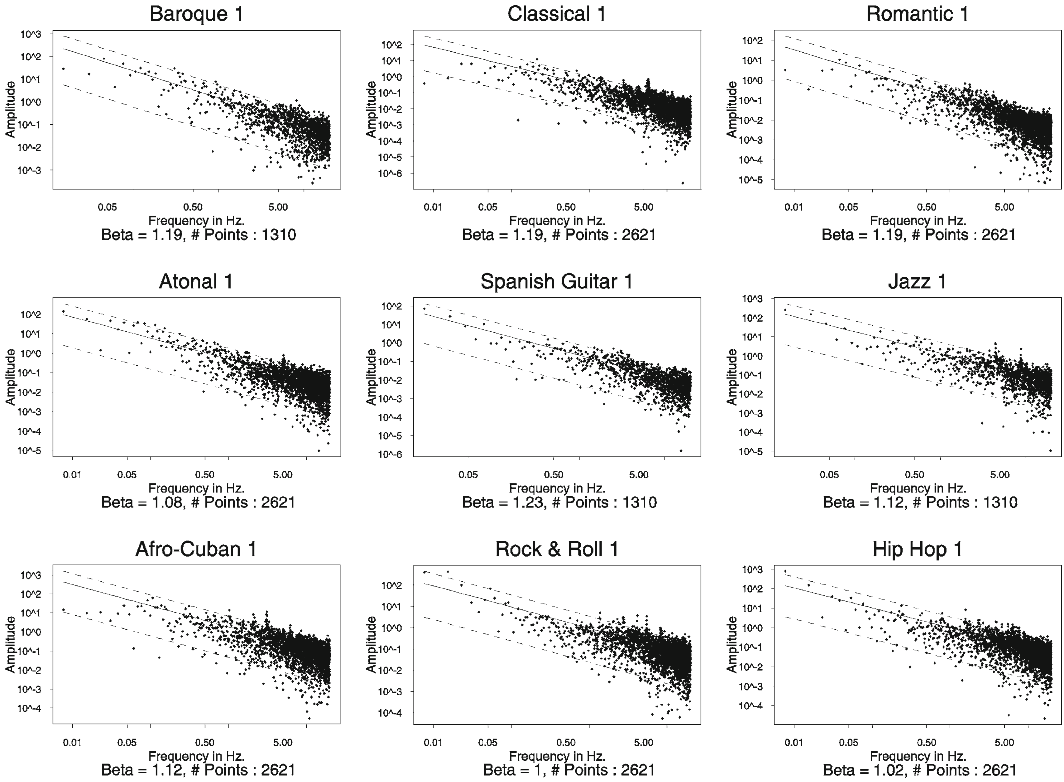


Fig. 4. Fitted $1/f$ model for the smoothed zero crossings of 9 style types.

varies between 2/4 3/4, Tempo = 120 (Allegro Moderato), Tatum = 1/8.

5. Jazz
   5.1. Miles Davis, Four, E♭-major, 4/4, Tempo = 178 (Medium Swing), Tatum = 1/8.
   5.2. Wayne Shorter, Footprints, E♭-major, 6/4, Tempo = 178 (Medium Swing), Tatum = 1/8. In both cases we use the score of the head.
6. Latin
   6.1. Pérez Prado, Mambo No. 5, E♭-major, 2/2, tempo = 240, tatum = 1/8.
   6.2. Pérez Prado, Mambo No. 8, F-major, 2/2, Tempo = 240, tatum = 1/8.

We used the frequency version of the time series representation of the score. Defining Y_t as follows:

$$Y_t = \text{frequency at tatum } t \quad (11)$$

$$= 0 \text{ if a rest occurred at time } t, \quad (12)$$

with respect to the representation Eq. (2), Y_t would be the λ_j of the τ_j near t.

First we calculated the periodogram of Y_t and minimized the negative of the approximate log likelihood give in Eq. (8) above. The results for Bach's *Coffee Tocata* can be seen in Fig. 5. Again the $1/f$ model is fitting well.

Table 1
Results of fitting the power spectrum

Song	$\hat{\beta}$	SE	Tot. Pts.	%not in c.i.
Baroque 1	1.192739	0.0391	1310	5.0
Baroque 2	1.066802	0.0391	1310	5.8
Classical 1	1.190916	0.0276	2621	6.5
Classical 2	1.244281	0.0276	2621	5.8
Romantic 1	1.185097	0.0276	2621	4.7
Romantic 2	1.085337	0.0276	2621	5.7
Atonal 1	1.083865	0.0276	2621	4.8
Atonal 2	1.143627	0.0276	2621	5.2
Spanish Guitar	1.228339	0.0391	1310	4.6
Spanish Guitar	1.236005	0.0391	1310	6.0
Jazz 1	1.121480	0.0391	1310	5.4
Jazz 2	1.217540	0.0276	2621	5.6
Afro-Cuban 1	1.116375	0.0276	2621	5.8
Afro-Cuban 2	1.169583	0.0276	2621	6.3
Rock & Roll	1.000069	0.0276	2621	6.1
Rock & Roll	1.109331	0.0276	2621	5.6
Hip Hop	1.022285	0.0276	2621	5.8
Hip Hop	1.075118	0.0391	1310	6.0

Next we fitted the $1/f$ model to the 12 score representations listed above. The results can be seen in Fig. 6. Again the $1/f$ model appears plausible. The values obtained for $\hat{\beta}$ are listed in Table 2. The lowest was for *Four* and the highest was for *Footprints* (the two Jazz tunes). Again the $1/f$ model appears to be performing reasonably.

5. Third-order spectra

Non-Gaussian aspects of music do not appear to have been investigated. In this connection the bispectrum and bicoherence are pertinent parameters. They are useful in both discerning non-Gaussianity and in examining for nonlinearity. Definitions and estimates are given in Appendix A.

Suppose the process Y_t is linear, that is

$$Y_t = \int a_{t-u}\, \mathrm{d}\varepsilon_u, \tag{13}$$

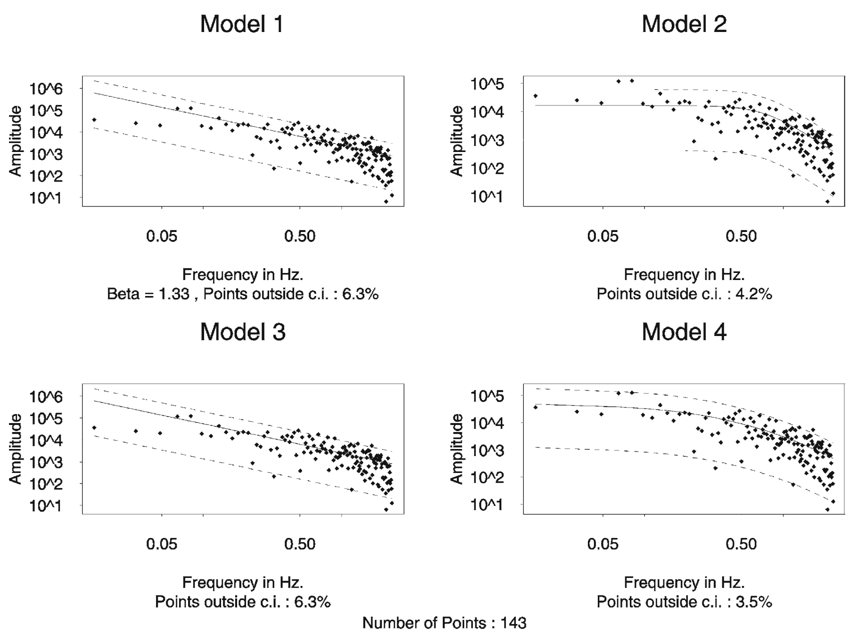


Fig. 5. Fitted models for the time series representation using frequencies of the score representation of the *Coffee Cantata*.

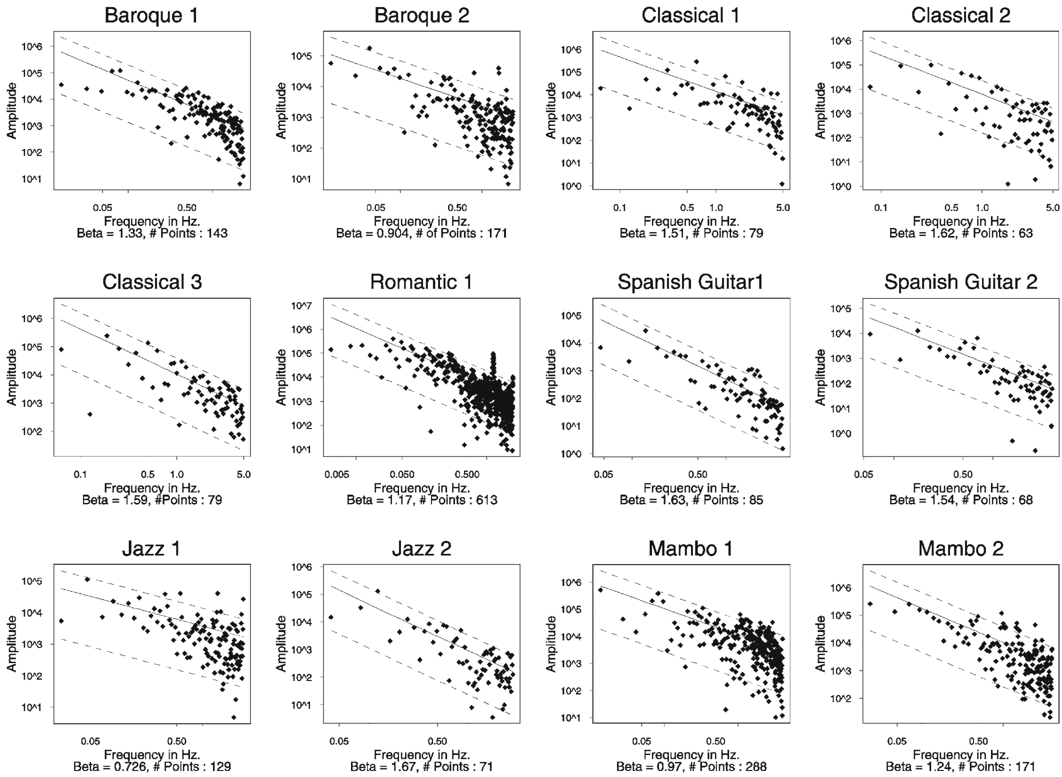


Fig. 6. Fitted $1/f$ model for the scores of the 12 pieces listed.

where ε_t is a process with independent increments having mean 0, variance σ^2, and third moment κ. (In the Gaussian case $\kappa = 0$.) Then the power spectrum of Y_t is

$$f_2(\lambda) = \frac{\sigma^2}{2\pi}|A(\lambda)|^2 \tag{14}$$

and the bispectrum is

$$f_3(\lambda, \mu) = \frac{\kappa}{(2\pi)^2} A(\lambda)A(\mu)\overline{A(\lambda + \mu)}, \tag{15}$$

where

$$A(\lambda) = \int e^{-i\lambda u} a_u \, du. \tag{16}$$

Table 2
Results of fitting the power spectrum

Song	$\hat{\beta}$	SE	Tot. Pts.	% not in c.i.
Baroque 1	1.330	0.1180	143	6.3
Baroque 2	0.904	0.1080	171	8.2
Classical 1	1.510	0.1590	79	14
Classical 2	1.620	0.1780	63	18
Classical 3	1.590	0.1590	79	3.8
Romantic	1.170	0.0571	613	7.8
Spanish Guitar 1	1.630	0.1530	85	5.88
Spanish Guitar 2	1.540	0.1710	68	5.88
Jazz 1	0.726	0.1250	129	7.8
Jazz 2	1.670	0.1680	71	8.4
Mambo 1	0.970	0.0833	288	9.3
Mambo 2	1.240	0.1080	171	9.4

If, for example, $A(\lambda) = 1/\lambda^{\beta/2}$ and the process is linear, then the power spectrum is $1/2\pi\lambda^{\beta}$ and the bispectrum

$$\frac{\kappa}{(2\pi)^2}\frac{1}{\lambda^{\beta/2}}\frac{1}{\mu^{\beta/2}}\frac{1}{(\lambda+\mu)^{\beta/2}}, \tag{17}$$

The spectrum may be estimated and examined to see if it is 0 (Gaussian process). Supposing that the denominator does not vanish, the bicoherence

$$|B(\lambda,\mu)|^2 = \frac{|f_3(\lambda,\mu)|^2}{f_2(\lambda)f_2(\mu)f_2(\lambda+\mu)} = \frac{\kappa^2}{(2\pi)^4}\frac{(2\pi)^6}{\sigma^6} = \gamma \tag{18}$$

is defined and constant for this linear process case, see [2]. In the case that the process Y_t is *reversible* (probabilistic properties of $\{Y_t\}$ and $\{Y_{-t}\}$ the same), the imaginary part of the bispectrum is identically 0. See [4]. Reversibility is not the property of most music.

The process (2) will have nonzero bispectrum to the extent that the frequencies λ_j, present for t near τ_j, satisfy relations such as $\lambda_j + \lambda_{j'} = \lambda_{j''}$.

Under regularity conditions (including stationarity and mixing) estimates $f_2^T(\lambda), f_3^T(\lambda,\mu)$ of the power and bispectra may be constructed that are asymptotically independent and normal. These may be used to form the bicoherence estimate, $|B^T(\lambda,\mu)|^2$, whose approximate statistical properties are indicated in Appendix A.

For a given sample value of the bicoherence, $|B^T(\lambda,\mu)|^2$, one may compute the approximate prob-value of achieving a value as large or larger in the null Gaussian case. The null distribution is an exponential, see Appendix A. Such

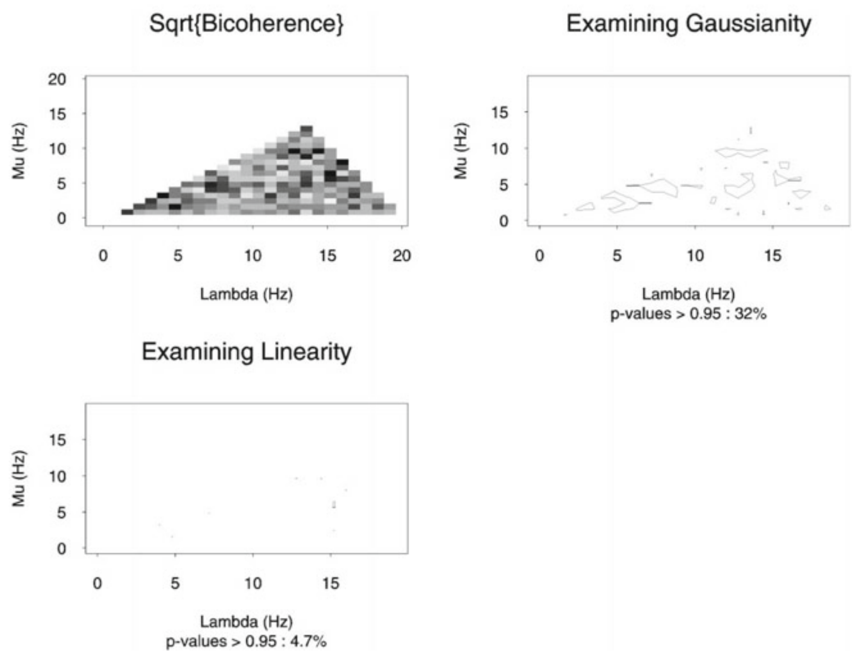


Fig. 7. Square root of the bicoherence estimate and contour plots of prob-values for non-Gaussianity and nonlinearity, respectively, for the smoothed zero crossings of the *Coffee Cantata*.

prob-values are contoured in Figs. 7–9, for the *Coffee Cantata.*

Likewise to assess the possibility that the basic process is linear, one may compute

$$\hat{\gamma} = \text{ave } |B^T(\lambda, \mu)|^2 , \tag{19}$$

with the average over all bicoherence estimates and then compute the approximate the prob-value corresponding to the deviate $||B^T(\lambda, \mu)|^2 - \hat{\gamma}|$. Again the prob-values are contoured for the *Coffee Cantata.* Details of the approximation are given in Appendix A.

These procedures, of using test statistics that are functions of (λ, μ), rather than some global statistic, have the advantage of indicating the character of departure if the null hypothesis appears rejected.

Nikias and Mendel [14] provide a review of higher order spectra and some of their uses.

5.1. Signal representation

We checked for non-Gaussianity and nonlinearity in the time series used in Section 4. The series studied, *Coffee Cantata*, lasted 64.15 s and was sampled at 8000 Hz. After applying the Voss filter every 200th observation was retained, 2566 data points in all. The spectra were estimated from this data. The resulting estimates can be seen in Figs. 7 and 8. For the zero crossings data, Gaussianity is being rejected, but not linearity. For the instantaneous power both Gaussianity and linearity are being rejected.

5.2. Score representation

Some similar computations were done for the score representation. In estimating the bicoherence

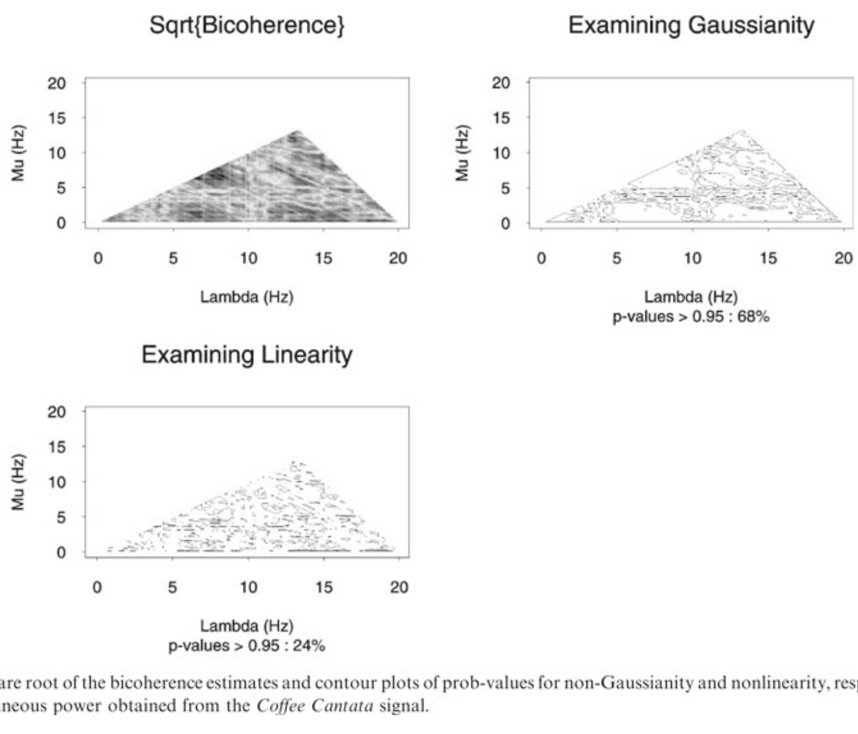


Fig. 8. Square root of the bicoherence estimates and contour plots of prob-values for non-Gaussianity and nonlinearity, respectively, for the instantaneous power obtained from the *Coffee Cantata* signal.

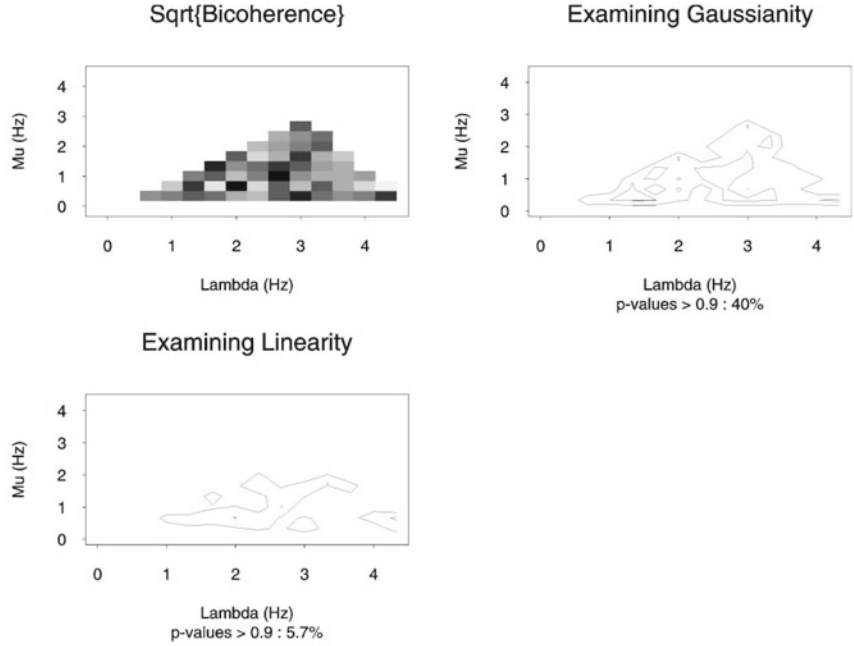


Fig. 9. Square root of the bicoherence estimates and contour plots of prob-values for non-Gaussianity and nonlinearity for the score of the *Coffee Cantata*.

we took 10 stretches. As a consequence the stretches were short, ranging from 12 to 120 points. The resulting estimates for the score of the *Coffee Cantata* can be seen in Fig. 9, now graphing the 50 and 90% contours. In this case Gaussianity appears rejected but not linearity.

6. Discussion and conclusions

We began with the question of what makes music, music. To address it we considered whether certain parametric forms fitted well, whether associated time series were Gaussian and whether they were linear. Broadly ranging selections of pieces were analyzed. The model $1/f^{\beta}$ with β near 1 appeared to fit the scores well, as opposed to alternatives allowing more curvature or flatness at low frequencies. The same can be said for the derived processes of zero crossings and instantaneous power. In each, the hypothesis of Gaussianity (really 0 bispectrum) was rejected. The conclusions regarding linearity were not so clear.

We have acted as if the processes involved were stationary. To the extent that they are not, the parameters and estimates may be treated as if they are focussed on an average of instantaneous spectra obtaining for the processes involved. The statistical packages of Matlab and S-plus were employed.

In future work the trispectrum will be considered. It will allow further assesment of linearity. Other values for the passband of the lowpass filter, here 20 Hz, will also be considered. The signal computations were carried through only for *Coffee Cantata*. The other pieces will be studied as well.

Acknowledgements

We thank David Wessel and Steven Clark for a variety of helpful remarks. We thank Ofer Licht for help in entering the CD music. The work of David R. Brillinger was supported by NSF grants DMS-9300002 and DMS-9625774. The work of Rafael A. Irizarry was supported by a National Science Foundation Graduate Research Fellowship.

Appendix A

A.1. Third-order spectra

We provide the basic definitions and properties in order that others can directly reproduce such a study.

A.1.1. Definitions and estimates

Bispectral Analysis is of use in discerning non-Gaussianity of a time series and also in examining the series for nonlinearity.

Let Y_t, $t = 0, \pm 1, \pm 2, \ldots$, denote a stationary time series. Let it have mean c_1, co-variance function $c_2(u)$ and third moment function

$$c_3(u, v) = E([Y_{t+u} - c_1][Y_{t+v} - c_1][Y_{t+u+v} - c_1]). \tag{A.1}$$

The *bispectrum* at *bifrequency* (λ, μ) is defined by

$$f_3(\lambda, \mu) = \frac{1}{(2\pi)^2} \sum \sum c_3(u, v) e^{-i(u\lambda + v\mu)}, \tag{A.2}$$

and the *bicoherence* by

$$|B(\lambda, \mu)|^2 = \frac{|f_3(\lambda, \mu)|^2}{f_2(\lambda) f_2(\mu) f_2(\lambda + \mu)}, \tag{A.3}$$

The fundamental domain of these parameters is $0 \leqslant \mu \leqslant \lambda$, $\lambda + \mu/2 \leqslant \pi$.

There are a variety of fashions by which the bispectrum may be estimated. A convenient one is: let the data be broken into L stretches of length V, so that $T = LV$. Next compute the tapered Fourier Transform of the lth stretch,

$$d^V(\lambda; l) = \sum_{v=0}^{V-1} h\left(\frac{v+1}{V+1}\right) Y_{lV+v} e^{-iv\lambda} \tag{A.4}$$

for $l = 0, \ldots, L-1$. Then form the third-order periodogram of the lth stretch,

$$I_3^V(\lambda, \mu; l) = \frac{1}{(2\pi)^2 V h_3} d_Y^V(\lambda; l) d_Y^V(\mu; l) \overline{d_Y^V(\lambda + \mu; l)}, \tag{A.5}$$

where $h_3 = \int h(u)^3 \, du$. The estimate of the bispectrum is now

$$f_3^T(\lambda, \mu) = \frac{1}{L} \sum_{l=0}^{L-1} I_3^V(\lambda, \mu; l). \tag{A.6}$$

One reference is [11]. In forming the estimate of the bicoherence, the power spectrum is estimated by similarly averaging the second-order periodograms of the L stretches.

In our empirical work no taper was employed, rather the series were prefiltered by fitting an autoregressive, prior to computing the spectral quantities. Such a linear filtering retains the 0 bispectral property of a Gaussian process and the linearity property of a linear process.

A.1.2. Statistical properties of the estimates

Suppose that the bispectrum is estimated, as above, by averaging the third-order periodograms of L contiguous segments of length V of a series of length $T = LV$. Then, for (λ, μ) not on the boundary of the fundamental domain, $f_3^T(\lambda, \mu)$ is asymptotically complex normal with mean $f_3(\lambda, \mu)$ and variance

$$\frac{h_6}{h_3^2} \frac{1}{2\pi} f_2(\lambda) f_2(\mu) f_2(\lambda + \mu) \frac{V}{L}, \tag{A.7}$$

provided $V, L/V \to \infty$ as $T \to \infty$. It is noteworthy that for consistency a large number, L, of individual stretches will be required. Further estimates at distinct frequencies are asymptotically independent.

It follows that when $f_3(\lambda, \mu) = 0$, $|f_3^T(\lambda, \mu)|^2$ is asymptotically

$$\frac{h_6}{h_3^2} \frac{V}{L} \frac{1}{2\pi} f_2(\lambda) f_2(\mu) f_2(\lambda + \mu) \chi_2^2/2, \tag{A.8}$$

which result may be used to examine the hypothesis $f_3(\lambda, \mu) = 0$. In the examples, prob-values based on this distribution are graphed.

In the case that $f_3(\lambda,\mu)\neq 0$ the variate $|f_3^T(\lambda,\mu)|^2$ will be approximately normal with mean $|f_3(\lambda,\mu)|^2$ and variance

$$2\frac{h_6}{h_3^2}\frac{V}{L}\frac{1}{2\pi}f_2(\lambda)f_2(\mu)f_2(\lambda+\mu)|f_3(\lambda,\mu)|^2. \quad \text{(A.9)}$$

In other words the large sample distribution of the bicoherence estimate

$$|B^T(\lambda,\mu)|^2=\frac{|f_3^T(\lambda,\mu)|^2}{f_2^T(\lambda)f_2^T(\mu)f_2^T(\lambda+\mu)} \quad \text{(A.10)}$$

will be approximately exponential with mean

$$\frac{h_6}{h_3^2}\frac{V}{L}\frac{1}{2\pi} \quad \text{(A.11)}$$

when $f_3(\lambda,\mu)=0$. It will be approximately normal with mean, the bicoherence,

$$|B(\lambda,\mu)|^2=\frac{|f_3(\lambda,\mu)|^2}{f_2(\lambda)f_2(\mu)f_2(\lambda+\mu)} \quad \text{(A.12)}$$

and variance

$$2\frac{h_6}{h_3^2}\frac{V}{L}\frac{1}{2\pi}\frac{|f_3(\lambda,\mu)|^2}{f_2(\lambda)f_2(\mu)f_2(\lambda+\mu)} \quad \text{(A.13)}$$

when $f_3(\lambda,\mu)\neq 0$. The quantity $|B^T(\lambda,\mu)|$ will then be approximately normal with mean $|B(\lambda,\mu)|$ and variance

$$\frac{1}{2}\frac{h_6}{h_3^2}\frac{1}{2\pi}\frac{V}{L}. \quad \text{(A.14)}$$

This approximation follows via the delta method.

A.1.3. Related work

Rosenblatt and Van Ness [18] developed various asymptotic properties of bispectral estimates, as did Brillinger [2] for higher-order spectral estimates. Huber et al. [10] considered the estimation of the bicoherence and in particular suggested approximating its distribution, when the population value was 0, by a χ_2^2. Elgar and Guza [7] investigated the accuracy of this approximation. Rao and Gabr [17] and Hinich [8] proposed global bispectrum-based tests for the non-Gaussianity and nonlinearity of a stationary process. Rao and Gabr structured the problem as assessing whether all components of a multivariate normal have the same mean. Hinich (see also [5]) based tests on the interquantile range of sample bicoherence values. Terdik and Math [20] note that the bispectrum of a process, such that the linear predictor is the quadratic, satisfies a particular algebraic identity and use this to assess possible linearity.

A.2. Obtaining the numerical representations

A.2.1. Signal representation

A song was chosen from a Compact Disc. It was down-loaded into a Mono .au file sampled at 8000 Hz using a CD-ROM and software for the Sparc machines. To go from stereo to a mono signal the two channels were averaged. Our statistical analysis was done mostly by S-Plus which cannot read .au files. We altered Thau's program `xplay`, a sound player for Sun Sparc machines, which can handle AIFF, .au, and some WAVE files, so that it would save a file with the numbers corresponding to the sampled signal in a file readable to S-Plus. Due to technical details of the way Compact Discs are recorded and the way `xplay` works the units of the sampled audio signal are completely arbitrary.

To obtain the smoothed zero crossings or 'instantaneous' pitch and power from the sampled signals we wrote C programs that performed the zero crossing calculation, the bandpass filtering, the squaring and the lowpass filtering relatively quickly. The filtering was done by calculating the FFT of the signal, setting the coefficients of the pertinent frequencies to zero and then performing the inverse FFT.

A.2.2. Score representation

A piece of music was selected. Then using an EMU Proteus Keyboard, a Mac Power Book 520 and a program we wrote in Max (see [19] for some information on Max) we saved the midi-numbers and duration into text files. We used MIDI-note number 36 (lower than the lowest note in any of the score) to denote rests. These files where made readable to S-Plus and Matlab using Pearl. Using S-Plus we created various functions that converted the raw data into objects of the time series and marked point process representations, respectively.

A.2.3. Time series representation

For each score we had two time series; one for the MIDI-numbers and the other for the frequencies. To do the analysis on the MIDI-number time series we took care of the rests by extending the previous notes over the duration of the rests. In the case of a song starting on a rest we simply ignored that part of the song.

For the case of the respective frequencies we assigned frequency 0 to the rests. This representation is probably more representative of the music, so we focused our attention on it.

A.2.4. Playing the data

To play the data we used Matlab. The command *sound* takes as an argument a vector. This vector is taken to be a signal sampled at 8192 Hz (at least on Sparcs, it varies for other computers). Each kth element of the vector is taken to be the sample at time $k/8192$ seconds. The following Matlab code performs the work.

```
%note is a vector containing the MIDI-
  note numbers(rests are 36)
%dur is a vector containing the durations
  of these notes
%rate is the sample rate
for i = 1:length(note);
     t = [0:rate*dur(i)];
     window = 1-cos(2*pi/(rate*dur(i))*t);
       %TUKEY's WINDOW
     t = t/rate*2*pi;
     if note(i) = = 36   %in our case MIDI-
                 number 36 in the raw data
                         %represented rests
       signal = [signal, t*0];
     else
       y = window * sin(note(i)*t);
       signal = [signal, y];
     end;
end;
sound (signal)
```

A.3. Notes on the Voss technique

Suppose the signal may be written

$$Y(t) = R(t)\cos(Z(t)t + \phi(t)) \tag{A.15}$$

with $Z(t)$ the instantaneous frequency, $R(t)$ a slowly changing amplitude and $\phi(t)$ a slowly changing phase. For $Z(t)$ in the passband of 0.1 to 10 kHz, after filtering the signal will remain essentially Eq. (A.15). With squaring it becomes

$$R(t)^2[1 + \cos(2Z(t)t + 2\phi(t))]/2. \tag{A.16}$$

After the lowpass filtering to [0, 0.01] kHz one has approximately

$$R(t)^2/2 \tag{A.17}$$

in other words essentially the squared envelope of the signal (A.15).

In terms of the representation (2) one has approximately

$$V_j^2 h\left(\frac{t-\tau_j}{\sigma_j}\right)^2/2 \tag{A.18}$$

for t near τ_j, provided λ_j is in the band [0.1,10] kHz. A spectrum analysis of this will bring out the periodicity properties of the point process $\{\tau_j\}$, which in music can be regarded as the rhythmic structure. If the probability distribution of the $(\tau_{j+1} - \tau_j)$ is long-tailed then the fitted process (2) can have $1/f$ spectra, see [12]. Also notice that V_j^2 is related to the accents or 'dynamics' of the music.

References

[1] J. Bilmes, Timing is of the essence: perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm, Master's Thesis, Massachusetts Institute of Technology, 1993.

[2] D.R. Brillinger, An introduction to polyspectra, Ann. Math. Statist. 36 (1965) 1351–1374.

[3] D.R. Brillinger, Time Series: Data Analysis and Theory, Holt, Reinhart and Winston, New York, 1975.

[4] D.R. Brillinger, M. Rosenblatt, Computation and interpretation of the kth order spectra, in: B. Harris (Ed.), Spectral Analysis of Time Series, Wiley, New York, 1967, pp. 189–232.

[5] P.L. Brockett, M.J. Hinich, D. Patterson, Bispectral-based tests for the detection of Gaussianity and linearity in time series, J. Amer. Statist. Assoc. 83 (1988) 657–664.

[6] K. Dzhaparidze, Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series, Springer, New York, 1986.

[7] S. Elgar, R.T. Guza, Statistics of bicoherence, IEEE Trans. Acoust. Speech Signal Process. 36 (1988) 1667–1668.

[8] M.J. Hinich, Testing for Gaussianity and linearity of a stationary time series, J. Time Ser. Anal. 3 (1982) 169–176.

[9] K.J. Hsu, A.J. Hsu, Fractal geometry of music, Proc. Natl. Acad. Sci. USA 87 (1990) 938–941.

[10] P.J. Huber, B. Kleiner, T. Gasser and G. Dumermuth, Statistical methods for investigating phase relations in stationary stochastic processes, IEEE Trans. Audio Electro AU-19 (1971) 78–86.

[11] K.S. Lii, M. Rosenblatt, C. VanAtta, Bispectral measurements in turbulence, J. Fluid Mech. 77 (1976) 45–62.

[12] S.B. Lowen, M.C. Teich, Fractal renewal processes generate $1/f$ noise, Phys. Rev. 47 (1993) 992–1001.

[13] G. Loy, Musicians make a standard: the MIDI phenomenon, Computer Music Journal 9 (4); reprinted in: C. Roads (Ed.), The Music Machine, MIT Press, Cambridge, 1989.

[14] C.L. Nikias, J.M. Mendel, Signal processing with higher-order spectra, IEEE Signal Process. Mag. (1993) 10–37.

[15] J.R. Pierce, The Science of Musical Sound, Freeman, New York, 1992.

[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, Cambridge, 1992.

[17] T. Subba Rao, M. Gabr, A test for linearity of stationary time series, J. Time Series 1 (1980) 145–148.

[18] M. Rosenblatt, J.W. Van Ness, Estimation of the bispectrum, Ann. Math. Statist. 36 (1976) 1120–1136.

[19] R. Rowe, Interactive Music Systems, MIT Press, Cambridge, MA, 1994.

[20] Gy. Terdik, J. Màth, A new test of linearity for the based on its bispectrum, University of Debrecen, Technical Report No. 95/152, 1996.

[21] D. Wessel, D. Bristow, Z. Settel, Control of phrasing and articulation in synthesis, ICMC Proc. (1987) 108–116.

[22] R.F. Voss, Fractals in nature: from characterizations to simulation, in: H.-O. Peitgen, D. Saupe (Eds.), The Science of Fractal Images, Springer, Berlin, 1988, pp. 29–69.

[23] R.F. Voss, J. Clarke, '$1/f$ Noise' in music and speech, Nature 258 (1975) 317–318.

[24] R.F. Voss, J. Clarke, $1/f$ Noise in music: music from $1/f$ Noise, J. Acoust. Soc. Amer. 63 (1978) 258–263.

1

Some Examples of Empirical Fourier Analysis in Scientific Problems

DAVID R. BRILLINGER University of California, Berkeley, California

"One can FT anything—often meaningfully."

J. W. Tukey

1. INTRODUCTION

As a concept and as a tool, the Fourier transform is pervasive in applied mathematics, computing, mathematics, probability and statistics as well as in substantive sciences such as chemistry, geophysics and physics. This chapter presents a review of such applications and then four personal analyses of scientific data based on Fourier transforms. Specific points made include: Fourier analysis is conceptually simple, its concepts often have direct physical interpretations, useful statistical properties are available, and there are various interesting connections between the mathematical and physical concepts.

By Fourier analysis is meant the study of spaces and functions, making substantial use of sine and cosine functions. The subject has a long and glorious history. In particular, fundamental work has been carried out by both mathematicians and applied scientists. Fourier analysis remains of interest to mathematicians because generalizations seem inexhaustible and because there are continual surprises. Classic works by mathematicians

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_16, © Springer Science+Business Media, LLC 2012

include: Wiener (1933), Bochner (1959, 1960) and Zygmund (1968). These particular authors are concerned with functions on the line or on a general Euclidian space. Works on extensions to general groups include: Loomis (1953), Rudin (1962), Hewitt and Ross (1963), Katznelson (1976). More recent books are Terras (1988) and Körner (1989), the former particularly addressing the nonabelian case, the latter presenting a variety of historical examples and essays on specific topics.

In contrast, the Fourier transform is of interest to statisticians because it proves inordinately useful in the analysis of data and eases the development of various theoretical results. Noteworthy contributions to statistics have been made by Slutsky (1934), Cramér (1942), Good (1958), Yaglom (1961), Tukey (1963), Hannan (1965, 1966), Priestley (1965), Bloomfield (1976), Diaconnis (1988, 1989). Slutsky developed some of the statistical properties of the Fourier transform of a stretch of time series values. Cramér set down a Fourier representation (see Sec. 4.1) for stationary processes. The representation admitted many extensions and made transparent the effect of a variety of operations on processes. Good and Tukey indicated how the transform could be computed recursively and hence quickly in certain circumstances. Yaglom extended the domain of application to processes defined on compact and locally compact groups. Hannan considered problems for other groups than Yaglom and presented material relevant to practical applications. Priestley provided a frequency domain representation to describe nonstationary processes. Bloomfield made complicated results available to a broad audience. Diaconnis considered symmetric and permutation groups and the analysis of ordered data.

Particular uses of the empirical Fourier transform include: the development of estimates of finite dimensional parameters appearing in time series models (Whittle (1952), Dzhaparidze (1986), Feuerverger (1990)), the assessment of goodness of fit of models (Feigin and Heathcote (1976)), and the deconvolution of random measurements (Fan (1992)). Fourier analysis has a special place amongst the tools of statistics for the concepts often have their own physical existence.

There are special computational, mathematical and statistical properties and surprises associated with the Fourier transform. These include: the central limit theorems for the stationary case with approximate independence at particular frequencies, the existence of fast Fourier transforms, (Good (1958), Tukey (1963), Cooley and Tukey (1965), Rockmore (1990)) the need for convergence factors, the ideas of aliasing.

Section 2 concerns some particular physical situations. Section 3 contains pertinent background from analysis. Section 4 contains stochastic background. Section 5 presents analyses of four data sets from the natural sciences and the author's experience. The examples highlight approximation, shrinkage estimation, the method of stationary phase, central limit theorems and uncertainty estimation. The first example, concerning crystalographic data, involves the empirical representation of a basic function on the plane by an expansion in sines and cosines. This makes sense because of periodicities inherent in the crystal structure. The example also involves shrinkage of the coefficients of the expansion in order to obtain improved estimates. The second analysis is of a record of an earthquake that took place in Siberia as recorded at Uppsala, Sweden. The oscillatory character of the data may be understood heuristically via the method of stationary phase, to be described below. A model of the transmission medium is constructed and model assessment carried out by a sliding or dynamic Fourier analysis. This last may be viewed as a form of wavelet analysis. The third analysis, concerned with nuclear magnetic resonance (NMR) spectroscopy, employs Fourier analysis to obtain physical insight into the behavior of an input-output system, and then makes use of cross-spectral analysis to estimate the transfer function of the system. The periodogram of the residuals is employed to assess the fit. The final example involves both wavelet and Fourier analysis. It is concerned with the question of whether a microtubule moves steadily or via jumps. The Fourier analysis is employed in this case to obtain uncertainty estimates in the presence of stationary noise. Section 6 contains conclusions and indicates open problems.

2. SOME PHYSICAL EXAMPLES OF FOURIER ANALYSIS

Cycles, periods, and resonances have long been noted in scientific discussions of astronomy, vibrations, oceanography, sound, light and crystalography amongst other fields. In technology oscillations occur often for example in telephone, radio, TV and laser engineering. Natural operations occur commonly that correspond with linear and time invariant systems as defined in Section 3 below. These are the eigenoperations of Fourier analysis.

Fourier analysis is sometimes tied specifically to the physics of a problem. For example Bazin et al. (1986) physically demonstrate the operations/concepts of translation, linearity, similarity, convolution and Parseval's theorem for the Fourier transform via diffraction experiments with laser light. The Fourier transform here is formed via a lens. See Goodman (1968) Shankar et al. (1982), Glaeser (1985) for a discussion of the optics involved.

An important example arises in radio astronomy. Suppose there is an array of receivers. Suppose there is a small incoherent source, at great distance, producing a plane travelling wave. If $Y(x, y, t)$ denotes the radio field measurement made at time t on a telescope located at position (x, y), then

$$E\{Y(x+u, y+\nu, t)\overline{Y(x,y,t)}\} = \int\int f(\alpha,\beta)e^{i(u\alpha+\nu\beta)}d\alpha\, d\beta \tag{2.1}$$

where (α, β) are the coordinates of the source of interest in the sky and $f(\alpha, \beta)$ is its brightness distribution as a function of (α, β). In other words, the Fourier transform is the quantity observed. The result Eq. (2.1) is known as the van Cittert–Zernike Theorem, see Born and Wolf (1964).

Linear time invariant systems abound in nature. They have the property of carrying cosinusoids into cosinusoids. Nowadays in science there is much concern with nonlinear operations and phenomena. Impressively, the classic trigonometric identity

$$[\cos \lambda t]^2 = \tfrac{1}{2}\cos 2\lambda t + \tfrac{1}{2} \tag{2.2}$$

is "demonstrated" in Yariv (1975) via a color plate showing red laser light becoming blue on passing through a crystal. The crystal involved squares the signal as in Eq. (2.2). A wavelength of 6940 Å (red) becomes one of 3970 Å (blue). Bloembergen (1982), Moloney and Newell (1989) discuss such nonlinear aspects of light. The appearance of harmonics such as in Eq. (2.2) leads to a consideration of higher-order spectra.

The Fourier transform is continually employed in the solution of equations of motion associated with physical phenomena and mathematicians have focussed on consequent cycles and harmonics. For example, Hirsch (1984) has remarked that "Dynamicists have always been fascinated (not to say obsessed) by periodicity." In that connection Ruelle (1989) makes effective use of the Fourier transform in the study of dynamic systems, specifically addressing aspects of chaos, periods and scaling.

The Fourier transform leads to entities with direct physical interpretations. One can point to a variety of success stories of the application of Fourier analysis. Michaelson (1891a, b) measured visibility curves, essentially the modulus of a Fourier transform, and after an inversion thereby inferred that the red hydrogen line was a doublet. This inference of splitting ultimately led to important developments in quantum mechanics. Tidal components caused by the sun, moon and planets have been isolated by Fourier analysis, see Cartwright (1982), Båth (1974), Bracewell (1989). Katz and Miledi (1971) inferred the mechanism of acetylcholine release via a Fourier analysis. Bolt et al. (1982) saw a fault rupturing in an earthquake by a frequency-wavenumber spectral analysis. Finally it may be noted that R. R. Ernst received the 1991 Nobel Prize in Chemistry for developing the technique of Fourier transform spectroscopy, see Amato (1991). A discussion of a variety of other physical examples may be found in Lanczos (1966), Båth (1974), Bracewell (1989).

3. SOME ANALYTIC BACKGROUND

3.1 The Fourier Case

Consider a square integrable function $g(x), 0 \leq x < 2\pi$. In this simple case Fourier analysis is built upon the values

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx} g(x)\, dx \tag{3.1}$$

$k = 0, \pm 1, \pm 2, \ldots$, and Fourier synthesis on expansions

$$g(x) \approx \sum_{k=-\infty}^{\infty} c_k e^{ixk} \tag{3.2}$$

The functions $\exp\{ikx\}$, $k = 0, \pm 1, \pm 2, \ldots$ here are orthogonal on $[0, 2\pi)$ and this connects Eqns. (3.1) and (3.2).

One important use of Fourier methods is the approximation of functions. If the values $c_k, k = 0, \pm 1, \ldots \pm K$ of Eq. (3.1) are available, a naive approximation to $g(x)$ is provided by

$$\sum_{k=-K}^{K} c_k e^{ixk} \tag{3.3}$$

However early researchers found that the approximation of Eq. (3.3) was often improved by inserting multipliers, w_k^K, such as $1 - |k|/K$, into the expansion and employing

$$g^K(x) = \sum_{k=-K}^{K} w_k^K c_k e^{ixk}. \tag{3.4}$$

instead of Eq. (3.3). Defining the kernel

$$W^K(x) = \sum_{k=-K}^{K} w_k^K e^{ixk}$$

Eq. (3.4) can be written

$$\int_0^{2\pi} W^K(y - x) g(x)\, dx \tag{3.5}$$

and one sees that Eq. (3.4) is a weighted average of the desired $g(\cdot)$. The effect of the multipliers, in some cases, is to improve the approximation by damping down the more rapidly oscillating terms in the expansion. This idea of damping down will recur below in the consideration of shrinking to improve estimates. The expression of Eq. (3.5) may be used to study directly the effect of the kernel function on the approximation. Timan (1963), Butzer and Nessel (1971) are books specifically concerned with approximations based on Fourier expressions.

In work with data values Y_t observed at $t = 0, \ldots, T-1$ one might replace Eq. (3.1) with

$$\frac{1}{T}\sum_{t=0}^{T-1} \exp\left\{\frac{-i2\pi kt}{T}\right\} Y_t$$

having written $g(2\pi t/T) = Y_t$. As referred to earlier there are fast algorithms to evaluate this.

A second important use of Fourier analysis is in the study of time invariant systems. A simple linear time invariant system is described by

$$Y_t = \sum_{t=-\infty}^{\infty} c_{t-s} X_s.$$

i.e., a convolution. The response of this system to the input $X_t = \exp\{i\lambda t\}$ is

$$Y_t = C(\lambda) X_t \tag{3.6}$$

with $C(\lambda)$ the Fourier transform

$$C(\lambda) = \sum_{s=-\infty}^{\infty} e^{-i\lambda s} c_s$$

for $0 \leq \lambda < 2\pi$. This function is referred to as the transfer function of the system. Cosinusoids, $\exp\{i\lambda t\}$, are seen to be carried into cosinusoids. A variety of physical systems have this property to a good approximation.

Nonlinear time invariant systems may sometimes be approximated by Volterra expansions of the form

$$Y_t = \sum_{s=-\infty}^{\infty} c_{t-s} X_s + \sum_{s=-\infty}^{\infty} \sum_{s'=-\infty}^{\infty} d_{t-s,t-s'} X_s X_{s'} + \cdots$$

The input $X_t = \exp\{i\lambda t\}$ here leads to the output

$$C(\lambda) e^{i\lambda t} + D(\lambda, \lambda) e^{i2\lambda t} + \cdots$$

where $C(\lambda)$ is given above and

$$D(\lambda, \mu) = \sum_s \sum_{s'} e^{-i\lambda s - i\mu s'} d_{s,s'}$$

In such a nonlinear system one sees harmonics of the frequencies in the input appearing in the output. The laser example of Sec. 2 involved a system that was quadratic.

Fourier analysis is useful in work with constant coefficient differential equations. These show the occurrence of oscillations and are often effective models of physical systems. Consider for example the linear system

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{A}\mathbf{S}(t) + \mathbf{B}X(t)$$

with $\mathbf{S}(\cdot)$ vector-valued and $X(\cdot)$ scalar. Supposing

$$\mathbf{S}(t) = \int e^{it\lambda}\mathbf{s}(\lambda)\, d\lambda$$

and

$$X(t) = \int e^{it\lambda}x(\lambda)\, d\lambda$$

by Fourier analysis one has the solution directly as

$$\mathbf{S}(\lambda) = (i\lambda\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}x(\lambda)$$

Supposing $x(\lambda)$ constant and the latent values, μ_j, of $\mathbf{A}$ to be distinct this may be written

$$\mathbf{S}(t) = \sum_j \mathbf{a}_j e^{i\mu_j t}$$

for some vectors $\mathbf{a}_j$.One sees the occurrence of oscillations at frequencies Re μ_j. One reference concerning such differential equations is Hochstadt (1964).

Turning to a further technique of Fourier analysis, that will be basic in one of the examples below, suppose that one is considering, for large x, an integral of the form

$$\int e^{ik(\lambda)x}R(\lambda)\, d\lambda$$

The method of stationary phase approximates this by

$$e^{\operatorname{sgn} k''(\lambda_0)i\pi/4}\sqrt{2\pi/(x|k''(\lambda^0|)}R(\lambda_0)e^{ik(\lambda_0)x}.$$

where λ_0 satisfies $k'(\lambda_0) = 0$. References include Barndorff-Nielsen and Cox (1989) and Aki and Richards (1980). The idea is that unless the $k(\lambda)$ is near 0 the rapidly oscillating multipliers $\cos k(\lambda), \sin k(\lambda)$ will give the integral value 0.

3.2 The Wavelet Case

Wavelet analysis is enjoying a surge of contemporary investigation and is a competitor of Fourier analysis. It may be viewed as Fourier analysis with the sine and cosine functions replaced by other families of (orthogonal) functions. There are many similarities between Fourier and wavelet analysis. Consider the expansion in Eq. (3.2) with the coefficients in Eq. (3.1). The expansion is based on the fact that the sine and cosine functions provide a basis for $L_2[0, 2\pi)$. In wavelet analysis other systems of functions are used, see e.g., Strichartz (1993), Benedetto and Frazier (1994). Wavelets are of practical importance because they can sometimes provide more parsimonious descriptions than Fourier ones.

Wavelets often focus on local versus global behavior and in particular can pick up transient behavior. Basic is a (mother) wavelet $\psi(\cdot)$ nonzero only on say the unit interval $[0, 1)$. Given a square-integrable function $g(x)$, one considers an expansion

$$g(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \beta_{jk}\psi_{jk}(x) \tag{3.7}$$

with

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$$

and

$$\beta_{jk} = \int \psi_{jk}(x)g(x)\,dx \tag{3.8}$$

The family $\{\psi_{jk}(\cdot)\}$ is taken to be orthonormal and complete, see e.g., Daubechies (1992), Walter (1992, 1994), Strichartz (1993), Benedetto and Frazier (1994).

The expansion in Eq. (3.7) represents $g(\cdot)$ in terms of functions with support individually on dyadic intervals $[k/2^j, (k+1)/2^j]$ for j, k integers. It suggests an approximation

$$g^{JK}(x) = \sum_{|j| \leq J} \sum_{|k| \leq K} \beta_{jk}\psi_{jk}(x) \tag{3.9}$$

to $g(x)$. This may be written as

$$g^{JK}(x) = \int W^{JK}(x, y)g(y)\,dy \tag{3.10}$$

the kernel being

$$W^{JK}(x, y) = \sum_{j,k} \psi_{jk}(x)\psi_{jk}(y) \tag{3.11}$$

This kernel will tend to a delta function in various circumstances, see Walter (1992). Equation (3.10) can be used to study the degree of approximation directly as could Eq. (3.5) in the Fourier case. Equations (3.10) and (3.11) are wavelet analogs of Eqns. (3.4) and (3.5).

In the case of a discontinuous function, as will occur in Example 5.4, a particular wavelet analysis is especially suitable, namely Haar wavelet analysis. This analysis is based on the function

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \le x < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

In the Haar case the kernel is

$$W_n(x,y) = 2^n \phi(2^n y - [2^n x]) \tag{3.12}$$

with $[\cdot]$ here referring to integral part and $g_n(x)$ of Eq. (3.10) a local mean, $g_n(x) = |I|^{-1} \int_I g(y)\,dy$, x being in the particular interval $I = [m/2^n, (m+1)/2^n)$, see Fine (1949), Walter (1992).

There are empirical versions of Eq. (3.8) for use when discrete time data $Y_t, t = 0, \ldots, T-1$ are available. One computes for example

$$\hat{\beta}_{jk} = \frac{1}{T} \sum_{t=0}^{T-1} \psi_{jk}(t/T)\, Y_t \tag{3.15}$$

Just as there are fast Fourier transforms, there are fast wavelet transforms, Strang (1993). Also one can write $p2^j$ for 2^j above, with no real change in concept, but improved approximations in practice. The dynamic spectrum analysis of Example 5.2 is one type of wavelet analysis with $j = j_0$ and $\psi(x) = \exp\{-i2\pi x\}$.

Insertion of multipliers, as in Eq. (3.4) for Fourier approximation, is fundamental. This will be discussed later.

4. STOCHASTICS AND STATISTICS

In this section the quantities being transformed will be random.

4.1 Stationary Processes

Fourier analysis is basic to dealing with stationary random processes. A process, Y_t, is said to be second-order stationary if $\text{cov}\{Y_{t+u}, Y_t\}$ exists for $t, u = 0, \pm 1, \pm 2, \ldots$ and does not depend on t. In practice this often appears a reasonable working assumption. In the case of

$Y_t, t = 0, \pm 1, \pm 2, \ldots$ a second-order stationary process, following Cramér (1942), one has the Fourier representation

$$Y_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dZ(\lambda) \tag{4.1}$$

with $Z(\cdot)$ a random function such that

$$\text{cov}\{dZ(\lambda), dZ(\mu)\} = \delta(\lambda - \mu) f(\lambda)\, d\lambda\, d\mu$$

$-\pi < \lambda, \mu \leq \pi, f(\cdot)$ being the power spectrum of Y and $\delta(\cdot)$ the Dirac delta function. The Cramér representation has the advantage of taking one directly to the Fourier domain and thereby making some operations on the process clearer. The series $\mathbf{Y}_t$ may be vector-valued. Then the cross-spectral density matrix, $\mathbf{f}(\cdot)$, is given by

$$\text{cov}\{d\mathbf{Z}(\lambda), d\mathbf{Z}(\mu)\} = \delta(\lambda - \mu)\mathbf{f}(\lambda)\, d\lambda\, d\mu$$

Cross-spectrum analysis is useful for system analysis, i.e., estimating for example the transfer function of a linear time invariant system.

Higher-order spectra may be defined directly via $Z(\cdot)$, e.g., the bispectrum $f(\lambda, \mu)$ at frequency λ, μ is given by

$$\text{cum}\{dZ(\lambda), dZ(\mu), dZ(\nu)\} = \eta(\lambda + \mu + \nu) f(\lambda, \mu) d\lambda\, d\mu\, d\nu$$

where $\eta(\lambda)$ is the 2π periodic extension of the Dirac delta function.

Empirical Fourier analysis, e.g., of residuals of a fit, provides a diagnostic using in particular the result that if the process is white noise, the power spectrum is constant in frequency, λ.

Blackman and Tukey (1959), Båth (1974), Brillinger (1975) and Bloomfield (1976) are books focussing on the empirical Fourier analysis of time series

4.2 Central Limit Theorems

In classic forms the central limit theorem is concerned with the distributions of sums of independent random variables

$$S_T = Y_0 + Y_1 + \cdots + Y_{T-1}$$

and their approximate normality with variance $T\sigma^2$ for large T. It is usual to assume that the Y's are identically distributed.

At some point engineers began promulgating a folk theorem to the effect that narrow-band noise is approximately Gaussian, [see Leonov and Shiryaev (1960), Picinbono (1960), Rosenblatt (1961)]. One fashion to formulate this remark is as a statement that

$$S_T(\lambda) = Y_0 + e^{-i\lambda} Y_1 + \cdots + e^{-i\lambda(T-1)} Y_{T-1} \tag{4.2}$$

$0 \leq \lambda < 2\pi$, is approximately (complex) normal for each λ. Under stationarity and mixing assumptions for the series Y_t, the variance of Eq. (4.2) is approximately

$$2\pi T f(\lambda) \tag{4.3}$$

with $f(\lambda)$ the power spectrum of Eq. (4.1) at frequency λ. Surprisingly, the values of $S_T(\lambda)$ at distinct frequencies of the form $\lambda = 2\pi j/T$, are approximately independent. Problems involving stationary mixing processes may thus be converted into ones involving (approximately) independent normal random variables. Empirical Fourier transforms such as Eq. (4.2) have many uses and several are indicated in this paper. A fundamental use is to estimate a power spectrum by smoothing the squared-modulus.

Early work on the asymptotic properties of finite Fourier transforms includes that of Slutsky (1934), Leonov and Shiryaev (1960), Rosenblatt (1961), Good (1963), Hannan (1969), Brillinger (1969), Hannan and Thomson (1971), Hannan (1972).

There has been some consideration of the cases of long range dependence and stable distributions. References include: Rosenblatt (1981), Freedman and Lane (1981), Fox and Taqqu (1986), Yajima (1989), Shao and Nikias (1993). The case of random generalized functions, which includes for example point processes and random measures, is considered in Brillinger (1982).

In the case of wavelets and a model

$$Y_t = g(t/T) + \varepsilon_t \tag{4.4}$$

with ε_t stationary noise having power spectrum $f(\lambda)$, under regularity conditions, the statistic $\hat{\beta}_{jk}$ of Eq. (3.13) may be shown to be asymptotically normal with mean β_{jk} and variance

$$\frac{2\pi}{T} f(0)$$

see Brillinger (1996). The variance is the same as that of Eq. (4.4). Further when the functions $\psi_{jk}(\cdot)$ and $\psi_{j'k'}(\cdot)$ are orthogonal, the coefficients $\hat{\beta}_{jk}, \hat{\beta}_{j'k'}$ are approximately independent for distinct (j, k) and (j', k'). This last results suggests that an estimate of $f(0)$ may be obtained by averaging the values $T|\hat{\beta}_{jk}|^2/T$ for which $\beta_{jk} = 0$.

4.3 Shrinking

Among surprises in working with Fourier transforms is the importance of convergence factors. These are the w_k^K of Eq. (3.4). In Eq. (3.4) they shrink the coefficients of the $\exp\{ixk\}$ towards 0 as k increases. Such multipliers are also important in the stochastic case, see: Tukey (1959), Brillinger (1975), Bloomfield (1976), Dahlhaus (1984, 1989).

A related concept is shrinking. In a regression context Tukey (1979) distinguishes three types of shrinking. Crudely: "first shrinkage" corresponds to pretesting and selection of regressor variables, "second shrinkage" corresponds to a type of Wiener filtering and "third shrinkage" corresponds to borrowing strength from other coefficients to improve the collection of coefficients. In this last case the multipliers are not meant for attenuating high frequencies, rather they are meant for attenuating uncertain terms. A common characteristic is that the estimates become biased; however, biased estimates have long been dominant in time series analysis.

Second shrinkage plays an important role in two of the examples that follow. A particular second shrinkage estimate, introduced in Tukey (1979), may be motivated as follows. Consider a classic simple regression model

$$y = \beta x + \varepsilon$$

with b an estimate of β and s an estimate of its standard error. Seek a multiplier m such that mbx is an improved estimate of βx. The mean-squared error of the new estimate is

$$x^2 E\{(\beta - mb)^2\}$$

which may be estimated by

$$x^2\{(1 - m)^2[b^2 - s^2] + m^2 s^2\}$$

This is minimized by the choice $m = 1 - s^2/b^2$. One would prefer to take m to be the positive part

$$(1 - s^2/b^2)_+ \tag{4.5}$$

This multiplier has the reasonable property of being 0 for b less than its standard error.

In Sec. 3.1 convergence factors, w_k^K, were inserted into trigonometric expressions to obtain improved approximation. In Example 5.1 such multipliers based on the reliability of estimated coefficients $\hat{c}_k$ will be inserted to obtain an improved estimate. To estimate $\hat{g}(x)$ of Eq. (4.4) one considers, for example,

$$\hat{g}(x) = \sum_k w(\hat{c}_k/s_k)\hat{c}_k e^{ixk} \tag{4.6}$$

where s_k^2 is an estimate of the variance of $\hat{c}_k$ and $w(u)$ is a function that is near 1 for large u and near 0 for small u. Examples of functions $w(\cdot)$ are given in Fig. 1.

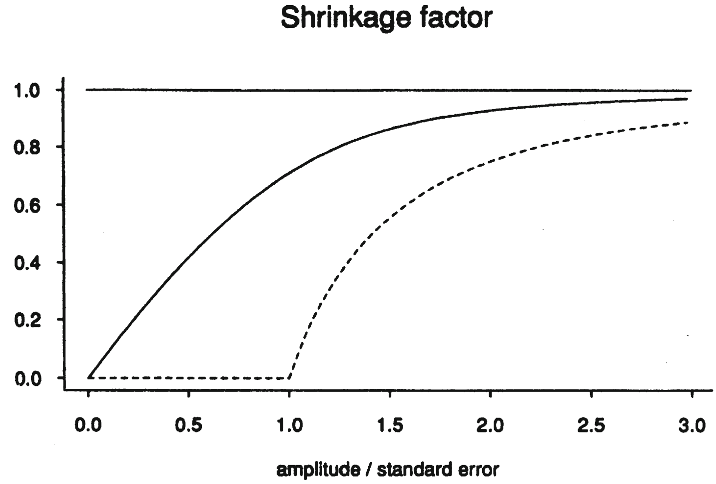


Figure 1. Graph of the multipliers Eqns (5.7) and Eq. (5.8), as a functions of the amplitude of the estimate divided by its estimated standard error.

In work to obtain improved wavelet-based estimates, Donoho and Johnstone (1990), Hall and Patil (1995) create shrinkage estimates involving multipliers, there referred to as "thresholders". The estimates take the form

$$\sum_{|j|\leq J} \sum_{|k|\leq K} w(|\hat{\beta}_{jk}|/s_{jk})\hat{\beta}_{jk}\psi_{jk}(x) \tag{4.7}$$

where s_{jk} is an estimate of var $\hat{\beta}_{jk}$ and $0 \leq w(\cdot) \leq 1$.

There are many classical references to selection of variables and pretesting, i.e., first shrinkage. References to second shrinkage include: Whittle (1962), Thompson (1968), King (1972), Ott and Kronmall (1976), Tukey (1979), Zidek (1983), Donoho and Johnstone (1990), Stoffer (1991), Hall and Patil (1993), Donoho et al. (1995). References to third shrinkage include: Stein (1955), Efron and Morris (1977), Copas (1983), Saleh (1992).

5. EXAMPLES

In this Section four biological and physical examples are presented.

5.1 Electron Microscopy

Electron microscopy is a tool for studying the placements of atoms within molecules. It is mainly carried out with crystalline (periodic) material. One problem is to obtain improved images and that is the concern of the present example. Glaeser (1985), Henderson et al. (1986), Hovmöller (1990) are references describing the basics of electron microscopy.

In the planar case, the principal theoretical concept is the projected (Coulomb) density distribution

$$V(x,y) = \sum_{h,k} F_{h,k} e^{2\pi i(hx+ky)/\Delta} \tag{5.1}$$

$h, k = 0, \pm 1, \pm 2, \ldots$ with (x, y) planar coordinates and with Δ the period of the crystal. The function $V(\cdot)$ is real-valued and has various symmetries. The h, k in Eq. (5.1) are referred to as the Miller indices, while the $F_{h,k}$ are referred to as structure factors. One wishes to estimate $V(x, y)$ over $0 \leq x, y < \Delta$.

The datum is an image, $Y(x, y)$, with $0 \leq x < X, 0 \leq y < Y$. The image may be written as

$$Y(x,y) = V(x,y) + \text{noise} \tag{5.2}$$

The empirical Fourier transform is

$$\hat{F}_{h,k} = \int_0^Y \int_0^X Y(x,y) e^{-2\pi i(hx+ky)/\Delta} dx\, dy \tag{5.3}$$

which may be written

$$\int_0^\Delta \int_0^\Delta \sum_{m,n} Y(x + m\Delta, y + n\Delta) e^{-2\pi i(hx+ky)/\Delta} dx\, dy \tag{5.4}$$

The synthesis corresponding to the analysis Eq. (5.3) is

$$\sum_{h,k} \hat{F}_{h,k} e^{2\pi i(hx+ky)/\Delta} \tag{5.5}$$

$0 \leq x < \Delta,\ 0 \leq y < \Delta$.

There has been concern to form an improved image. In this connection Blow and Crick (1959), Hayward and Stroud (1981) introduced "multipliers", $w(\cdot)$, into expressions like Eq. (5.5), forming

$$\hat{V}(x,y) = \sum_{h,k} w(|\hat{F}_{h,k}| \hat{\sigma}_{h,k}) \hat{F}_{h,k} e^{2\pi i(hx+ky)/\Delta} \tag{5.6}$$

where the $\hat{\sigma}_{h,k}$ are estimates of the standard errors of the $\hat{F}_{h,k}$. This is a second shrinkage estimate. Consideration of the mean-squared error, as in Eq. (4.5), leads to the multiplier

$$w(|\hat{F}|/\hat{\sigma}) = \left(1 - \frac{\hat{\sigma}^2}{|\hat{F}|^2}\right)_+ \tag{5.7}$$

which by analogy with Wiener filtering will be called the Wiener multiplier. By Bayesian arguments Blow and Crick (1959) and Hayward and Stroud (1981) were lead to the multiplier

$$w(\gamma) = \frac{\sqrt{\pi}}{2}\gamma\left[I_0\left(\frac{\gamma^2}{2}\right) + I_1\left(\frac{\gamma^2}{2}\right)\right]e^{-\gamma^2/2} \tag{5.8}$$

with $\gamma = |\hat{F}|/\hat{\sigma}$, and I_0, I_1 modified Bessel functions, see Brillinger et al. (1989, 1990). It and Eq. (5.7) are graphed in Fig. 1. These multipliers approach 1 as the uncertainty approaches 0.

Estimates employing Eqns (5.7) and (5.8) are illustrated in Fig. 2 for images of the protein bacteriorhodopsin. This substance occurs naturally

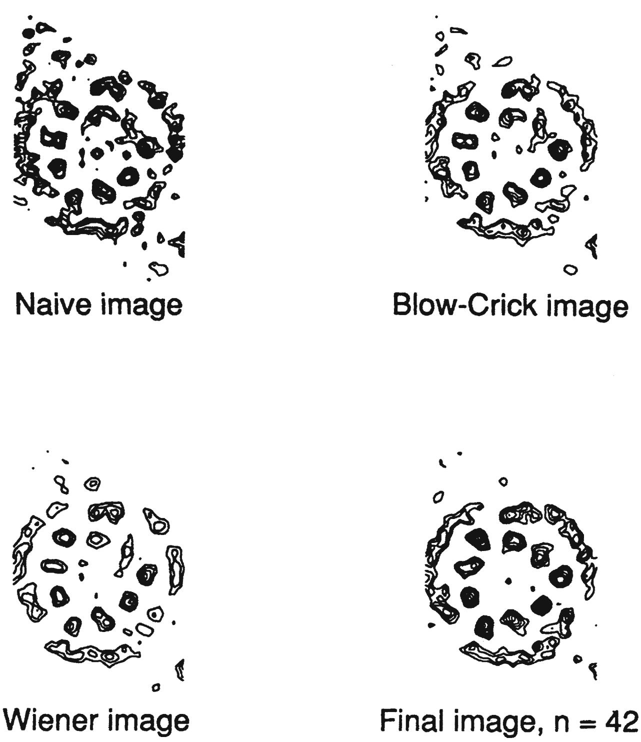


Figure 2. Estimates of the basic cell of bacteriorhodopsin. The upper left panel is the naive estimate as shown in Eq. (5.5). The upper right panel is the estimate Eq. (5.6) with the multiplier, Eq. (5.8). The bottom left panel is the estimate Eq. (5.6) with the multiplier, Eq. (5.7). The last panel is Eq. (5.6), with Eq. (5.8), obtained by combining 42 individual images.

as a two-dimensional crystalline array within the cell membrane of *Halobacterium halobrium*. Together with accompanying lipid molecules, it is known as "purple membrane". This crystal is based on a hexagonal lattice. In Fig. 2 only the positive contours are shown. (Negative density features signify the absence of atoms and thus have no direct usefulness when the density map is interpreted.) The first panel of Fig. 2 shows the elementary estimate of Eq. (5.5). The top right shows Eq. (5.6) with $w(\cdot)$ of Eq. (5.7). The third, lower left, shows Eq. (5.6) with $w(\cdot)$ of (5.8). The final panel provides an estimate based on combining 42 individual images. This last image may be viewed as what the earlier estimates based on a single image ascribe to be.

Through the inclusion of the multipliers, the peaks have become more substantial and better separated. Also, the estimates show better approximations to a three-fold symmetry. Details of the data collection and further details of the analysis may be found in Brillinger et al. (1989, 1990).

The Fourier transform is useful in this example firstly because of the lattice periodicities and secondly for the central limit theorem result suggesting specific estimates of the s_{hk} of Eq. (5.6) namely for s_{hk}^2 one takes the average of the squared moduli of Fourier coefficients thought to be signal free.

There are extensions to the 3D case, see Henderson et al. (1990), Wenk et al. (1992).

5.2 Seismic Surface Waves

Various sound waves are transmitted through the Earth following a seismic disturbance, in particular surface (or Rayleigh) waves. These are vibrations whose energy is trapped and propagated just under the surface. The waves have sinusoidal form and are prominent in the later part of a seismogram. For example see Fig. 3 for an event that was recorded in Uppsala, Sweden. These waves have the interesting aspect of having been discovered mathematically. For basic details see Aki and Richards (1980) and Bullen and Bolt (1985).

Consider modelling that part of a seismogram where the Rayleigh waves occur. Let $Y(x,t)$ denote the vibrations recorded at distance x from the earthquake source, as a function of time t. With a layered crust model the theoretical seismogram is a solution of a system of differential equations with associated boundary conditions and may be represented as

$$\int e^{-i(\lambda t - k(\lambda)x)} R(\lambda)\, d\lambda$$

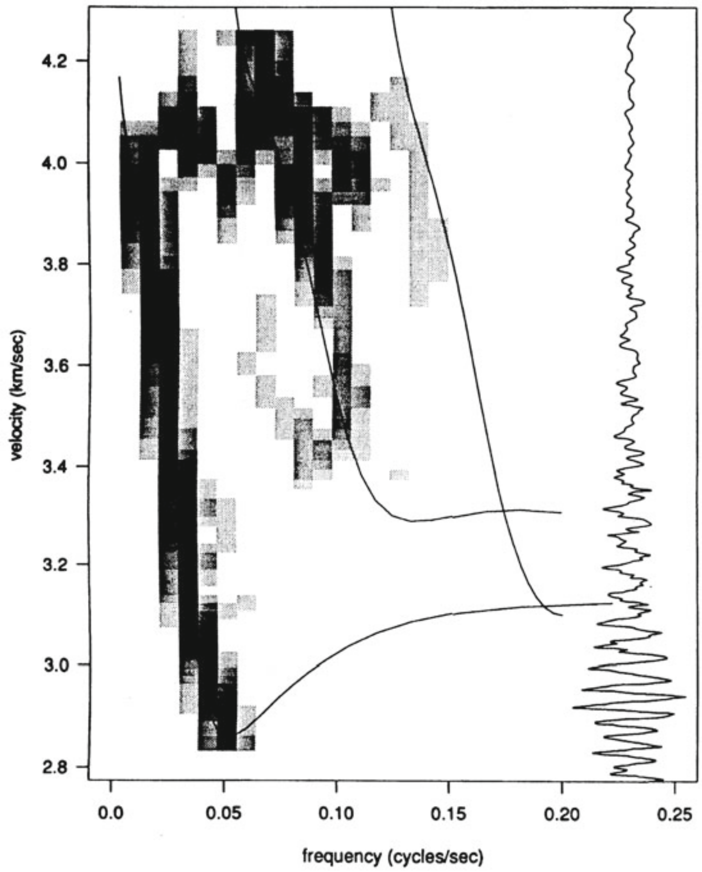


Figure 3. The Siberia–Upsalla dynamic spectrum as a function of frequency and velocity as computed from Eq. (5.11). The vertical trace is the seismogram as a function of velocity.

Here, when $x = 0$

$$\int e^{-i\lambda t} R(\lambda)\, d\lambda$$

represents the vibrations at the earthquake source. The solution in Eq. (5.9) comes from substituting a particular solution $\exp\{-i(\lambda t - kx)\}$ into the differential equations and matching boundary conditions, see Aki and Richards (1980). One writes $k(\lambda) = \lambda/c(\lambda)$ with $c(\lambda)$ the (phase) velocity with which the wave of frequency λ travels. The functions $k(\cdot)$ and $c(\cdot)$ depend on the transmission medium.

In the case that x is large one can use the method-of-stationary phase, described in Section 3.1, to see the sinusoidal form of the waves. Specifically for large x, Eq. (5.9) is approximately

$$R(\lambda_t)\exp\{-i(\lambda_t t - k(\lambda_t)x)\} \tag{5.10}$$

with λ_t the solution of

$$\frac{d}{d\lambda}\{\lambda t - k(\lambda)x\} = 0$$

that is $k'(\lambda_t) = t/x = 1/U(\lambda_t)$. Here $U(\lambda)$ is the group velocity, the velocity with which the energy travels, at frequency λ. The phenomenon of waves with different frequencies travelling with different velocities, as occurs here, is called dispersion.

Given an earth model, θ, that is a collection of layer depth, velocity and density parameters, one can compute the group velocity $U(\lambda|\theta)$, see Bolt and Butcher (1960), Aki and Richards (1980). For frequency λ and parameter θ there may be several possible dispersion curves $U_n(\lambda|\theta), n = 0, 1, 2, \ldots$ called modes. Dynamic Fourier analysis provides a way to see these modes, and is presented in Fig. 3. The concern of the example of this section is to estimate θ.

The event studied originated in Siberia, 20 April 1989, and the trace was recorded at Uppsala, Sweden. Figure 3 provides a grey scale display of energy as a function of velocity and frequency. It is computed as

$$\left|\sum_{s=-S}^{S} h(s/S)\,Y(t-s)e^{-is\lambda}\right|^2 \tag{5.11}$$

with $t = x_0/v$, v velocity, x_0 distance to source and $h(\cdot)$ a convergence factor. One sees waves of about 0.07 cycles/second arriving first. Figure 3 also shows the dispersion curves $U_n(\lambda|\hat{\theta})$ for one fitted earth model. Some further details are given in Brillinger (1993).

The velocity-frequency curves, $U_n(\lambda|\theta)$, may be inverted to frequency-time curves $\lambda = \lambda_t(t|\theta)$. To estimate θ one can then consider choosing θ, α to minimize

$$\sum_t \left|Y(t) - \int e^{-i(\lambda t - k(\lambda|\theta)x_0)} R(\lambda|\alpha)\, d\lambda\right|^2$$

where α is some parametrization of the source motion. One approach is to approximating the integral in Eq. (5.9), is to take $R(\cdot)$ piecewise constant, linear in α. Figure 4 provides the results of such an analysis. Graphed are

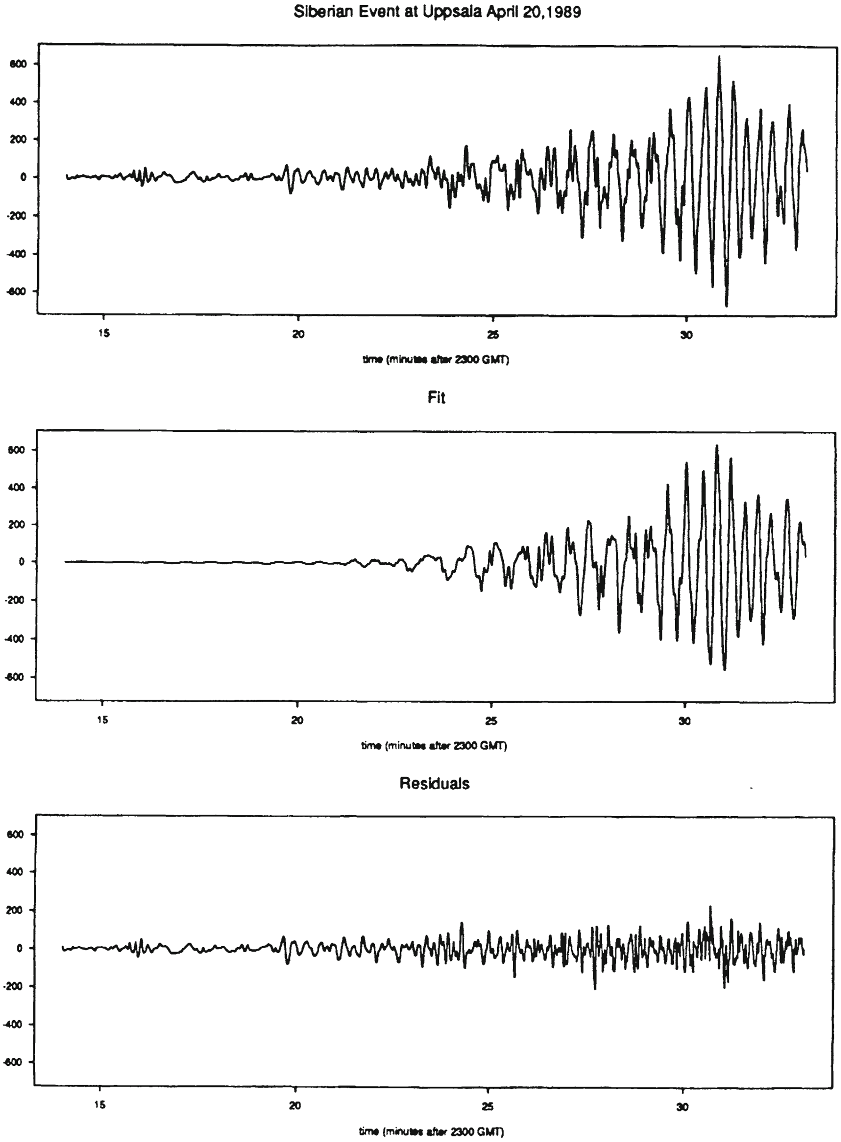


Figure 4. The top trace is the seismogram as a function of time. The middle is the fit based on Eq. (5.9). The bottom is the difference of these two.

the series, the fit and the residuals. The standard errors might be computed as in Richards (1961), focussing on the nonlinear parameters θ and acting as if the noise series was white. An improved estimation procedure is needed, for the residual series of Fig. 4 suggests the presence of signal-generated noise.

Even though this particular situation is clearly nonstationary, Fourier analysis has been basic to addressing it. This is a consequence of the presence of dispersion. The example is also of additional interest since one has a Fourier transform of two variables whose support lies on several curves, see Fig. 3. This type of plot allows inference of the presence of higher modes and assessment of the fit as well.

5.3 NMR Spectroscopy

Nuclear magnetic resonance is a quantum mechanical phenomenon employed to study the structure of various molecules. In an experiment, one creates a fluctuating magnetic field, $X(t)$, encompasing a substance and then observes an induced voltage, $Y(t)$. Hennel and Klinowski (1993) is one general reference.

If $\mathbf{S}(t)$ is a vector describing the state of the system at time t, then the fluctuations are described by the Bloch equations

$$\frac{d\mathbf{S}(t)}{dt} = \mathbf{a} + \mathbf{A}\mathbf{S}(t) + \mathbf{B}\mathbf{S}(t)X(t) \tag{5.12}$$

and the measurements by

$$Y(t) = \mathbf{c}^T\mathbf{S}(t) + \text{noise} \tag{5.13}$$

with $\mathbf{c}$ depending on the geometry of the experiment. The principal parameters are frequencies of oscillation and decay rates. The parameters of interest sit in the matrices $\mathbf{A}$ and $\mathbf{B}$, see Brillinger and Kaiser (1992). The entries of $\mathbf{A}$ and $\mathbf{B}$ have physical interpretations, e.g., the diagonal entries of $\mathbf{A}$ represent occupancy probabilities.

Equations (5.12) is interesting for being bilinear. It can be solved, symbolically, by successive substitutions, obtaining

$$\mathbf{S}(t) = \mathbf{C} + \int^t e^{\mathbf{A}(t-s)}\mathbf{C}X(s)\,ds + \int^t\int^s e^{\mathbf{A}(t-s)}\mathbf{B}e^{A(s-r)}\mathbf{C}X(r)X(s)\,dr\,ds + \cdots$$

with $\mathbf{C} = -\mathbf{A}^{-1}\mathbf{a}$. If $\mathbf{A}$ is written $\mathbf{U}e^{\Lambda}\mathbf{U}^{-1}$ with Λ diagonal, then the pulse response, $\mathbf{S}(t)$, is seen to be a sum of complex exponentials and various of their powers and products. The real parts of the entries of Λ will lead to the decay of these components while the imaginary parts represent resonance frequencies.

The problem is to estimate the parameters of Eq. (5.12) and thereby, to characterize the substance. Some of the parameters may be estimated by cross-spectral analysis and others by likelihood analysis.

Brillinger and Kaiser (1992) present results from a study of 2,3-dibromothiophene. The matrices **A** and **B** are 4×4 with complex-valued entries. The parameters include a coupling constant, J and frequencies ω_A and ω_B. In the experiment the input employed was a sequence of pulses

$$X(t) = \sum_j M_j \delta(t - j\Delta)$$

with $\Delta = 1/150\,\mathrm{s}$, t in seconds and M_j the m-sequence given by $M_j = M_{j-1}M_{j-4}M_{j-8}M_{j-12}$ starting at $M_j = -1$ for $j = 1, \ldots, 12$.

Figure 5 presents corresponding stretches of input and output together with the results of a cross-spectral analysis. Specifically the first-order transfer function estimate

$$\hat{A}(\lambda) = \frac{\underset{\mu\approx\lambda}{\mathrm{smooth}}\{[\sum_t Y(t)e^{-i\mu t}]\overline{[\sum_t X(t)e^{-i\mu t}]}\}}{\underset{\mu\approx\lambda}{\mathrm{smooth}}\{|\sum_t X(t)e^{-i\mu t}|^2\}} = \hat{f}_{YX}(\lambda)\hat{f}_{XX}(\lambda)^{-1}$$

is given in Fig. 5. Theoretically its peaks are located at the frequencies

$$(\omega_A + \omega_B)/2 \pm J \pm \sqrt{J^2 + (\omega_A - \omega_B)^2/2}$$

and the widths of the peaks relate to a damping constant T_2.

In a more detailed analysis the parameters of the model, including initial state values, were estimated by least squares seeking

$$\min_\theta \sum_t |Y(t) - \mathbf{c}^T\mathbf{S}(t|\theta)|^2 \qquad (5.14)$$

θ referring to the unknown parameters. In the computations the state vector, $\mathbf{S}(t|\theta)$ was evaluated recursively. Figure 6 shows the amplitude of the Fourier transform of the data and of the corresponding fit. (It is usual to graph an unsmoothed estimate in the NMR literature in order to obtain high resolution of peaks.) There is an intriguing small peak just above 60 Hz which recurs when the time series is broken down into contiguous segments. NMR researchers refer to such a phenomenon as a "birdie", but had no explanation for its source in the present case. Further details may be found in Brillinger and Kaiser (1992).

There are extensions of the cross-spectral approach to the 2, 3, 4, ... and higher dimensional cases, see Blümich (1985).

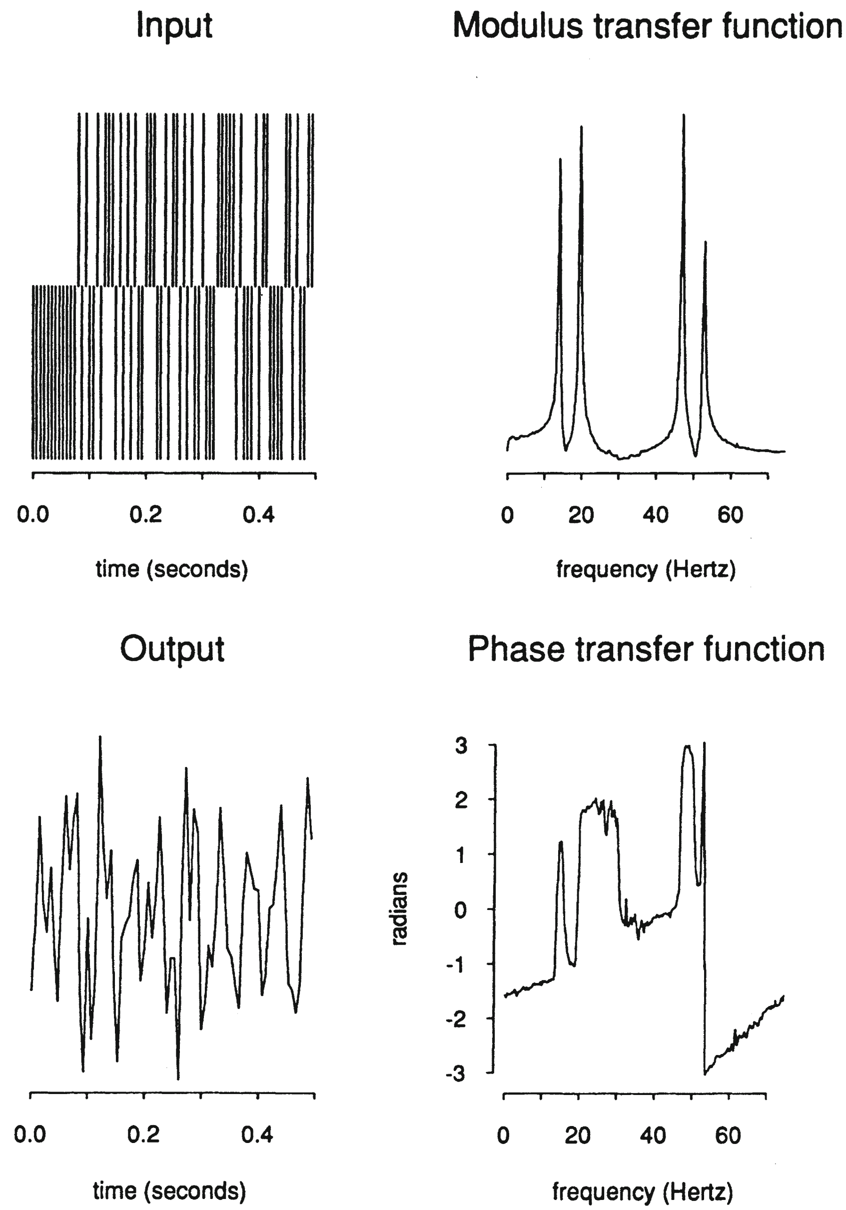


Figure 5. Results of a nuclear magnetic resonance study of 2,3-dibromothiophene. The top left is a segment of the input and below is the corresponding output. The right column provides the estimated amplitude and phase of the (linear) transfer function.

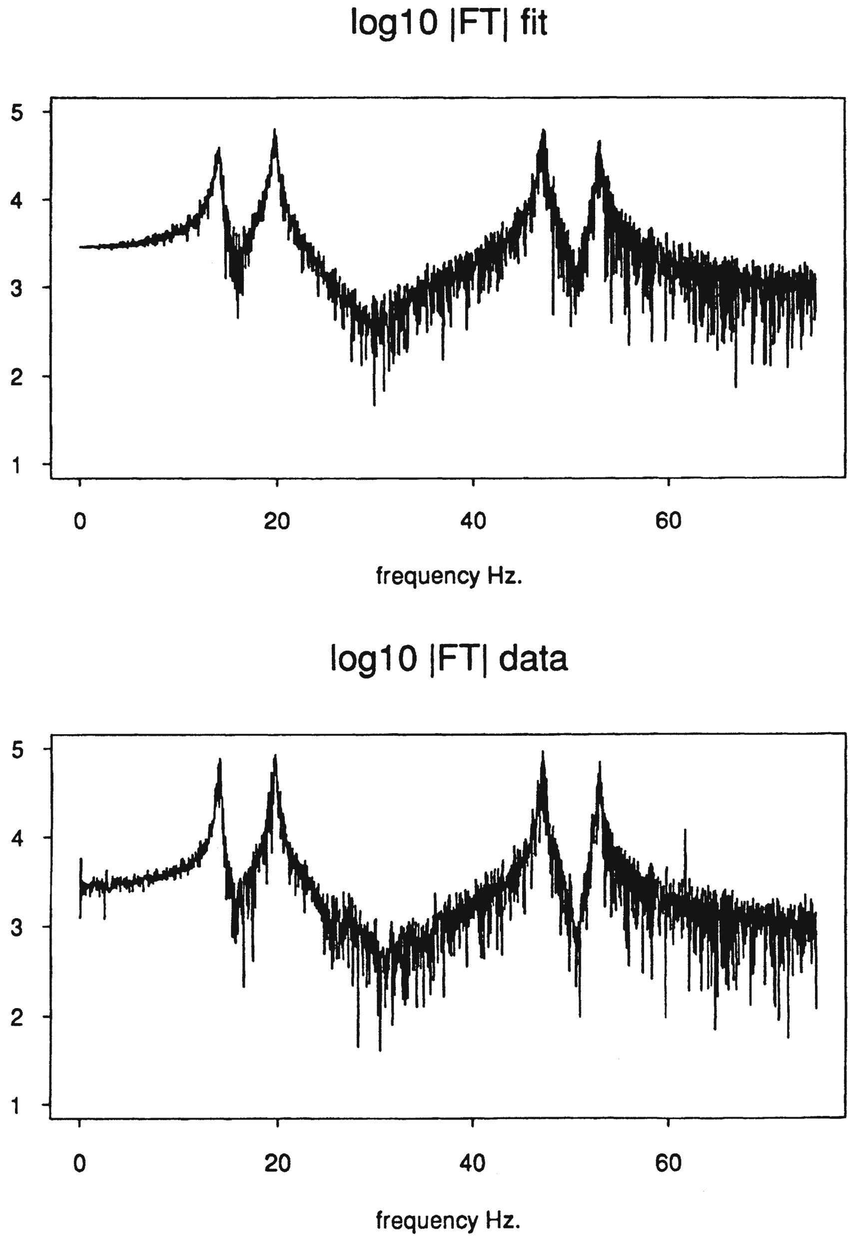


Figure 6. The modulus of the Fourier transform of the output and of the corresponding fit derived from Eq. (5.14).

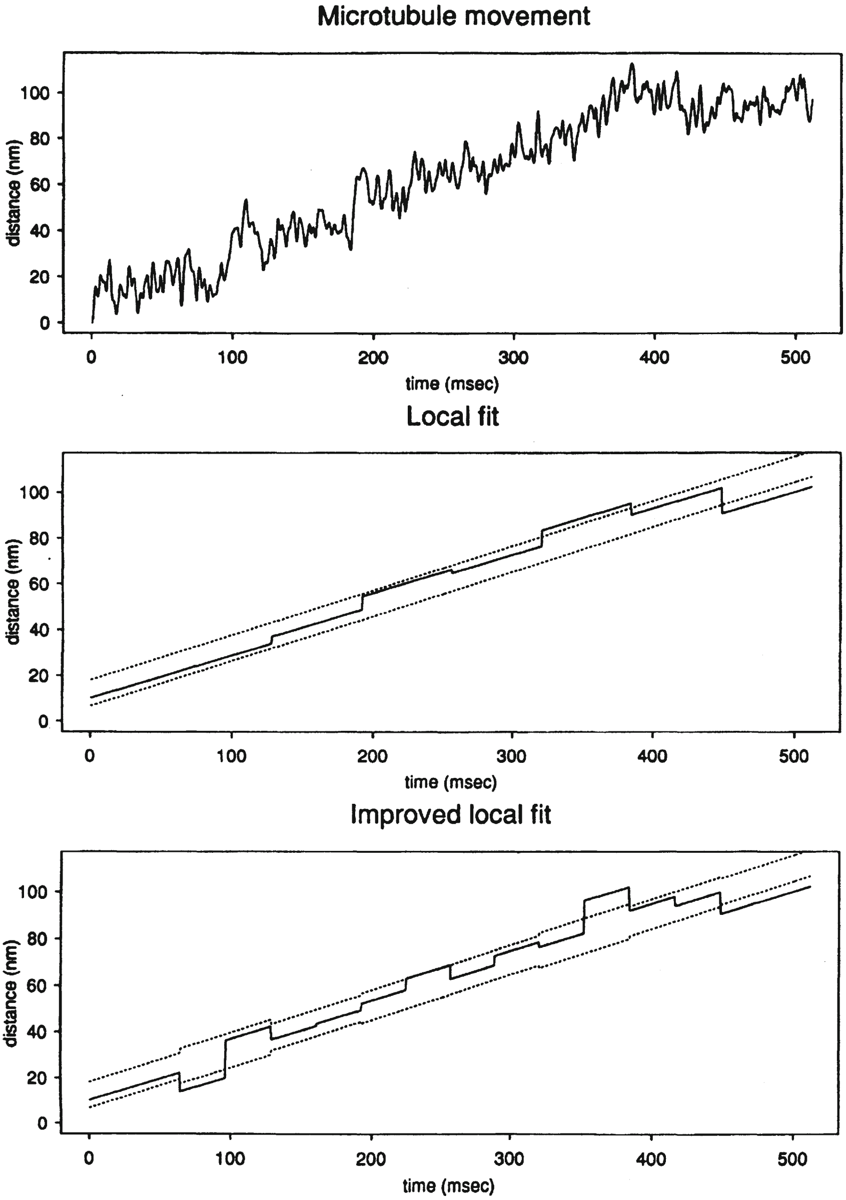


Figure 7. The top trace is the estimated movement of a microtubule as a function of time. The middle provides the fit with no shrinkage. The bottom panel provides a shrunken fit. The dashed lines provide approximate ±2 standard error limits.

In this example the Fourier transform is useful for examining resonance, for assessing goodness of fit and for understanding the nonlinearity involved.

5.4 Microtubule Movement

As an illustration of wavelet analysis, consider the problem of searching for jumps in records of microtubule movement. Microtubules are linear polymers basic to cell motility. A concern is whether movement is smooth, or rather via a series of jumps, see Malik et al. (1994).

If $Y(t)$ denotes the distance a microtubule has travelled at time t, the model considered is

$$Y(t) = \alpha t + g(t/T) + \text{noise} \tag{5.15}$$

$t = 0, \ldots, T-1$ with α a parameter related to diffusion motion and $g(\cdot)$ a step function. The model in Eq. (5.15) will be approximated by

$$Y(t) = \alpha t + \sum_k \gamma_{nk}\phi_{nk}(t/T) + \text{noise} \tag{5.16}$$

for some n. Because of the presence of the term αt in Eq. (5.16) the analysis in the present case is not so immediate, but still all that one needs are local means. The least squares estimates are obtained by regression of Y on the $\phi_{nk}(t/T)$ and on t made orthogonal to the ϕ_{nk}. Further details on the fitting are given in the Appendix to this chapter.

In the experiments of concern samples were taken from the bovine brain. Specifics may be found in Malik et al. (1994). The top panel of Fig. 7 provides a data trace. Next is an estimate $\hat{g}_n(t/T)$ with $w(u)$, of Eq. (4.7), identically 1. The final panel an improved estimate based on the multiplier $w(u) = (1 - 1/u^2)_+$. The value of $n = 3$ was chosen having in mind a search for isolated jumps for this particular data set. Also indicated are approximate ± 2 standard error limits around the fitted straight line. There is little evidence for the presence of isolated jumps. The construction of the standard error estimate is described in the Appendix to this chapter.

The Fourier transform was used here to develop uncertainty estimates, following on an assumption that the noise was stationary.

6. SOME OPEN PROBLEMS

This Section briefly lists a number of topics, motivated by the examples of the paper, that appear fruitful for more development.

Foremost among the topics calling out for further research is the theoretical and practical development of shrinkage estimates. The ideas are basic.

The effects are substantial, see Fig. 2 for example. One wonders about "optimal" choice of the multipliers/shrinkage factors. Perhaps optimal rates of convergence may be determined and then it be checked which multipliers lead to those. This paper has focused on second shrinkage. Berger and Wolpert (1983) develop third shrinkage estimates in random function cases. Lillestol (1977) studies time series in one case.

In both the surface wave and nuclear magnetic resonance examples, examination of residuals suggests the presence of signal-generated noise. Better estimates are needed. Either because the ones used are inefficient or because the signal-generated noise is basic. In the latter case an appropriate likelihood function needs to be developed. Ihaka (1993) does this for one case in seismology. If the noise is indeed nonstationary and autocorrelated, then a novel form of uncertainty estimation technique will be needed. In the case of the "improved" wavelet estimate, the uncertainty was estimated as if the shrinkage factors were constant, see Appendix to this chapter. Perhaps a useful bootstrap procedure could be developed, based on an assumption of stationary innovations being present.

Quite a different type of problem is the following: develop the aliasing structure for higher-order spectra in the case of a process observed on a lattice. This will be particularly complicated in the case of lattices in R^p with $p > 1$. Another problem in the case of image estimates, is how to visualize the associated uncertainty.

The Fourier transforms studied have all been scalar-valued. There are central limit theorems for processes taking on values in a group. It would be of interest to obtain corresponding results for group-valued Fourier transforms, e.g., in the p-adic case.

7. DISCUSSION AND SUMMARY

The principal interest of the examples of the paper has been in problem formulation and in addressing particular scientific questions. In each of the examples, an empirical Fourier transform has played a central role. With its broad collection of understood properties this transform has assisted the analyses greatly. The usefulness of second shrinkage, analogous to the use of convergence factors in Fourier approximation, is also noteworthy.

The particular groups of the examples have been abelian. General group theoretic ideas and empirical Fourier analysis have been discussed for other groups. For the case of the symmetric group see Diaconnis (1988, 1989) and Kim and Chapman (1993). For the locally compact abelian case see Brillinger (1982). For p-adics see Brillinger (1992). The use of p-adics in signal processing is discussed in Gorgui-Naguib (1990). For other cases see Hannan (1969). Key distinctions that arise are abelian versus nonabelian groups, compact versus locally compact groups, and whether t is in a group or Y is in a group.

There are other transforms that are useful in practice. These include: the Laplace, Hilbert, Stieltjes, Mellin, with some work having been done for abstract groups, see Loomis (1953).

The case of lacunary trigonometric series is somewhat like the case of point processes. Here the Fourier transform has a different form, e.g., for point process data $\{\tau_1 < \tau_2 < \cdots < \tau_N\}$ it is given by

$$\sum_{j=1}^{N} \exp\{-i\lambda\tau_j\}$$

$-\infty < \lambda < \infty$. Such a transform is used in Rosenberg et al. (1989) for example.

Unemphasized, but important, topics include: the Poisson summation formula useful in understanding aliasing and the sampling theorem (Hannan (1965)), abstract fast algorithms (Rockmore (1990)), spherical functions (Terras (1988)), uncertainty principles (Smith (1990)).

In conclusion we quote J. B. Fourier (1822), *Théorie Analytique de la Chaleur*: "L" étude approfondie de la nature est la source la plus féconde des découvertes mathématiques." There is so much evidence in favor of this remark today.

Acknowledgments

The author thanks his collaborators in the examples of Sec. 5: Ken Downing, Bob Glaeser, Guy Perkins, Bruce Bolt, Reinhold Kaiser, Fady Malik and Ron Vail respectively. This paper is based on talks presented at the 1993 opening ceremonies of the Fields Institute for Research in Mathematical Sciences at the University of Waterloo, Ontario, Canada and at the 1993 Winter Meeting of the Canadian Mathematical Society at Ottawa. The work was partially supported by the NSF Grants DMS-9208683, DMS-9300002, DMS-9625774 and the Office of Naval Research Grant N00014-94-1-0042. Comments of Bob Kass and the referee helped improve the layout and style of the paper.

REFERENCES

Aki, K. and P. G. Richards, *Quantitative Seismology I & II.* Freeman, San Francisco, 1980.

Amato, I., Nobel prizes 1991 *Science 254*: 518–519, 1991.

Barndorff-Nielsen, O. E. and D. R. Cox, *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London, 1989.

Båth, M., *Spectral Analysis in Geophysics*. Elsevier, Amsterdam, 1994.

Bazin, M. J., P. H. Lucie, and S. Moss de Olivierra, Experimental demonstrations of the mathematical properties of Fourier transforms using diffraction phenomena, *Eur. J. Phys. 7*: 183–188 (1986).

Benedetto, J. J. and M. W. Frazier, (eds), *Wavelets*, CRC Press, Boca Raton, 1994.

Berger, J. and R. Wolpert, Estimating the mean function of a Gaussian process and the Stein effect, *J. Mult. Anal. 13*: 401–424, (1983).

Blackman, R. B. and J. W. Tukey, *The Measurement of Power Spectra. Dover*, New York, 1959.

Bloembergen, N., Nonlinear optics and spectroscopy, *Science 215*: 1057–1064 (1982).

Bloomfield, P., *Fourier Analysis of Time Series: An Introduction*, Wiley, New York, 1976.

Blow, D. M. and F. H. C. Crick, The treatment of errors in the isomorphous replacement method, *Acta Cryst. 12*: 794–802 (1959).

Blümich, B., Stochastic nmr spectroscopy, *Bull. Magnet. Resonance 7*: 5–26 (1985).

Bochner, S., *Lectures on Fourier Integrals*, Princeton Press, Princeton, 1959.

Bochner, S., *Harmonic Analysis and the Theory of Probability*, Univ. Calif. Press, Berkeley, 1960.

Bolt, B. A. and J. Butcher, Rayleigh wave dispersion for a single layer on an elastic half space, *Australian J. Physics 13*: 498–504 (1960).

Bolt, B. A., Y. B. Tsai, K. Yeh and M. K. Hsu, Earthquake strong motions recorded at a large near-source array of digital seismographs, *Earthquake Eng. Structural Dynam. 10*: 561–573 (1982).

Born, M. and Wolf, E. (1964). Principles of Optics. Second Edition. Macmillan, New York.

Bracewell, R. N., The Fourier transform, *Scientific Ameri.*, June: 86–95 (1989).

Brillinger, D. R., A search for a relationship between monthly sunspot numbers and certain climatic series, *Bull. Inter. Statist. Inst. 43*: 293–306 (1969).

Brillinger, D. R., *Time Series: Data Analysis and Theory*, Holt, New York, 1975.

Brillinger, D. R., Asymptotic normality of finite Fourier transforms of stationary generalized processes, *J. Mult. Analysis 12*: 64–71 (1982).

Brillinger, D. R., Some asymptotics of finite Fourier transforms of a stationary p-adic process, *J. Comb. Inf. Sys. Sci. 16*: 155–169 (1991).

Brillinger, D. R., An application of statistics to seismology: dispersion and modes. In *Developments in Time Series Analysis* (ed. T. Subba Rao), Chapman and Hall, London, 1993, pp. 331–340.

Brillinger, D. R., Some uses of cumulants in wavelet analysis, *Nonpar. Statist. 6*: 93–114 (1996).

Brillinger, D. R., K. H. Downing and R. M. Glaeser, Some statistical aspects of low-dose electron imaging of crystals, *J. Stat. Planning Inf. 25*: 235–259 (1990).

Brillinger, D. R., K. H. Downing, R. M. Glaeser and G. Perkins, Combining noisy images of small crystalline domains in high resolution electron microscopy, *J. App. Stat. 16*: 165–175 (1989).

Brillinger, D. R. and R. Kaiser, Fourier and likelihood analysis in NMR spectroscopy. In *New Directions in Time Series I* (eds. D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. Taqqu), Springer, New York, 1992, pp. 41–64.

Bullen, K. E. and B. A. Bolt, *An Introduction to the Theory of Seismology*, Cambridge Univ. Press, Cambridge, 1985.

Butzer, P. L. and R. J. Nessel, *Fourier Analysis and Approximation*, Academic, New York, 1971.

Cartwright, D. E., Tidal analysis—a retrospect in *Time Series Methods in Hydrosciences* (eds A. H. El-Shaarawi and S. R. Esterby), Elsevier, Amsterdam, 1982, pp. 170–188.

Cooley, J. W. and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Comp. 19*: 297–301 (1965).

Copas, J. B., Regression, prediction and shrinkage, *J. Roy. Statist. Soc.* B 45: 311–335 (1983).

Cramér, H., On harmonic analysis in certain function spaces, *Arkiv Math. Astr. Fysik. 28*: 1–7 (1942).

Dahlhaus, R., Parameter estimation of stationary processes with spectra containing strong peaks in *Robust and Nonlinear Time Series Analysis* (eds J. Franke, W. Haerdle and D. Martin), Springer, New York, 1984, pp. 50–67.

Dahlhaus, R., Efficient parameter estimation for self-similar processes, *Ann. Statist. 17*: 1749–1766 (1989).

Daubechies, I., *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

Diaconnis, P., *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, 1988.

Diaconnis, P., A generalization of spectral analysis with applications to ranked data, *Ann. Statist. 17*: 949–979 (1989).

Donoho, D. L. and I. M. Johnstone, *Wavelets and optimal nonlinear function estimates*, Tech. Report 281, Statistics Dept. Univ. California, Berkeley, 1990.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian and D. Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc.* B *57*: 301-369 (1995).

Dzhaparidze, K., *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*, Springer, New York, 1986.

Efron, B. and C. Morris, Stein's paradox in statistics, *Scientific Amer. 236*: 119–127 (1977).

Fan, J., Deconvolution with supersmooth distributions, *Canadian J. Statist. 20*: 155–169 (1992).

Feuerverger, A., Efficiency in time series, *Canadian J. Statist. 18*: 155–162. (1990).

Feigin, P. D. and C. R. Heathcote, The empirical characteristic function and the Cramér-von Mises statistic, *Sankhya A 38*: 309–325 (1976).

Fine, N. J., On Walsh functions, *Trans. Amer. Math. Soc. 65*: 372–414 (1949).

Fox, R. and M. S. Taqqu, Large-sample properties of parameter estimates for strongly dependent stationary time series, *Ann. Statist. 14*: 517–532 (1986).

Freedman, D. and D. Lane, The empirical distribution of the Fourier coefficients of a sequence of independent, identically distributed long-tailed random variables, *Z. Wahrschein. Ver. Geb. 55*: 123–132 (1981).

Glaeser, R. M., Electron crystallography of biological macromolecules, *Ann. Rev. Phys. Chem. 36*: 243–275 (1985)

Good, I. J., The interaction algorithm and practical Fourier series, *J. Roy. Statist. Soc. B 20*: 361–372 (1958).

Good, I. J., Weighted covariance for detecting the direction of a Gaussian source in *Time Series Analysis* (ed. M. Rosenblatt), Wiley, New York, 1963, pp. 447–470.

Goodman, J. W., *Introduction to Fourier Optics*, McGraw-Hill, San Francisco, 1968.

Gorgui-Naguib, R. N., p-adic transforms in digital signal processing in *Mathematics in Signal Processing II* (ed. J. G. McWirter), Clarendon, Oxford, 1990, pp. 43–53.

Hall, P. and P. Patil, On wavelet methods for estimating smooth functions, *Bernoulli 1* (1995), pp. 41–58.

Halvorson, C., A. Hays, B. Kraabel, R. Wu, F. Wudl and A. J. Heeger, A 160-femtosecond optical image processor based on a conjugated polymer, *Science 265*: 1215–1216 (1994).

Hannan, E. J., *Aliasing*. Tech. Report 25, Statistics Dept., Johns Hopkins University, 1965.

Hannan, E. J., *Group Representations and Applied Probability*, Methuen, London, 1966.

Hannan, E. J., Fourier methods and random processes, *Bull. Internat. Statist. Inst. 42(1)*: 475–496 (1969).

Hannan, E. J., Spectra changing over narrow bands in *Statistical Models and Turbulence* (eds M. Rosenblatt and C. Van Atta), Springer, New York, 1972, pp. 460–469.

Hannan, E. J. and P. J. Thomson, Spectral inference over narrow bands, *J. Appl. Prob. 8*: 157–169 (1971).

Hayward, S. B. and R. M. Stroud, Projected purple membrane determined to 3.7 Å resolution by low temperature electron microscopy, *J. Molec. Biol. 151*: 491–517 (1981).

Henderson, R., J. M. Baldwin, K. H. Downing, J. Lepault, and F. Zemlin, Structure of purple membrane from *Halobacterium halobium*, *Ultramicroscopy 19*: 147–178 (1986).

Henderson, R., J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann and K. H. Downing, Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy, *J. Mol. Biol. 213*: 899–929 (1990).

Hennel, J. W. and J. Klinowski, *Fundamentals of Nuclear Magnetic Resonance*, Wiley, New York, 1993.

Hewitt, E. and K. A. Ross, *Abstract Harmonic Analysis I&II*, Academic, New York, 1963.

Higgins, J. R., Five short stories about the cardinal series, *Bull. Amer. Math. Soc. 12*: 45–89 (1985).

Hirsch, M. W., The dynamical systems approach to differential equations, *Bull. Amer. Math. Soc. 11*: 1–64 (1984).

Hochstadt, H., *Differential Equations, A Modern Approach*, Holt-Rinehart, New York, 1964.

Hovmoller, S., Structure analysis by crystallographic image processing—Hommage à Jean Baptiste Joseph Fourier (1768–1830), *Microsc. Microanal. Microstruct. 1*: 423–431 (1990).

Ihaka, R., Statistical aspects of earthquake source parameter estimation in the presence of signal generated noise, *Commun. Statist. A 22*: 1425–1440 (1993).

Katz, B. and R. Miledi, Further observations on acetylcholine noise, *Nature 232*: 124–126 (1971).

Katznelson, Y., *An Introduction to Harmonic Analysis*, Dover, New York, 1976.

Kim, P. and G. R. Chapman, Group action on a lattice and an application to time series, *J. Stat. Planning Inf. 34*: 183–195 (1993).

King, N., An alternative for the linear regression equation when the predictor variable is uncontrolled and the sample size is small, *J. Amer. Statist. Assoc. 67*: 217–219 (1972).

Korner, T. W., *Fourier Analysis*. Cambridge Univ. Press, Cambridge, 1989.

Lanczos, C., *Discourse on Fourier Series*, Hafner, New York, 1966.

Leonov, V. P. and A. N. Shiryaev, Some problems in the spectral theory of higher moments, *Theory Prob. Appl. 5*: 460–464 (1960).

Lillestol, J., Improved estimates for multivariate complex normal regression with application to analysis of linear time-invariant relationships. *J. Mult. Anal.* *7*: 512–524 (1977).

Loomis, L., *An Introduction to Abstract Harmonic Analysis*, Van Nostrand, New York, 1953.

Malik, F., D. R. Brillinger, and R. D. Vale, High resolution tracking of microtubule motility driven by a single kinesin motor, *Proc. Natl. Acad. Sci.* USA 91, pp. 4584–4588 (1994).

Michaelson, A. A., On the application of interference methods to spectroscopic methods—I. *Phil. Mag.* *33*: 338–346 (1891a).

Michaelson, A. A., On the application of interference methods to spectroscopic methods—II. *Phil. Mag.* *34*: 280–299 (1891b).

Moloney, J. V. and A. C. Newell, *Nonlinear optics*. Tech. Report 574. IMA, University of Minnesota, Minneapolis, 1989.

Ott, J. and R. A. Kronmall, Some classification procedures for multivariate binary data using orthogonal functions, *J. Amer. Statist. Assoc.* *71*: 391–399 (1976).

Picinbono, M. B., Tendance vers le caractère gaussien par filtrage sélectif, *Comptes Rendus Acad. Sci.*, 1174–1176 (1960).

Priestley, M. B., Evolutionary spectra and non-stationary processes, *J. Roy. Statist. Soc. B* *27*: 204–237 (1965).

Richards, F. S. G., A method of maximum-likelihood estimation, *J. Roy. Statist. Soc. B* *23*: 469–475 (1961).

Rockmore, D., Fast Fourier analysis for abelian group extensions, *Adv. Appl. Math.* *11*: 164–204 (1990).

Rosenberg, J. R., A. M. Amjad, P. Breeze, D. R. Brillinger and D. M. Haliday, The Fourier approach to the identification of functional coupling between neuronal spike trains, *Prog. Biophys. Molec. Biol.* *53*: 1–31 (1989).

Rosenblatt, M., Some comments on narrow band-pass filters, *Quart. Appl. Math.* *18*: 387–394 (1961).

Rosenblatt, M., Probability limit theorems and some questions in fluid mechanics in *Statistical Models and Turbulence* (eds M. Rosenblatt and C. Van Atta), Springer, New York, 1972, pp. 27–40.

Rosenblatt, M., Limit theorems for Fourier transforms of functionals of Gaussian sequences, *Z. Wahrsch. Verw. Gebiete. 55*: 123–132 (1981).

Rudin, W., *Fourier Analysis on Groups*, Wiley, New York, 1962.

Ruelle, D., *Chaotic Evolution and Strange Attractors*, Cambridge Univ. Press, Cambridge, 1989.

Saleh, A. K., *Contributions to Preliminary Test and Shrinkage Estimation,* Department of Mathematics and Statistics, Carleton University, Canada, 1992.

Shao, M. and C. L. Nikias, Signal processing with fractional lower order moments: stable processes and their applications, *Proc. IEEE 81*: 986–1010 (1993).

Shankar, P. M., S. N. Gupta, and H. M. Gupta, Applications of coherent optics and holography in biomedical engineering, *IEEE Trans. Biomed. Eng. BME-29*: 8–15 (1982).

Slutsky, E., Alcuni applicazioni di coefficienti di Fourier al analizo di sequenze eventuali coherenti stazionari, *Giorn. d. Instituto Italiano degli Atuari 5*: 435–482 (1934).

Smith, K. T., The uncertainty principle on groups, *SIAM J. Appl. Math. 50*: 876–882 (1990).

Stein, C., Inadmissibility of the usual estimator for the mean of a multivariate normal distribution in *Proc. Third Berk. Symp. Math. Statist. Prob. Vol. 1*, Univ. Calif. Press, Berkeley, 1955, pp. 197–206.

Stoffer, D. S., Walsh-Fourier analysis and its statistical applications, *J. Amer. Statist. Assoc. 86*: 481–482 (1991).

Strang, G., Wavelet transforms versus Fourier transforms, *Bull. Amer. Math. Soc. 28*: 288–305 (1993).

Strichartz, R. S., How to make wavelets, *Amer. Math. Monthly 100*: 539–556 (1990).

Tarter, M. E. and M. D. Lock, *Model-free Curve Estimation*, Chapman and Hall, New York, 1993.

Terras, A., *Harmonic Analysis on Symmetric Spaces and Applications I & II,* Springer, New York, 1988.

Thompson, J. R., Some shrinkage techniques for estimating the mean, *J. Amer. Statist. Assoc. 63*: 113–122 (1968).

Timan, A. F., *Theory of Approximation of Functions of a Real Variable*, Pergamon, Oxford, 1963.

Tukey, J. W., An introduction to the frequency analysis of time series in *The Collected Works of John W. Tukey I* (1984). (ed. D. R. Brillinger), Wadsworth, Pacific Grove, 1963, pp. 503–650.

Tukey, J. W., Equalization and pulse shaping techniques applied to determination of initial sense of Rayleigh waves in *The Collected Works of John W. Tukey I* (1984). (ed. D.R. Brillinger), Wadsworth, Pacific Grove, 1959, pp. 309–358.

Tukey, J. W., *Introduction to the dilemmas and difficulties of regression*. Unpublished, 1979.

Walter, G. G., Approximation of the delta function by wavelets, *J. Approx. Theory 71*: 329–343 (1992).

Walter, G. G. (1994). Wavelets and Other Orthogonal Systems with Applications. CRC Press, Boca Raton.

Wenk H. R., K. H. Downing, M. S. Hu and M. A. Okeefe, 3D structure determination from electron-microscope images—electron crystallography of staurolite, *Acta Crystallographica A1, 48*: 700–716 (1992).

Whittle, P., Estimation and information in time series analysis, *Skand. Aktuar. 35*: 48–60 (1952).

Whittle, P., Discussion of C. M. Stein "Confidence sets for the mean of a multivariate normal distribution", *J. Roy. Statist. Soc. B 24*: 294 (1962).

Wiener, N., *The Fourier Integral and Certain of Its Applications*, Dover, New York, 1933.

Yaglom, A. M., Second-order homogeneous random fields in *Proc. Fourth Berkeley Symp. Math. Statist. Prob. 2*, Univ. Calif. Press, Berkeley, 1961, pp. 593–622.

Yajima, Y., A central limit theorem of Fourier transforms of strongly dependent stationary processes, *J. Time Series Analysis 10*: 375–384 (1989).

Yariv, A., *Quantum Electronics*, Second Edition, Wiley, New York, 1975.

Zidek, J., Discussion of Copas (1983), Regression, prediction and shrinkage, pp. 347–48. *J. Roy. Statist. Soc. B 45*: 311–335 (1983).

Zygmund, A., *Trigonometric Series*, Cambridge Univ. Press, Cambridge, 1968.

APPENDIX

The estimate presented in the middle panel of Fig. 7 is ordinary least squares. (In many time series situations such estimates are asymptotically efficient.)

The model shown in Eq. (5.16) is linear in α and the γ_{nk}. It may be written

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\alpha + \varepsilon$$

taking $\mathbf{Z} = [t - \bar{t}]$ and $\mathbf{X} = [X_{tk}]$, with $X_{tk} = 1$ for $k/2^n \leq t/T < (k+1)/T$ and 0 otherwise. It is seen to have the form of an analysis of covariance model. The least squares estimates may be written

$$\hat{\alpha} = (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}\mathbf{y} \tag{A.1}$$

$$\hat{\gamma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\hat{\alpha}) \tag{A.2}$$

with $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Some Examples of Random Process Environmental Data Analysis

David R. Brillinger

1. Introduction

Data of process type are now routinely collected and analyzed in the environmental sciences. This is a consequence, in part, of today's general availabilty of sophisticated computing, storage, display and analysis equipment. At the same time stochastic models have been developed that take detailed note of the special characteristics of such data and hence allow more appropriate and efficient analyses to be carried through. The problems can be difficult, but often an aproach is suggested by basic scientific background and the parameters have physical interpretations. Recognizing a process type is an important step along the way to its analysis. The goal of this work is to bring out some basic ideas by presenting a number of elementary examples of random process data analysis.

The work proceeds by describing some basic types of stochastic processes and then presenting some techniques for addressing general problems arising. The emphasis is on processes, their characteristics and understanding their nature by descriptive statistics and elementary analyses rather than by developing background theory. By presenting examples, from different fields, and doing so in comparative fashion the intention is to bring out both similarities and differences. The examples have differing goals.

Concern will be with how the data might be presented visually and described analytically. The next section presents a few basic formal concepts. Section 3 is concerned with temporal point and marked point

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_17, © Springer Science+Business Media, LLC 2012

processes and an application to a risk assessment problem in space science. Section 4 is concerned with a count-valued time series relating to concerns with childbirth risk. Section 5 focuses on spatial-temporal processes with an example from neuroscience. Section 6 focuses on particle processes with an example from marine biology and the beginnings of an example from pest management. Finally there are some general remarks and discussion.

Techniques highlighted include: plotting the data, likelihood analysis, the EM method, generalized linear modeling, Fourier inference and state space modelling.

2. Some Basic Concepts and Methods

A classical and effective approach for addressing a broad variety of environmental problems is to view the data that have come to hand as part of a realization of a stochastic process. In simplest terms a *random process* is a family of random variables indexed by a label. In the present work the label will refer to time or space-time. The data of concern may be real-valued, vector-valued, categorical-valued, or generalized function-valued amongst other possibilities.

A random process may be described as a family of jointly distributed random variables. The values it takes on and the character of the index labelling the members of the family are what provide special features. For example a *temporal point process*, referring to the occurence times of some event of interest, might be described by providing the joint distributions of the count-valued random variables $N(A_1), ..., N(A_k)$ where $A_1, \cdots, A_k$, k *in* $Z = \{0, \pm 1, \pm 2, ...\}$, refer to any Borel subsets of $R = (-\infty, \infty)$ and $N(A)$ is the number of occurences of the event in the set

A. Of course the distributions must be consistent for the process to be well-defined. In the *stationary* case these distributions will be invariant under translations in time. Stationarity is basic to the definition of important parameters describing processes and to the derivation of the statistical properties of quantities computed from process data. In Section 6 an example of points distributed in both the plane and time is considered. Then the Borel subsets are contained in R^3.

Two general approaches will be made use of in the analyses presented. In the *Method of Moments* basic use is made of moments and cumulants to define parameters of importance and to develop properties of polynomial-type statistics based on data at hand. In *Likelihood Analysis* a serious attempt is made to set down a full model and thereby obtain efficient procedures. Bayesians would recomend multiplying the likelihood further by a prior distribution.

3. Point Processes and Marked Point Processes

3.1. Background

A *temporal point process* is a collection of occurrence times of events, $\{\tau_j\}$, supposed distinct and ordered by $\tau_j < \tau_{j+1}$, j *in* Z and τ_j *in* R.

Practically, it is often useful to describe a point process via its *conditional intensity* function. Among those introducing this approach into statistics were Cox and Lewis (1972), Rubin (1972) and Snyder (1975). To describe the conditional intensity of a temporal point process write

$$N(t) = \#\{\tau_j \text{ in } [0,t)\} = N[0,t)$$

and $H_t = \{\tau_j \; with \; \tau_j \le t\}$. This last is referred to as the *history* of the process up to time t. When it exists, the *conditional intensity*, $\mu(t \mid H_t)$, is given by

$$Prob\{dN(t) = 1 \mid H_t\} = \mu(t \mid H_t)dt$$

with the interpretation that $\mu(t \mid H_t)$ is the rate of occurence of events at time t given the history until and including then. With occurrence times $0 \le \tau_j < T$ and supposing the process distribution to depend on a parameter θ the *likelihood function* is given by

$$\prod_j \mu(\tau_j \mid H_t,\theta) \exp\{-\int_0^T \mu(t \mid H_t,\theta)dt\}$$

It may be used to make inferences concerning θ.

Important characteristics of a stationary point procss may sometimes be inferred from an estimate of its *autointensity* function

$$m(u) = Prob\{dN(t+u) = 1 \mid dN(t) = 0\}/dt \qquad (3.1)$$

giving the rate at which points occur at lag u after an existing point. It may be estimated directly, see Brillinger (1978).

A *marked temporal point process* is a sequence of pairs $\{(\tau_j, M_j)\}$, with $\tau_j \; in \; R$ refering to the j-th time and M_j an associated quantity (or mark) at that time. The likelihood function may be based on the probability element

$$Prob\{dN(t) = 1 \text{ and } m < M_{N(t)} < m+dm \mid H_t\} = \nu(t,dm \mid H_t)dt \qquad (3.2)$$

see Fishman and Snyder (1976). For example one then has

$$Prob\{no \; point \; in \; (t,t+u) \; with \; mark \ge m \mid H_t\} = \exp\{-\int_t^{t+u}\int_m^{\infty} \nu(s,dm \mid H_s)ds\}$$

(3.3)

Writing

$$o(s, dm \mid H_t) = E\{\nu(s, dm \mid H_t)\} \quad (3.4)$$

the probability (3.3) is bounded by

$$\int_t^{t+u} \int_m^{\infty} o(s, dm \mid H_t) ds \quad (3.5)$$

see Brillinger (1982).

A common question is whether the temporal and mark variations are statistically independent.

3.2. An Example From Space Science

Astronauts living and working in space are subject to a wide variety of risks of which an important one is that they, or their space craft, may be hit by orbiting debris. To assess this risk NASA sampled the population of orbiting objects, see Committee on Orbital Debris (1995). A narrow radar beam was used to detect and estimate characteristics of debris, data being collected over a number of observation periods when the Haystack telescope was avaiable.

Figure 1a displays, $N(t)$, the cumulative count of times at which pieces of orbital debris, at altitudes between 700 and 1100 km, passed through the field of view of a radar beam for one observation period. For the data graphed 33 pieces were detected passing through in 160.1 minutes. In the stationary case the step function should fluctuate around a straight line, as appears reasonably the case here.

Were the process homogeneous Poisson the intervals would be independent exponentials with the same mean. Figure 1b provides a plot of the points $(Y_{(j)}, j/(n+1))$ where the $Y_{(j)}$ are the order statistics of the intervals, $\tau_{j+1} - \tau_j$, between successive times. In preparing the figure the data for all the observation periods were employed. The plot would be approximately linear were the intervals homogeneous exponentials. A straight line has been applied to the plot as a reference and the exponential appears a reasonable working hypothesis, but there is a hint of departure.

A point process may be a *renewal* process, that is the intervals independent and identically distributed. Turning to this possibility the presence of serial correlation amongst the intervals is assessed. Given a stretch of values, Y_j, $j=0,\ldots,J-1$, the *periodogram* is defined by

$$\frac{1}{2\pi J} \mid \sum Y_j \exp\{-i\lambda j\} \mid^2$$

In the case of independence this statistic will fluctuate about a constant level. Figure 2a provides the average of the periodograms of the sequences of intervals between objects averaging over the observation stretches. Also included on the plot is an estimate of that constant level and approximate 90% confidence interval lines assuming the basic process stationary and mixing. There is little evidence against the assumption of a renewal process for this data set.

The autointensity function (3.1) may also be used to examine the Poisson assumption. In the Poisson case it would be constant at the mean rate of the process. Figure 2b provides an estimate of the square root. The estimate employed merges the data from all observation periods. Approximate 90% confidence limits are indicated by the dotted lines. One sees a suggestion that the intensity is raised at lags .04 to .08, but the suggestion is not strong. A possibility is that the process could be renewal with a non-exponential interval distribution. The square root has been graphed here because in the case of reasonably lengthy stretches of data the sampling fluctuations are approximately constant.

In summary, a homogeneous Poisson process appears a plausible working hypothesis for the point process of passage times of these objects.

3.3 Continuing the Space Science Example

Figure 3a provides the times of passage of the same particles as in Figure 1a, but now the estimated altitudes of the particles are also indicated by the heights of the vertical lines. In collecting the data the sizes of the objects were also estimated, by the so-called radar cross section. This measure has an (imperfect) connection with the physical size, see Levanon (1988). Figure 3b extends Figure 3a by including the sizes of the particles. The altitude is still indicated by the y-axis height but the sizes of the objects are indicated by the radii of circles. The data here may be viewed as part of a realization of a marked point process with mark $M = (altitude, radar\ cross\ section)$.

Questions of interest include whether the sequence of marks $\{M_j\}$ is independent of the sequence of times $\{\tau_j\}$, and whether the sequences of altitudes and sizes are themselves independent and identically distributed (i.i.d.). The first question was raised in the context of earthquake sequences by Vere-Jones (1970). As will be seen below it may be addressed by spectrum analysis. Figures 4a and 4b provide average periodograms of the altitude and size values in the manner of Figure 2a. They both have the character of white noise processes. The estimates are almost totally within the approximate 90% confidence limits.

As mentioned above it is of interest to ask whether the sequence of mark values is independent of the temporal point process. One might wonder for example do larger sized objects tend to follow longer gaps? This question may be addressed via cross-spectral analysis in the case that it appears plausible to assume the inherent process stationary. Figures 4c and 4d are estimates of the coherences of the sequence of intervals, $Y_j = \tau_{j+1} - \tau_j$, with the altitude and size sequences respectively. These figures provide no evidence of substantial dependence.

Having an approximate model one can now use expressions (3.3), (3.4) to estimate some risks of interest. Taking as working model the basic point process to be Poisson of rate μ, the altitude sequence as independently i.i.d. with density $f_A(a)$ and the sizes as further independently i.i.d. with density $f_S(s)$, the intensity function o of (3.2) is given by

$$\mu f_A(a) f_S(s)$$

as is o of (3.4). The quantities appearing here may be estimated simply and thereby bounds such as (3.5) estimated.

The strength of the data analyses presented is that a broad class of alternative possibilities have been considered prior to obtaining a very simple working model.

4. Time Series

4.1. Background

A *time series* is a wiggly line, $\{Y(t)\}$, with $Y(t)$ *in* R and t *in* Z or *in* R. If $Y(t)$ is binary, taking on the values 0, 1, and the 1's are rare then the series Y appears like a temporal point process.

Given expressions for the conditional mass or density functions, such as

$$Prob\{y \le Y(t+1) < y+dy \mid H_t, \theta\} = p(t+1 \mid H_t, \theta)dy$$

in the case of t *in* Z, one can express the likelihood as

$$\prod_t p(t+1 \mid H_t, \theta) \tag{4.1}$$

Here H_t is the history $\{Y(u), u \le t\}$. The likelihood can be used to make inferences concerning θ.

4.2 Example from Public Health

The United States has a worrying high level of cesarean deliveries compared to most developed nations, Clarke and Taffel (1995). This shows itself in a substantially reduced number of babies born on weekends. Besides cesareans, the increased proportion of weekday births may be due to the number of births that are induced. Both of these are causes of concern because of increased maternal and infant health risks, *ibid*. The preceding authors list average numbers of births each day of the week for the whole USA from Sunday to Saturday as 8754, 11398, 12333, 11957, 11895, 11957, 9420 respectively for the year 1992.

In this section the dependence of delivery day on the day of the week is studied for the city of Toronto in 1986. Figure 5a graphs the number of births for each day of the year. One notes a rapid oscillation and a bowing up in the middle. The bowing corresponds to more births in the summer. The apparent dependence of the number of births on the time of year, in addition to day of the week, is something that has been noted various times before. The smooth curve added is an estimate of an underlying slowly changing rate as obtained by the function lowess() of the statistical package S, see Becker, Chambers and Wilks (1988), Cleveland et al (1992). Figure 5b provides parallel box plots of the birth counts for each day of the week. The lower counts for the first and last days correspond to the reduced number of Saturday and Sunday births. The weekday-weekend phenomenon mentioned is quite pronounced here.

The series values, $Y(t)$, are actually counts and so it appears sensible to employ a model taking some note of this. Consider modelling the count on day t as Poisson with mean

$$\mu_t = \exp\{T_t + S_{<t>}\}$$

where T_t is a slowly varying trend component and $S_{<t>}$ refers to day of week effect, $<t>$ being the day of the week date t falls on. Further consider modelling the sucessive daily counts as statistically independent. Because of the assumed independence expression (4.1) simplifies to

$$\prod_t \frac{1}{Y(t)!}(\mu_t)^{Y(t)}\exp\{-\mu_t\}$$

This model may be fit directly via the function gam() of S, Hastie (1992).

The estimates $\hat{T}_t$, $\hat{S}_{<t>}$ obtained are similar to those suggested by Figure 5. To examine the Poisson assumption the residuals of the fit may be examined for overdispersion. The estimate of the overdispersion parameter is 1.07502 so any overdispersion appears mild.

Figure 6a provides the periodogram of the original count values. The weekly effect is apparent through the presence of the peaks near 1/7 and 2/7. The trend shows itself in the higher values near frequency 0. Figure 6b is the periodogram of the Pearson residuals, having removed the estimated trend and day of the week effects. Included on the plot are approximate 95% confidence limits about the independent noise level. There is no strong suggestion of autocorrelation amongst the residuals. The model of independent Poisson counts therefore appears useful.

Were some autocorrelation suggested in the latter plot one could use the function gam() of S to include it by approximating the series by 0-1 series and including lagged values in the predictor, see Brillinger and

Segundo (1979). Becker (1986) fits an epidemic process via a conditional binomial generalized linear model.

A model such as the one obtained may be used, for example, to estimate possible changes in costs resulting from fewer elective cesareans.

5. Spatial-temporal Processes

5.1 Background

Spatial−temporal process data may be written $Y(\mathbf{r},t)$ with $\mathbf{r} = (x,y)$ or (x,y,z) and $(\mathbf{r},t)$ in some subset of R^3 or R^4. One argument, t, has a privileged character. Such data may often be conveniently displayed by a sequence of images, by a video or by spinning a surface.

The process may be available everywhere in a lattice or correspond to irregularly placed points. The latter case corresonds to a spatial-temporal point process $\{(\mathbf{r}_j,t_j) \equiv (x_j,y_j,t_j)\}$. Let H_t denote the history of this process up to and including time t. The conditional intensity is given by

$$Prob\{dN(x,y,t) = 1 \mid H_t\} = \nu(x,y,t \mid H_t)dxdydt$$

An example of spatial-temporal point process data will be presented in Section 6.3 . References include: Fishman and Snyder (1976), Vere-Jones and Thomson (1984), Rathburn (1993).

5.2 Example from Neuroscience

The next example concerns the olfactory system, that is the sense of smell. Data were collected of the response of a rabbit's sniffing an odor. The rabbitt was conditioned to respond to a particular odor. An array of sensors was applied to the brain above the olfactory bulb and electroencepholograms recorded. The array had 64 sensors laid out in an 8 by 8 latice, 3.5 mm by 3.5 mm. Bursts between breaths were measured with values taken 2 ms apart. There were $T = 38$ temporal values recorded for each burst and $J = 12$ replicates. For a description of the experiments see Freeman and Grajski (1987).

The data may be written $[Y_j(x,y,t)]$ with (x,y) location, t time and j replicate, $x,y = 1,...,8$, $t = 1,...,38$, $j = 1,..,12$. Figure 7 shows the data of the first replicate. One notes oscillations at a possibly common frequency, with the amplitude of the oscillations varying with position in the array.

The analysis presented focuses on the separation of space and time variation. In spatial-temporal circumstances separation of variables is an important analytic technique.

For the j-th replicate, consider the model

$$Y_j(x,y,t) = a(x,y)\gamma_j(t) + \varepsilon_j(x,y,t) \tag{5.1}$$

with $a(.)$, fixed and standardized by $\sum a(x,y)^2 = 1$, with the $\gamma_j(.)$ independent stationary time series of common power spectrum $f(.)$ and with the $\varepsilon_j(.)$ independent white noise processes of variance σ^2.

The relationship (5.1) is a form of random effects model. In the case that the γ_j, ε_j are Gaussian one can consider estimating the unknowns by maximum likelihood. This fitting is conveniently carried out in the frequency domain employing the EM method (Dempster et al (1977). The steps are: first Fourier transform with respect to t to obtain

$$Y_j(x,y,\lambda) = a(x,y)\Gamma_j(\lambda) + E_j(x,y,\lambda)$$

with λ taking the values $2\pi j/38$. Then having some initial values compute

$$\hat{a}(x,y) = \sum_{j,\lambda} Y_j(x,y,\lambda)\overline{\Gamma_j(\lambda)}/C$$

followed by

$$\hat{\Gamma}_j(\lambda) = [\sum_{x,y} Y_j(x,y,\lambda)a(x,y)]f(\lambda)/(f(\lambda) + \sigma^2)$$

and

$$\hat{f}(\lambda) = \sum_j |\Gamma_j(\lambda)|^2/J$$

$$\hat{\sigma}^2 = \sum_{j,x,y,\lambda} |Y_j(x,y,\lambda) - a(x,y)\Gamma_j(\lambda)|^2/JXY\Lambda$$

respectively. Finally iterate to convergence. The divisor C is chosen so that $\sum \hat{a}(x,y)^2 = 1$, while $J = 12$, $\Lambda = 19$ and $X,Y = 8$.

Figures 8 and 9 show the results of the computations. Figure 8 displays the fitted spatial function, $\hat{a}(x,y)$ in both perspective and contour fashion. It shows an apparent focus of activity. Figure 9 contains the twelve estimated time series components, $\hat{\gamma}_j(t)$. The estimated replicate time series show oscillations, as was to be anticipated from Figure 7. The amplitudes do vary noticeably with replicate.

Figure 10 shows the residual series $Y(x,y,t) - \hat{a}(x,y)\hat{\gamma}_1(t)$, for the first replicate on the same scale as Figure 7. One sees the amplitudes to be much reduced and the series to be noisier.

Following a classic approximation one can act as if the empirical Fourier transform values E_j are approximately Gaussian, with values at the Fourier frequencies independent, but the assumption of Gaussian γ is basic to the maximum likelihood analysis presented.

This type of work may be seen as establishing base values preparatory to seeking possible changes from base values resulting from some treatment.

6. Particle Processes

6.1 Background

A *particle process* represents the path or trajectory of an object moving along a line, around in a plane or about in space. In the case of the plane the trajectory may be represented by $(X(t), Y(t))$ where $X(t)$ and $Y(t)$ give the $x-y$ coordinates of the particle's location at time t. The particle could be meandering or pole-seeking. The representation $(X(t), Y(t))$ is that of a bivariate time series, but the conceptualization of the problem is often quite different.

An example on a grand scale is provided in Eddy and Que (1995) where there is discussion of how to display and analyse aircraft flights over the continental United States. The process is a collection of paths, $(X_j(t), Y_j(t), Z_j(t))$, in R^3.

6.2 An Example from Ecology

Next an example from ecology is presented - the migration path of an elephant seal. These animals were near extinct at the turn of the century so there is a societal need to learn more about their behavior. Figure 11 graphs the path of one animal as an example. The animal starts from the Channel Islands off Santa Barbara, California, proceeds to the northwest and then returns. A great circle route has been added to the figure for reference. This animal seems to know surprisingly well where she is going. The problem of how to describe such paths is of interest.

In Brillinger and Stewart (1998) a pole seeking model on the sphere is considered. Suppose θ, ϕ, δ respectively denote longitude, colatitude and speed in a coordinate system such that the animal is traveling to the North Pole. Since the method of estimating the noontime location is quite indirect, (based on times of sunrise, noon, sunset recovered when the animal returns) there is measurement error involved. The equations set down in Brillinger and Stewart (1998) have the form

$$\theta_t' = \theta_t + \tau\varepsilon_t' \tag{6.1}$$

$$\phi_t' = \phi_t + \tau\gamma_t'/\sin\theta_t' \tag{6.2}$$

$$\theta_{t+1} - \theta_t = \frac{\sigma^2}{2\tan\theta_t} - \delta + \sigma\varepsilon_{t+1} \tag{6.3}$$

$$\phi_{t+1} - \phi_t = \frac{\sigma}{\sin\theta_t}\gamma_{t+1} \tag{6.4}$$

with ε, γ, ε', γ' unit variance independent Gaussian noise processes. The latter two processes correspond to measurement error.

First the model (6.3-4) for the case of no measurement error is fit by maximum likelihood. The values obtained are:

$$\hat{\delta} = .0112(.0011)\textit{radians}$$

$$\hat{\sigma} = .00805\textit{radians}$$

The full model (6.3-4) is a nonlinear state space model, see eg. Harvey (1989). A likelihood may be found based on it once one has an expression for the conditional density, $p(\theta_{t+1}',\phi_{t+1}' \mid H_t,\delta,\sigma,\tau)$, as in (4.1). This is not directly available but may be estimated by Monte Carlo by generating realizations of the processes θ, ϕ and then averaging.

In the case with measurement error, and supposing the outbound speed is δ while the inbound is δ_1, the estimates are:

$$\hat{\delta} = .0126(.0001)$$

$$\hat{\delta}_1 = .0109(.0001)$$

$$\hat{\sigma} = .000489(.000004)$$

$$\hat{\tau} = .0175(.0011)$$

all in radians. Now the measurement errors, ε_t', γ_t' appear dominant, not the foraging movement given by the ε and γ of (6.3-4).

More complex Monte Carlo sampling schemes are available to use here, see eg. Stoffer and Wall (1991), Kitagawa (1996), but were unnecessary because of the small sample size and number of parameters involved.

6.3 Example from Pest Management

This last example is not developed as much as the preceding ones. It is meant to illustrate a data type and the beginnings of model development. It could also have been presented in Section 5.

In 1975 a medfly epidemic took place in the Los Angeles area of California, see Routhier (1977), Hagen et al. (1981). It covered the period September 24 to December 1. Figure 12a plots the locations at which medflies or their larvae were discovered. The map is the area north of Santa Monica and the shaded area is the Pacific Ocean. The incident began in the lower part of the figure in Culver City, and ended with trappings at the lower and upper reaches of the region. Figure 12b graphs the numbers observed each day, with a peak the 37th day. These data are once again spatial-temporal. Figure 12 separates the spatial and temporal variables.

To bring out the joint spatial-temporal character Figure 13 shows the locations of medfly sightings for successive nine day periods. Figure 14 plots the distances from the location of the initial sighting for the successive sightings. One sees sightings at a distance as the epidemic ends. A lowess line has been added as a reference.

The interpretation of this data set is complicated by many things including: the locations of the traps, the inefficencies of the traps (often described as low), the timing of visits to the traps and the eradication treatment. Treatment began early in October with application of malathion to host plants. It was soon realized that this approach was too slow to acieve eradication. In November a sterile fly release program began and eventually about 20 million such flues were being released each week. Alltold 500 million sterile flies were released and some 280,000 of these were trapped in the eight months of the program, see Routhier (1977).

The basic entomological and geographical processes are interesting. The bugs will be spreading by flying, sometimes assisted by the wind. Also the numbers will be increasing rapidly as eggs are laid and become adults. Adults will be dieing. If at some time a bug flies near a trap they may be attracted and caught. The data will become available only when a trap is examined. The bug trajectories might be modelled as particle processes, as the elephant seal migration was above. The locations of the traps, the ranges of the traps and the timing of trap examinations all affect the data obtained. The eradication effort will need to be included in the models. The mathematics of the spatial-temporal birth and death process are pertinent, see eg. Cox and Isham (1980).

One problem is how to use such data to estimate the characteristics of the overall population.

7. Other Types of Processes, Data and Techniques

There are various other data sets and associated processes that could well have been discssed in the spirit of the paper. Perhaps foremost are the spatial processes, $Y(\mathbf{r})$ with $\mathbf{r} = (x,y)$ or (x,y,z). One can also mention line processes, hybrids eg. sampled values, $Y(\tau_j)$, processes on graphs, trees, shapes, tesselations and other geometric entities. Stoyan et al. (1987) is one reference.

Another topic that might have been presented is the case where the values of a process are discrete, falling into categories. The categories may be ordered, that is the values ordinal.

Difficulties arising in working with data have not been discussed. There are problems with: biased estimates, long range dependence, outliers, missing values, ...

8. Discussion

The goal of this paper has been to present in comparative, parallel fashion examples where the basic data may be seen as part of a realization of a random process. Statistics texts often contain substantial material on descriptive statistics, focusing on numerical quantities and figures separately from any stochastic modelling. In part this has been the approach of the present paper. The classic problems of uncertainty estimation and goodness of fit are ever present, but the paper has not focused on these.

9. Acknowledgements

This paper was prepared with the partial support of NSF Grants DMS-9625774 and DMS-9704739. The author thanks Mark Matney, Walter Freeman and Brent Stewart for providing the data sets of Sections 3, 5.2, 6.2 respectively.

References

[1] Becker, N. G. (1986). A generalized linear modelling approach to the analysis of data from a single epidemic. Pp. 464-467 in Pacific Statistical Congress (Eds. I. S. Francis, B. F. J. Manly and F. C. Lan). Elsevier, Amsterdam.

[2] Becker, R A., Chambers, J. M. and Wilks, A. R. (1988). The New S Language. Wadsworth, Pacific Grove.

[3] Brillinger, D. R. (1978). Comparative aspects of the study of ordinary time series and point processes. Pp. 33-133 in Developments in Statistics, Vol. 1 (Ed. P.R. Krishnaiah). Academic, New York.

[4] Brillinger, D. R. (1982). Some bounds for seismic risk. Bull. Seismol. Soc. America 72, 1403-1410.

[5] Brillinger, D. R. and Segundo, J. P. (1979). Empirical examination of the threshold model of neuron firing. Biol. Cybernetics 35, 213-228.

[6] Brillinger, D. R. and Stewart, B. S. (1998). Elephant seal movements: modelling migration. Canad. J. Statistics, to appear.

[7] Clarke, S. C. and Taffel, S. (1995). Changes in Cesarean delivery in the United States, 1988 and 1993. Birth 22, 63-67.

[8] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1992). Local regression models. Pp. 309-376 in Statistical Models in S (Eds. J. M. Chambers and T. J. Hastie). Wadsworth, Pacific Grove.

[9] Committee on Space Debris (1995). Orbital Debris: a Technical Assessment. National Academy Press, Washington.

[10] Cox, D. R. and Isham, V. (1980). Point Processes. Chapman and Hall, London.

[11] Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events. Methuen, London.

[12] Cox, D. R. and Lewis, P. A. W. (1972). Multivariate point processes. Sixth Berkeley Symp. Math. Stat. Prob., 3, 401-448.

[13] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. Royal. Statist. Soc. B 39, 1-22.

[14] Eddy, W. F. and Que, S. (1995). Dynamic three-dimensional display of U.S. air traffic. J. Comp. Graph. Stat. 4, 261-280.

[15] Fishman, P. M. and Snyder, D. L. (1976). The statistical analysis of space-time point processes. IEEE Trans. Inf. Theory IT-22, 257-274.

[16] Freeman, W. J. and Grajski, K. A. (1987). Relation of olfactory EEG to behavior: factor analysis. Behavioral Neuroscience 101, 766-777.

[16] Hagen, K. S., Allen, W. W. and Tassan, R. L. (1981). Mediterranean fruit fly: the worst may be yet to come. California Agriculture, University of Californa.

[17] Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge U. Press, Cambridge.

[18] Hastie, T. J. (1992). Generalized additive models. Pp. 249-308 in Statistical Models in S (Eds. J. M. Chambers and T. J. Hastie). Wadsworth, Pacific Grove.

[19] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian state space models. J. Comp. Graph. Stat. 5, 1-25.

[20] Levanon, N. (1988). Radar Principles. Wiley, New York.

[21] Rathburn, S. L. (1993). Modelling marked spatio-temporal point patterns. Bull. ISI, Vol. 55, Book 2, 379-396.

[22] Routhier, W. (1977). Mediteranium fruit fly. Memorandom, California Dept. of Food and Agriculture.

[23] Snyder, D. L. (1975). Random Point Processes. Wiley, New York.

[24] Snyder, D. L. and Miller, M. I. (1991). Random Point Processes in Time and Space. Springer, New York.

[25] Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. J. Amer. Statist. Assoc. 86, 1024-1033.

[26] Stoyan, D., Kendall, W. S. and Mecke, J. (1987). Stochastic Geometry and Its Applications. Wiley, New York.

[27] Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. J. Royal Stat. Soc. B 32, 1-62.

[28] Vere-Jones, D. and Thomson, P. J. (1984). Some aspects of space-time modelling. Proc. XIIth Int. Biom. Conf. Tokyo, 265-275.

Figure Legends

Figure 1. The top graph, 1a, plots the cumulative count of debris pieces passing through the field of view of the radar after observing has started for some period. Figure 1b is an exponential probability plot for all the observed intervals between successive objects passing.

Figure 2a is the average of the periodograms for the data of the various observation periods. Also included are approximate 95% marginal confidence limits. Figure 2b is the estimated autointensity function with approximate 95% confidence limits.

Figure 3a represents the times of objects passing through and the corresponding altitude. Figure 3b is as Figure 3a, but now circles are included to represent the sizes of the objects.

Figures 4a and 4b are the averages of the periodograms of the altitudes and sizes, averaging over the available observation periods. Figures 4c and 4d are coherence estimates for the intervals beween successive passages with the altitude and size series repectively. The upper null 95% marginal confidence line has been added.

Figure 5a provides the number of births in Toronto 1986 for each day of the year. A smooth lowess curve has been superposed. Figure 5b presents parallel stem-and-leafs split by day of the week.

Figure 6a is the periodogram of the series of Figure 5a. Figure 6b is the periodogram of the residuals having removed an estimated trend and the daily effects.

Figure 7. The electroencephalograms of the first replicate of the experiment for the 8 by 8 array.

Figure 8. The estimated spatial function, $a(x,y)$, of the model (5.1).

Figure 9. The estimated latent series for the 12 replicates.

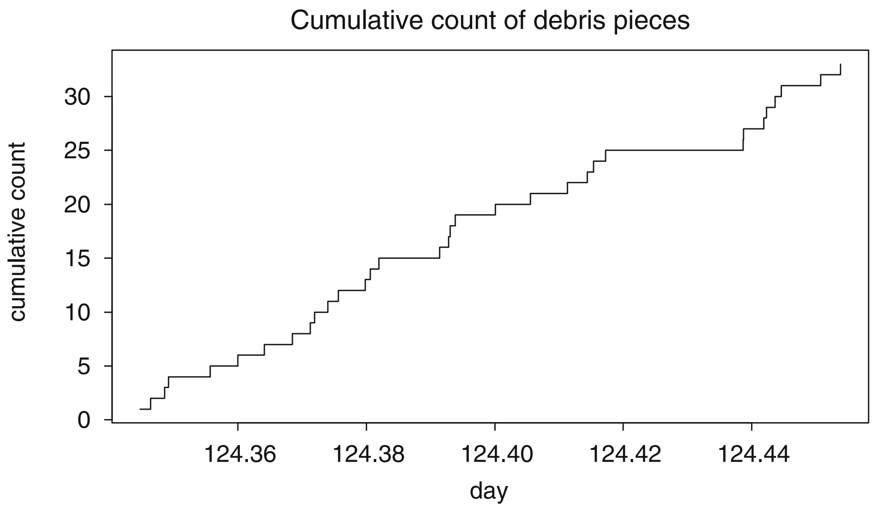
Figure 10. The residual series having fit the model (5.1).

Figure 11. The outbound and inbound tracks of an elephant seal heading into the Northwest Pacific from near Santa Barbara, California.

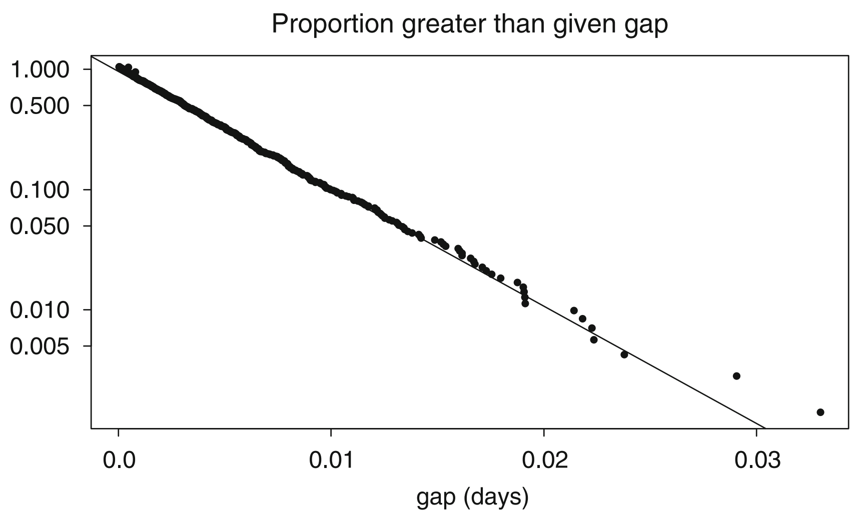
Figure 12a is a plot of the locations of trappings of medflies, adults and larvae, during an outbreak in 1975. The shaded region is the Pacific Ocean. Figure 12b graphs the counts noted each day during the epidemic.

Figure 13. The plot of Figure 12a, but for successive 9 day periods.

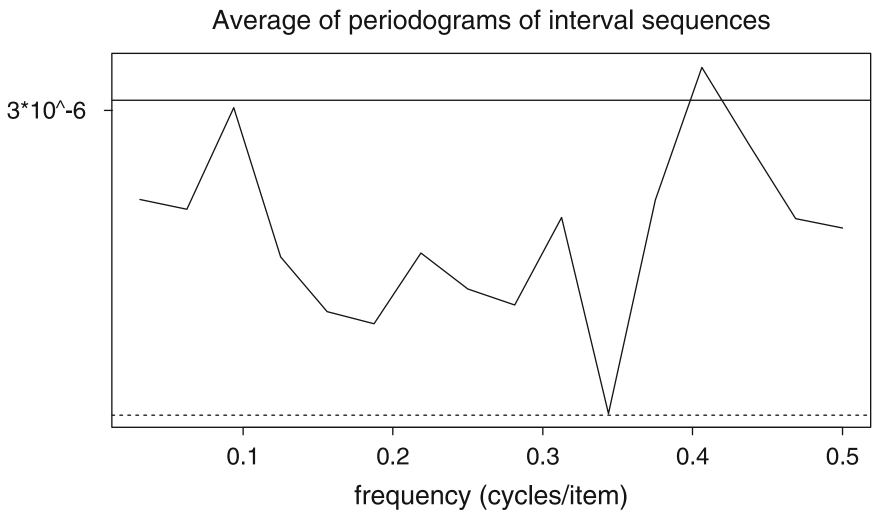
Figure 14. The distances of each sighting from the original, for each day.



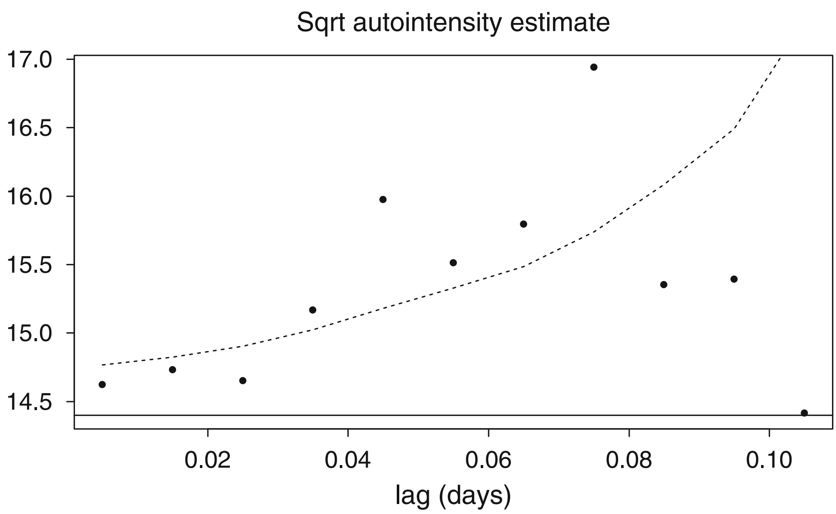
Cumulative count of debris pieces
cumulative count
0
5
10
15
20
25
30
124.36
124.38
124.40
124.42
124.44
day



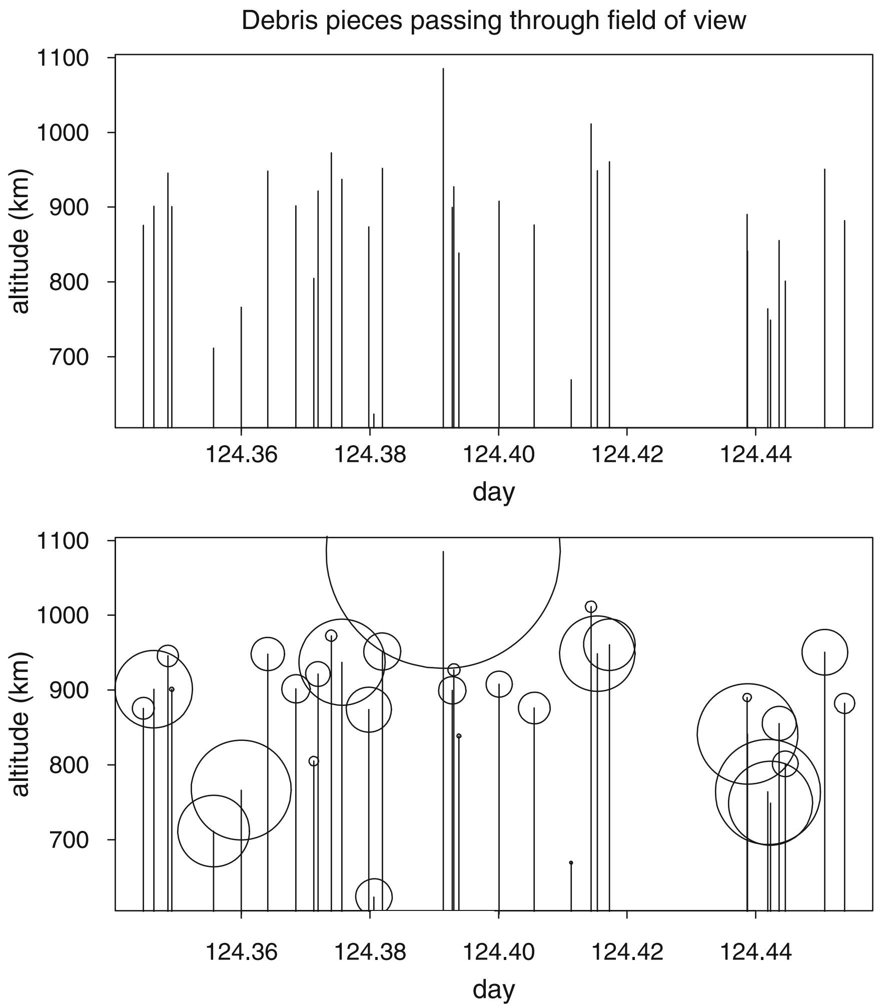
Proportion greater than given gap
1.000
0.500
0.100
0.050
0.010
0.005
0.0
0.01
0.02
0.03
gap (days)



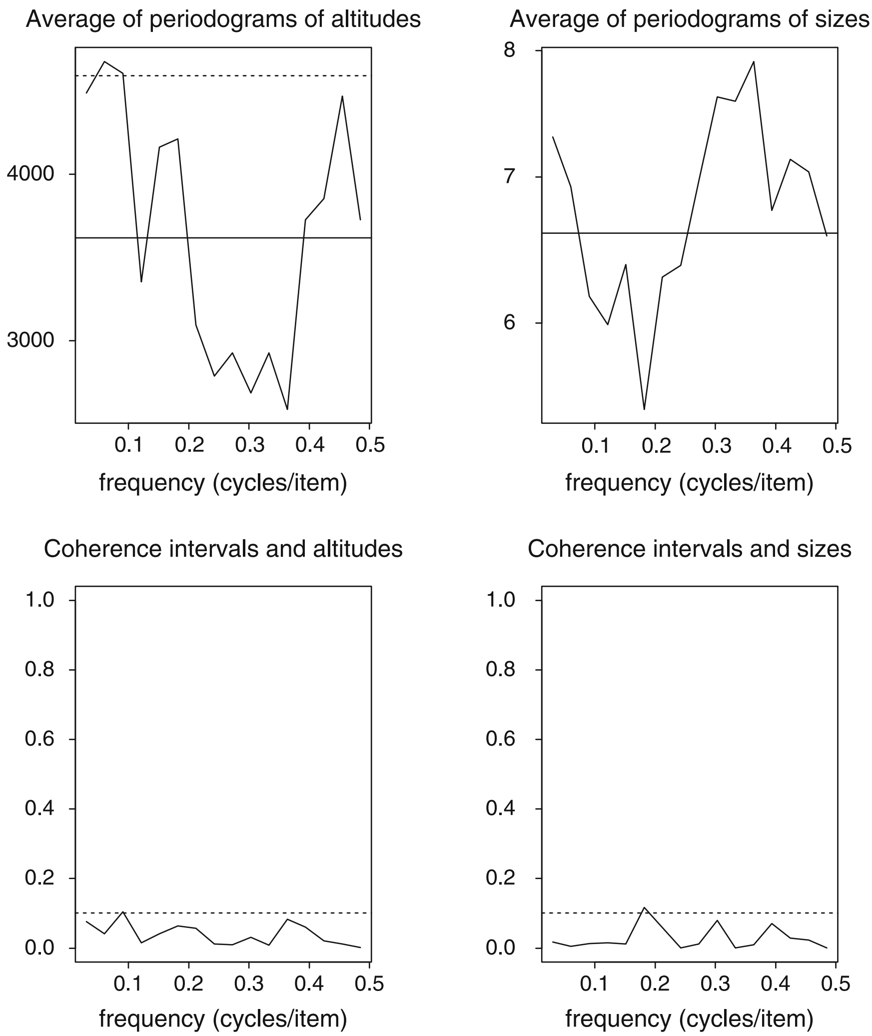
Average of periodograms of interval sequences
3*10^-6
0.1
0.2
0.3
0.4
0.5
frequency (cycles/item)

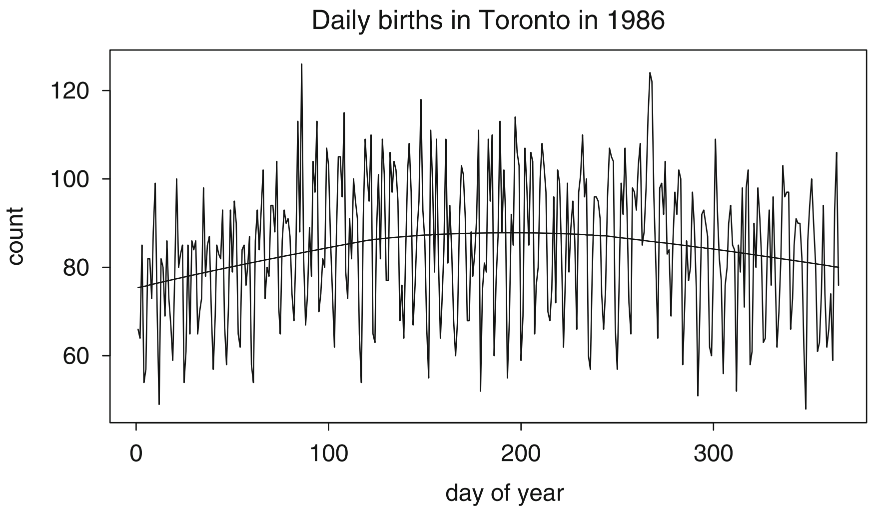


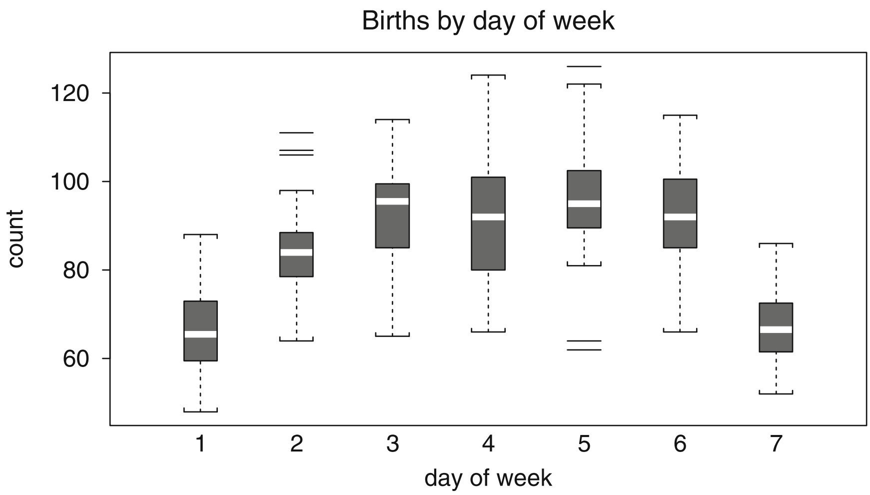
Sqrt autointensity estimate
17.0
16.5
16.0
15.5
15.0
14.5
0.02
0.04
0.06
0.08
0.10
lag (days)

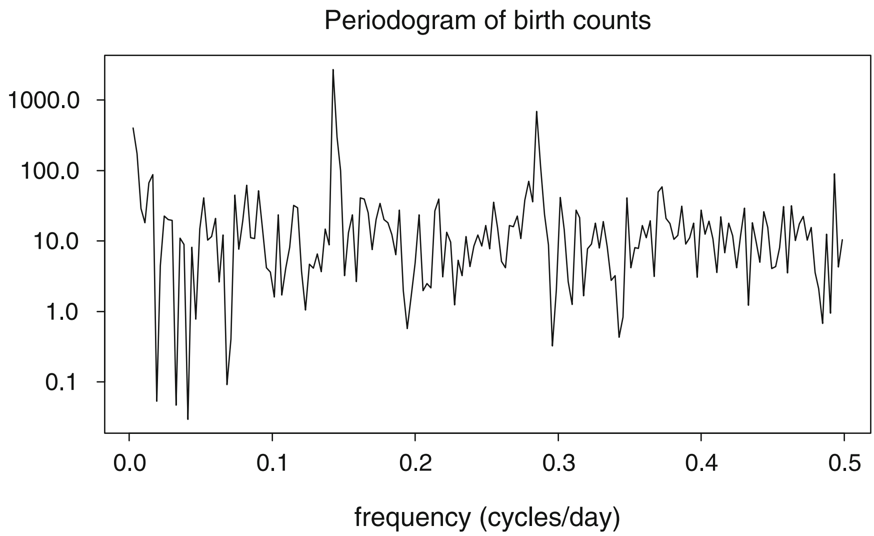


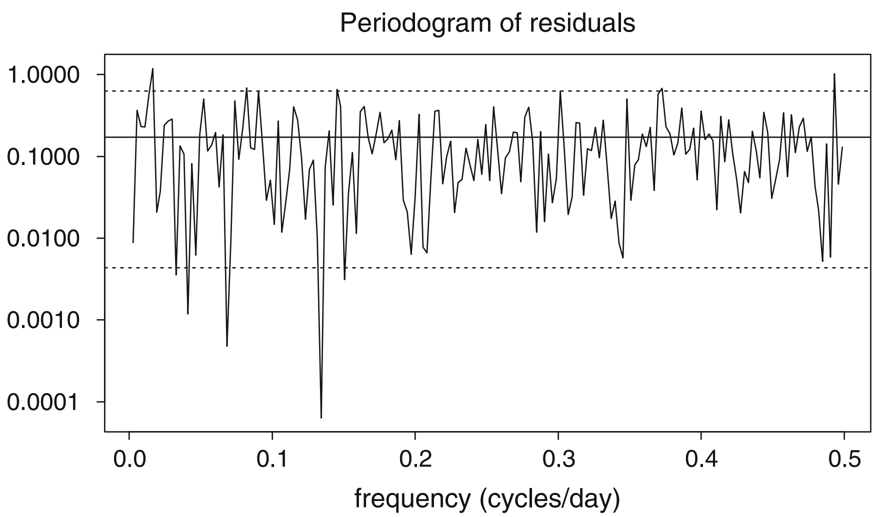
Radius of circle is proportional to RCS

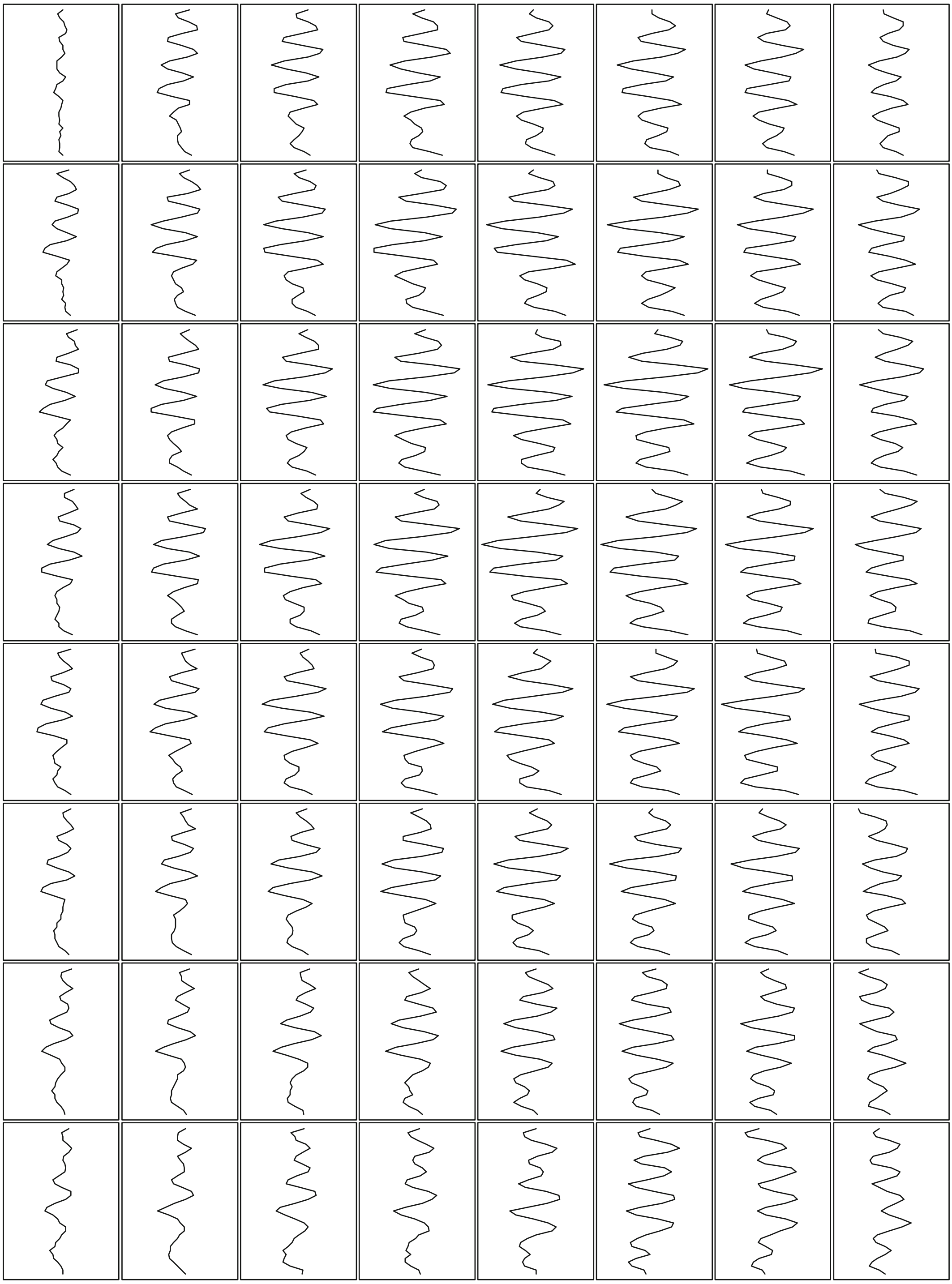

Average of periodograms of altitudes
4000
3000
0.1 0.2 0.3 0.4 0.5
frequency (cycles/item)
Average of periodograms of sizes
8
7
6
0.1 0.2 0.3 0.4 0.5
frequency (cycles/item)
Coherence intervals and altitudes
1.0
0.8
0.6
0.4
0.2
0.0
0.1 0.2 0.3 0.4 0.5
frequency (cycles/item)
Coherence intervals and sizes
1.0
0.8
0.6
0.4
0.2
0.0
0.1 0.2 0.3 0.4 0.5
frequency (cycles/item)

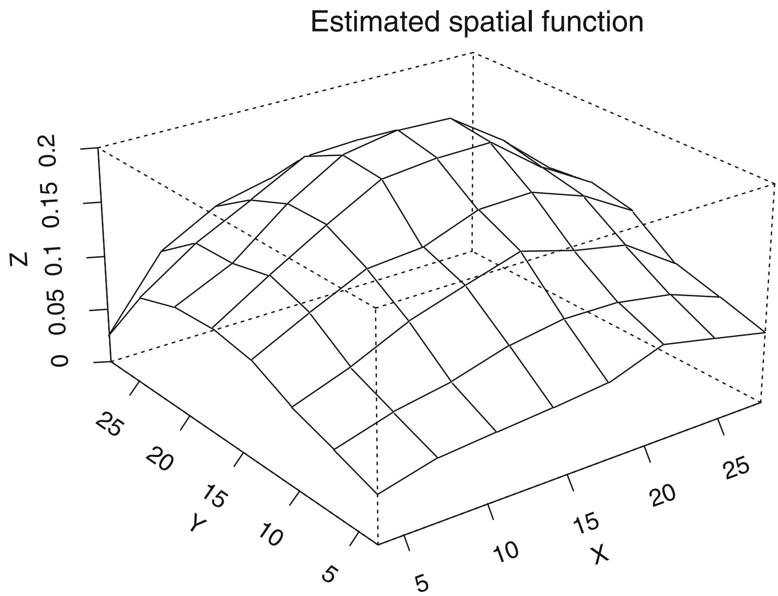

Daily births in Toronto in 1986
count
120
100
80
60
0
100
200
300
day of year

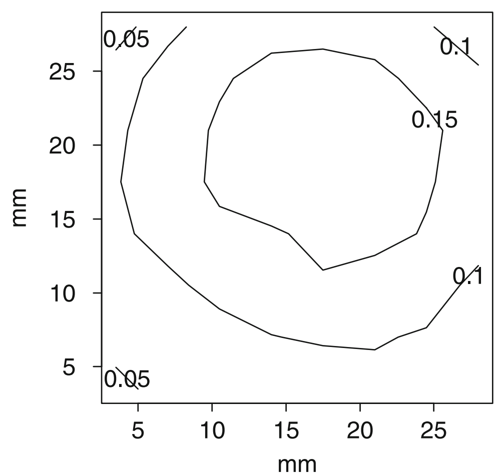

Births by day of week
count
120
100
80
60
1
2
3
4
5
6
7
day of week


Periodogram of birth counts
1000.0
100.0
10.0
1.0
0.1
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/day)

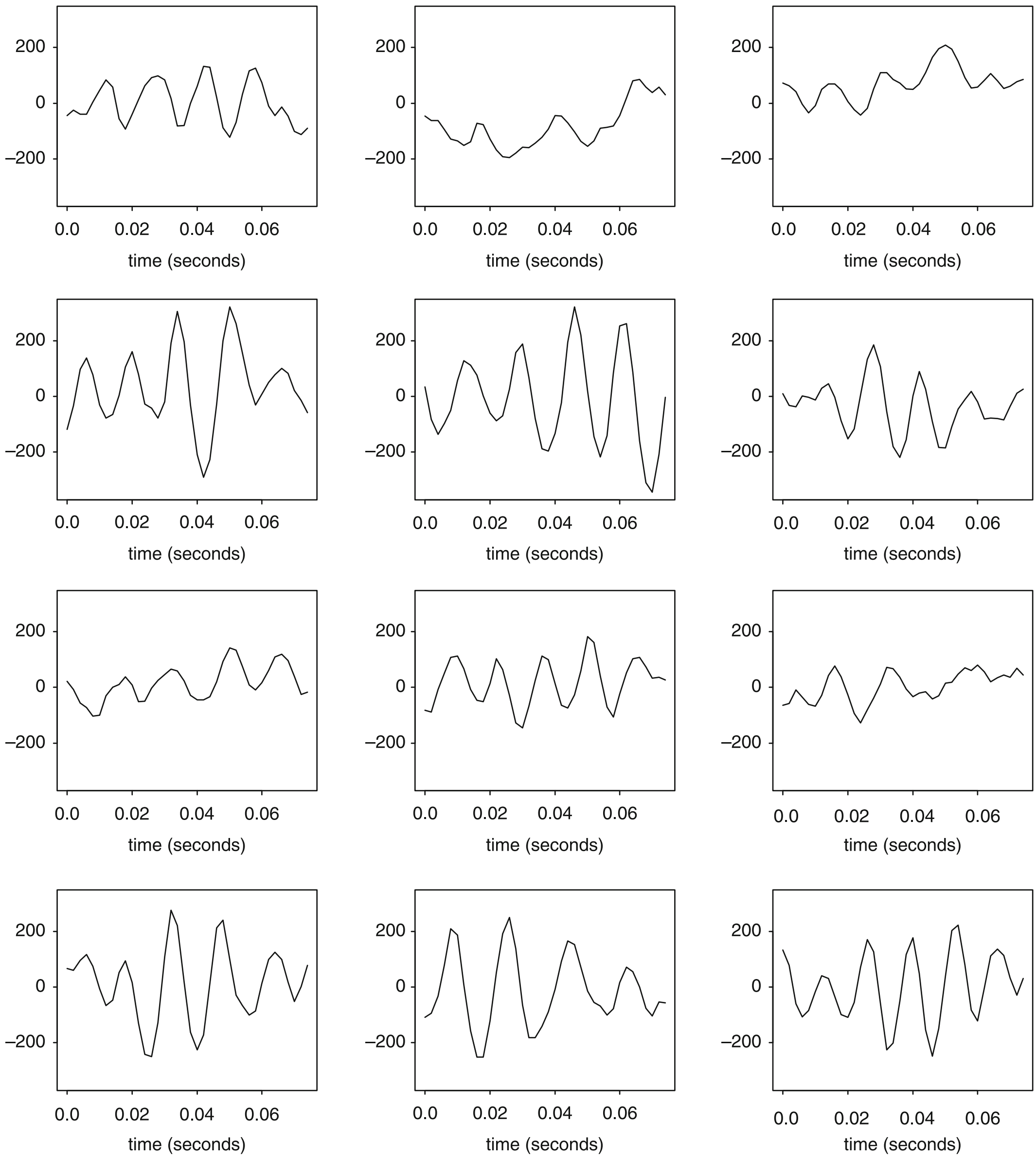

Periodogram of residuals
1.0000
0.1000
0.0100
0.0010
0.0001
0.0
0.1
0.2
0.3
0.4
0.5
frequency (cycles/day)

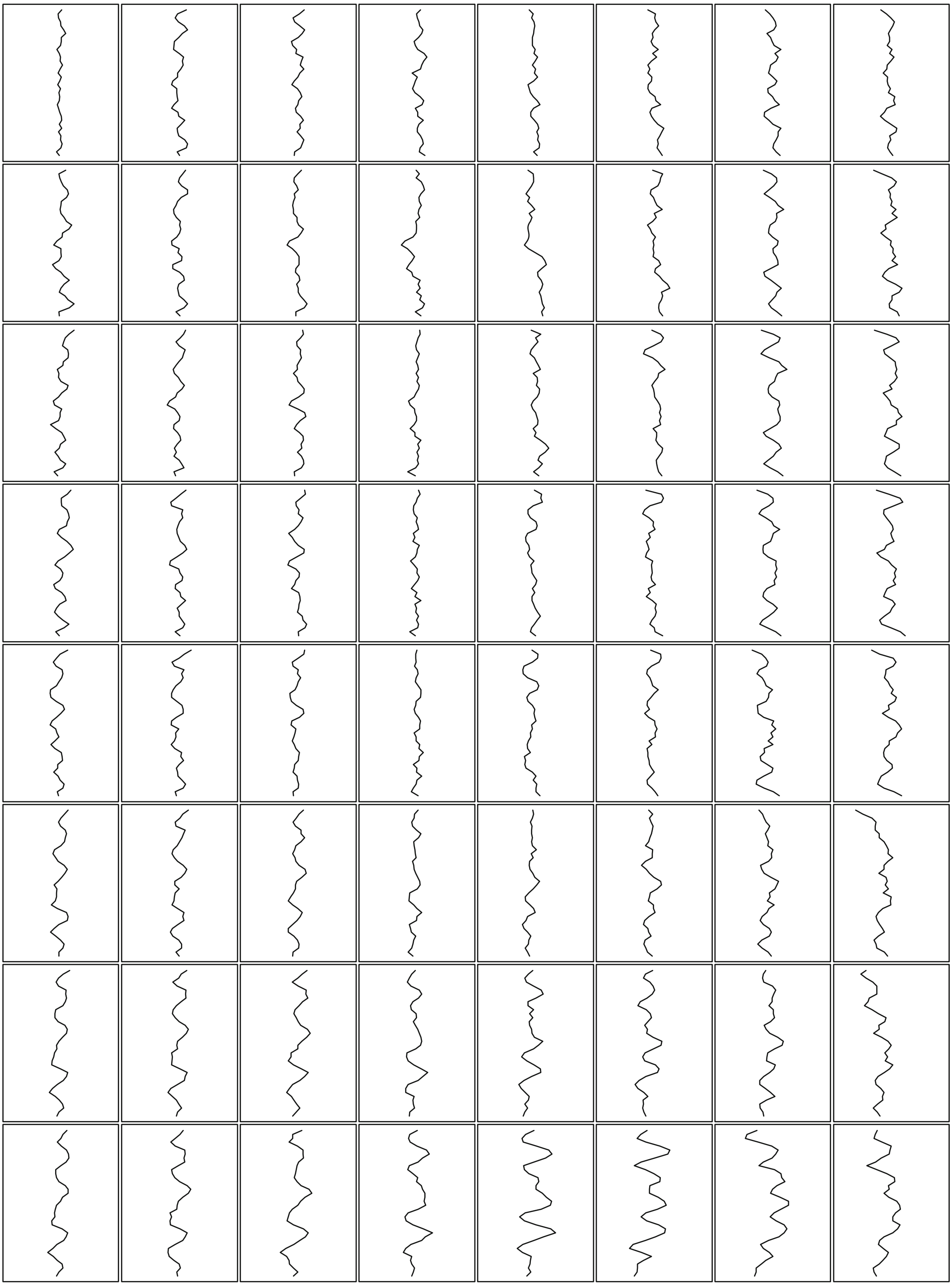


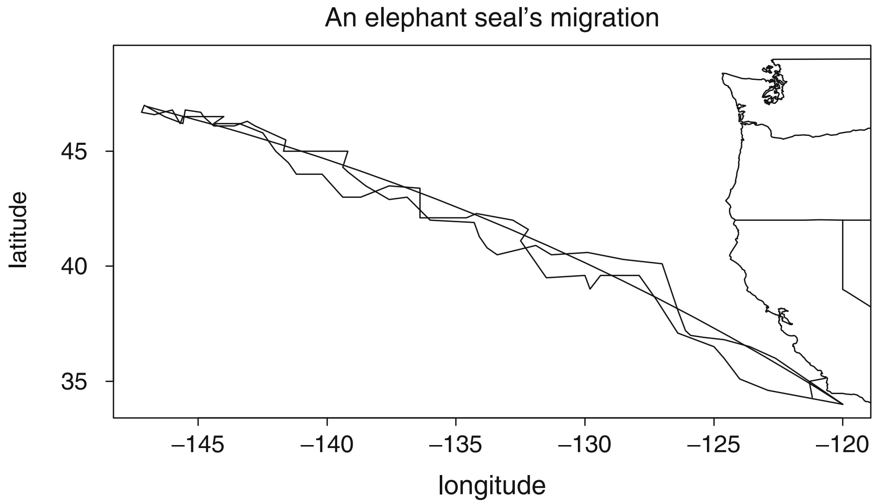

Estimated spatial function
Z
0
0.05
0.1
0.15
0.2
Y
25
20
15
10
5
X
5
10
15
20
25


0.05
0.1
0.15
0.1
0.05
mm
25
20
15
10
5
5
10
15
20
25
mm

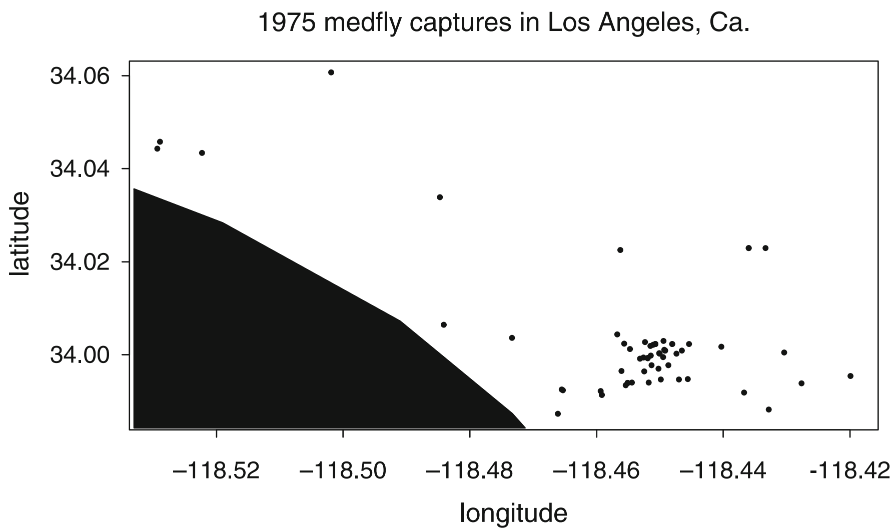
Estimated replicate time functions

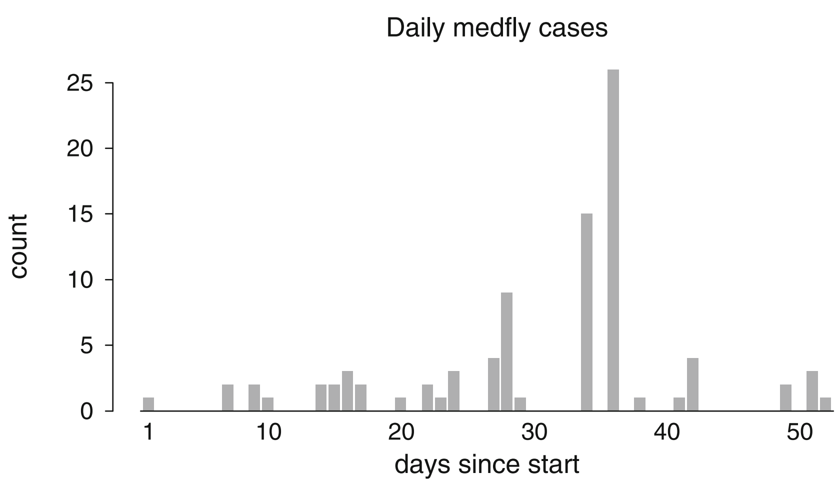


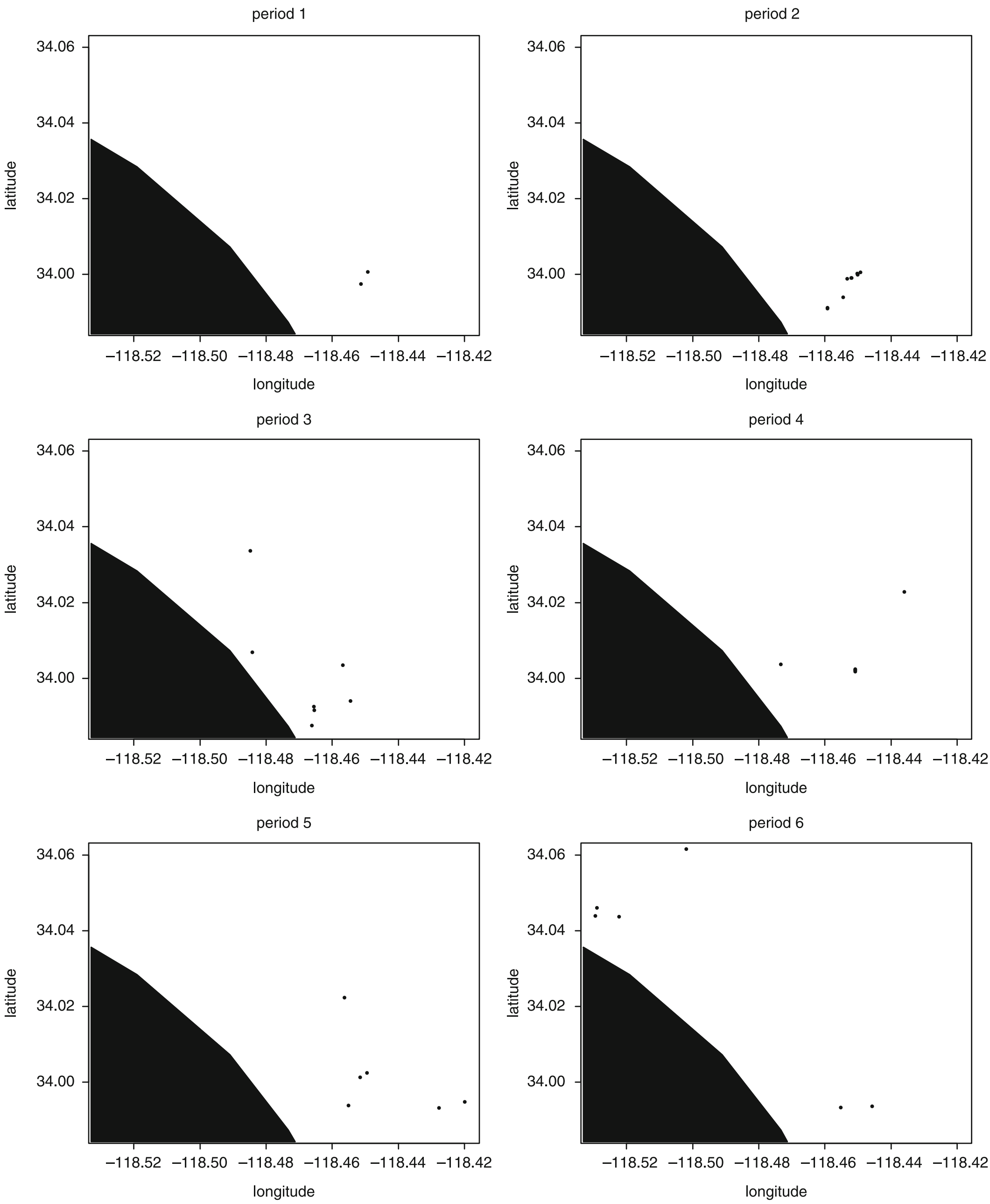


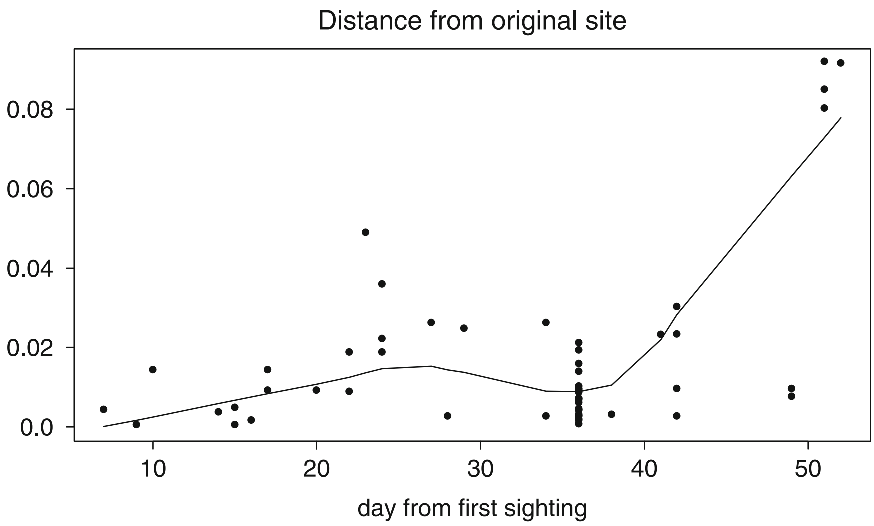


A great circle route is superposed






period 1
period 2
period 3
period 4
period 5
period 6
latitude
longitude
34.06
34.04
34.02
34.00
−118.52
−118.50
−118.48
−118.46
−118.44
−118.42



Distance from original site
0.08
0.06
0.04
0.02
0.0
10
20
30
40
50
day from first sighting

Part III
Population Biology and Environment

Commentary: Introductory Comments to Some Applied Papers by David R. Brillinger, by Tore Schweder and Haiganoush Preisler

In addition to statistics, David took care in developing my attitude as a scientist - and he wrote a poem of his own in my draft thesis about whales and statistics. He also cared for us personally. We were invited to use Lorie's and David's house when they went to New Zealand in the summer of 1973. Our newborn child spent her first time out of Alta Bates hospital in their house. David also gave me support in a more touchy matter. I was on a US Navy grant, and felt uneasy when I realized that I had to acknowledge the grant in a publication. Strike it in the last galley, was David's advice – which I in the end did not follow. And there was fun, also outside the soccer field. David suggested the movie "The harder they come". My son, an aspiring reggae musician, was happy to find the Jimmy Cliff LP in my old stock. In the last couple of years we have been lucky to have David as an advisor in our Centre for Ecological and Evolutionary Synthesis in Oslo, and to have David repeatedly visiting.

Empirical modelling of population time series data: The case of age and density dependent vital rates [1980]

A stochastic matrix model is used to study a population of sheep blowfly observed over two years in a lab. The flies were kept in a cage, and fed on a constant diet. The population experienced substantial fluctuations in size over the observational period. Matrix models for stage-structured populations like the sheep blowfly have become popular (Caswell (2000) *Matrix Population Models: Construction, Analysis, and Interpretation*, is cited some 1900 times).

The hypothesis behind the experiment was that competition for resources would occur only in egg laying and that the population fluctuations were due to variability in recruitment. Mortality was thought only to depend

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_18, © Springer Science+Business Media, LLC 2012

on age. By fitting a product model to age-and abundance specific survival probabilities by weighted least squares, it is found that age-specific mortality does depend on abundance, and also on abundance two days earlier. By residual analysis it is also found that there are further dependencies.

The present paper is an early study with vital parameters, particularly the mortality rate, depending on population size and being affected by random variation. It has influenced the field of population dynamics to make more use of rather standard statistical methodology.

Learning a Potential Function from a Trajectory [2007] This short Signal Processing Letter presents stochastic differential equation models for moving objects where the drift term is the negative gradient of a potential function, providing a formal background and a general discussion missing in the literature. These models can have regions of attraction, absorbtion or repulsion. Various interesting and useful potential functions that are linear combinations of given differential functions are considered. Being linear in unknown parameters, they can be estimated by linear regression although with stochastic regressors. Asymptotic theory is presented for the potential function estimator based on standard assumptions on conditional independence and zero mean residuals. A similar regression model is used in the soccer study (Brillinger 2007b) discussed below. For curved potential functions one might wonder whether the residual terms really have zero mean. Another question is how the asymptotics of the estimated potential depends on the potential function itself. If the attraction, say to a point, is sufficiently strong, a single trajectory might for example not provide enough information to allow first order convergence to normality.

A Potential Function Approach to the Flow of Play in Soccer [2007] Soccer, or football as we say outside USA and Canada where this game is by far the most popular of sports, is a game played by two teams of 11 players each. The play field is rectangular about 105 by 68 m, with a goal at each short end. The purpose is for each team to have the ball inside the goal of the other team as many times as possible, and have the ball inside its own goal as few times as possible. The ball is passed from player to player within a team, usually by the kick of a foot, until a goal is scored, the ball gets off the field or is picked up by the other team or is lost because an arm or hand is used or some other rule is broken.

David got interested in soccer during his years at LSE. In the early 1970s we were several Norwegian students at the Berkeley department. We got a Norwegian newspaper to the coffee room, and David was quick to grab it to get news of soccer in England.

The purpose of this article is to establish a statistical framework for describing and simulating how a game of soccer develops. This is done by breaking the game out in spells of ball occupancy. A spell is a succession of passes of the ball within a team. An unusually long spell in the 2006 World Cup game between Argentina and Serbia-Montenegro is studied in detail. It had 25 passes and ended with a goal for Argentina. The spell is characterized by the position r_i of the ball when pass i is initiated, and also the time t_i. A potential function H that describes the spatial succession $r_i \to r_{i+1}$ for given times is assumed:

$$(r_{i+1} - r_i) / (t_{i+1} - t_i) = -\nabla H(r_i) + noise.$$

The potential function is assumed linear in some parameters, and so is the gradient. These parameters are estimated by least squares from the positions and times of the spell. The estimated potential function is a bit skewed towards the left of the field seen from the Argentinian side, and might be symmetrized when used to simulate a game. To simulate a game, a potential function is also needed for the other team, and also a way to simulate time points within spells and a stopping time for spells.

Since passes generally are made towards the opposing goal where the potential function is steepest, the least squares approach taken in the paper will bias the estimated potential function towards less steepness. A partial fix is to also include the curvature (the Hessian) in the regression. Another point to be made is that simulating soccer games are done in different computer games. How are these simulations carried out relative to the method proposed here?

Analyzing games of soccer, i.e. pulling them apart in their basic elements like spells, and characterizing teams by their estimated potential functions and other characteristics might be useful for understanding games and for the training of a team. One point of particular interest might be to identify to what degree and in what aspects game results are determined by simply combining the features of each team without further interaction than independent succession of spells, and in what aspects more complex interaction must be invoked. If no such further interactions are of importance, a team should be trained without regards to qualities of the next opposing team.

The use of potential functions in modelling animal movement [2001] Potential functions are used in the physical sciences to model the motion of planets or particles in a field of gravitation or other forces determining velocity and direction. A two-dimensional stochastic differential equation (SDE) provides a stochastic version of the deterministic potential function model. The drift term in the SDE is then the negative gradient of the potential function. The authors use an SDE to model the motion of elk (and mule deer) in an enclosed forest. The extensive data was obtained from researchers who fitted a number of animals with collars containing Loran -C receivers. The position of a tagged elk is recorded about every minute, but with a measurement error of some 50 meters. Under the SDE model, the assumed potential function is estimated from the spatial distribution of elk positions, assuming this distribution to be the stationary distribution of the SDE. An interesting question is whether the function estimated this way by a kernel method really could be an estimate of a proper potential function. This is tested somewhat informally by the Student statistic found by comparing the two cross differentials of the estimated function. The authors find that the existence of a potential function cannot be rejected. Gray scale graphs depict the estimated potential function showing that elks tend to be in the northern end of the area during day, while more to the south during night.

Do individual elks move about according to a Markov process, say an SDE, and independently of each other? The potential function model assumes this. Although these basic assumptions cannot be tested within the SDE model, the model is very useful in summarizing position data for tagged animals to elucidate questions about habitat selection and foraging behavior, and also the effects of vehicular traffic and fences on animal behavior.

The paper shows how the SDE model might be used for animal motion data, and it finds its place in the sequence of related papers on models and numerical methods that David has together with biologists and fellow statisticians.

Elephant-seal movements: Modelling migration [1998] Elephant seals migrate between their rookeries at the California coast and their feeding areas in the North Pacific close to the Aleutian islands twice a year. Data from individuals fitted with tags measuring light conditions by time of the day when surfacing to breath, in addition to other variables were considered. The data allow the migration route to be tracked by estimating the position each morning by the length of daylight and the time of sunrise. The seals are essentially following a great circle between foraging area and haul-out. How they manage to navigate as precisely as they do is not known, but the authors speculate that the seals utilize the global magnetic field. In order to develop testable hypotheses for seal migration, formal models are developed. These models are cast in the form of stochastic differential equations for diffusion on the surface of the globe. They are of the Ornstein-Uhlenbeck type. From Brillinger (1997), the basic model for the case of steady drift along a meridian is set down:

$$d\theta_t = \left(\frac{\sigma^2}{2\tan(\theta_t)} - \delta\right) dt + \sigma dU_t, \quad d\phi_t = \frac{\sigma}{\sin(\theta_t)} dV_t$$

for latitude θ and longitude ϕ and standard Brownian motions U and V, all measured in radians. In this formulation the speed δ is negative when the motion is towards the North Pole. An alternative model is discussed below. If the target and the point of departure are at different meridians, the equations for the diffusion along a great circle are obtained by a trigonometric transformation of the above equations.

In this model, including measurement errors in the daily positions, approximate maximum likelihood estimators are obtained. Based on data from a seal migrating from the foraging area towards its rookery, estimates are obtained for speed along the great circle δ and for the diffusion standard deviation σ. The point of departure for this seal was estimated as the mean position over the days thought to be spent on foraging ahead of the migration towards the California rookery. The target, the seal's rookery, is a precise point, but the point of departure needs perhaps not be a well defined point. It might be preferable to model the foraging area as an area and not a point, say by a bivariate normal distribution located at the center of the foraging area.

In case navigation is done continuously by the magnetic field, continuous time modelling as above is appropriate. If however navigation is done celestially, discrete time modelling might be preferable. In that case navigation is prevented during daylight when the stars are invisible, and also those nights with clouds on the sky.

Tag technology has developed considerably since the elephant seal data were obtained. Modern tags would measure position by GPS at each surfacing, with very little measurement error. Such tags are widely used to understand animal behavior, and diffusion models for migration on a sphere like the one above should be in considerable demand.

An alternative model for an Ornstein-Uhlenbeck process on the sphere with drift towards the North Pole and with attraction to a meridian at longitude m is $d\theta_t = \delta dt + \sigma dU_t$, $d\phi_t = \gamma(m - \phi_t) + dV_t$. Here θ_t is latitude and ϕ_t is distance from (θ_t, m) along a great circle perpendicular to the meridian, and dU is random latitudinal disturbance and dV is the same along the perpendicular great circle. In this model, θ is causally independent of ϕ, and is therefore a one-dimensional linear SDE process. The exact likelihood for observations at discrete points in time is available. The expected velocity along the great circle is δ and the push of reverting back to the meridian at m along a perpendicular great circle at position ϕ_t is γ. This model is, perhaps, a more transparent model than that in the paper.

Random process methods and environmental data: the 1996 Hunter lecture [1997] Environmental processes like weather, river flow, earth quake damage etc. are essentially dynamic and nearly always affected by random variation, and random processes are fundamental in modelling them. This is the basic message of the paper, and three environmental processes are considered to illustrate the use of stochastic process models and statistical inference in environmental science.

In the first analysis nearly a century of daily river height of Rio Negro at Manaus, Brazil are analyzed. The question is whether there is an increasing trend in the flow of water out of the Amazon basin. There is considerable seasonal variation in river flow, and the approach is to fit a model with a trend plus a seasonal component and one for daily variability. Rather than modelling these components parametrically, year specific seasonal component are estimated by the median annual curve. The trend is only assumed to be non-decreasing and is estimated non-parametrically from the seasonally adjusted series. The estimated trend curve is by construction non-decreasing, and to judge whether it estimates a truly increasing trend, data were simulated from the model assuming no trend. The simulated data were analyzed in the same way as the observed data, and from visual comparison, the conclusion is that there is a soupçon (touch of) significance. This ingenious non-parametric time series analysis could be generalized. One might ask whether it yield other conclusion than more traditional parametric analyzes?

Damage due to the Loma Prieta earthquake with epicenter close to Santa

Cruz, California, is the theme of the second analysis. At various localities in the greater Bay area the earth quake damage was measured on an ordinal scale with 12 levels of increasing severity. The purpose of the analysis is to extrapolate these damage measurements to the whole of the affected area. The degree of damage at an affected locality at (x, y) is modelled as a multinomial based on a smooth spatial function $g(x, y)$ plus an extreme value distributed random variable. The contours of the estimated damage function g are shown. Perhaps the predicted value of the ordinal damage score might have been be more interesting. Other distributions than the extreme value could actually have been chosen. The predicted damage score is broadly invariant to the choice of distribution, while g has the scale of the chosen distribution.

In the third analysis the problem is to estimate the average velocity at which weather moves from west to east on the Globe. The data consist of 500 millibar pressure fields across the surface of the earth over a five day period, with two measurements (pressure fields) per day. The pressure field is modelled as $Y(x, t) = g(x) + h(x - vt) + noise$, where x is longitude in radians, v is velocity in radian/hour and t is time in hours. The longitude specific component g cancels in the difference $Y(x, t+1) - Y(x, t)$ and these differences are used to estimate the velocity by least squares.

David describes statistics as the science of using data wisely. He further makes the basic general remark that for problems such as those considered in the paper, the importance of collaboration and learning the pertinent subject matter cannot be overemphasized. Agreed! Whether data are used wisely is in fact not only a statistical matter. A statistical application is good to the extent it is statistically sound and is also helpful for the subject matter field. Without basic language and understanding of the field the risk of irrelevance is high. In his Hunter lecture David provides three wise analyzes, as he also does in his many other applied papers. He has evidently an intimate knowledge of important areas of environmental science as well as the areas in biology, earth science and other fields where he has contributed, and he cooperates well with subject matter scientists.

The 2005 Neyman Lecture: Dynamic Indeterminism in Science [2008] Jerzy Neyman (1894 to 1981) founded the Statistics Department at Berkeley in 1938 after 4 successful but turbulent years in London following his early years in Poland. Neyman's life history and some of his contributions to applied statistics are reviewed, with emphasis on his use of dynamic stochastic modelling in his applied work in astronomy, fisheries science and weather modification. Neyman was concerned with phenomena developing in time and space. David briefly presnts stochastic differential equations (SDEs) as a background to Neyman's applied work, and as a common thread in his own applied work, not the least what he presents in the paper to expand on Neyman's work and to support the case for dynamic indeterminism in science. "Indeterministic" was for Neyman broadly synonymous with "stochastic" and "statistical". Chaos or other non-probabilistic indeterminism are not mentioned, perhaps for good reasons since Neyman was a practical man who sought empirical knowledge in the many fields of science where he worked. David is also a practical man, and he uses finite differences to obtain likelihood functions in his SDE models.

Two of the three examples of Neyman's applied statistics works were done together with Elizabeth Scott. Together with astronomers they developed the Neyman-Scott model for clustered point processes when studying the spatial distribution of galaxies. From graphical comparison of photographic images of the sky and images obtained by simulating their model, they found that more clustering was needed, and they developed a two-stage Neyman-Scott model. Weather modification was another area they investigated together. Through a randomized experiment of cloud seeding in Switzerland they discovered a "far-away" effect of increased rainfall far away from the seeding. The third study was done before Scott entered the Berkeley Department as a PhD. Here, Neyman developed and estimated an age-structured model for the population dynamics of Californian sardines that was subject to heavy exploitation. This study predates the seminal book: Beverton, R. J. H.; Holt, S. J. (1957) *On the Dynamics of Exploited Fish Populations*.

David follows up on Neyman's sardine study with his own study of sheep blowflies by way of a population matrix model. This is discussed above (Brillinger, Guckenheimer, Guttorp and Oster, 1980) . David expands on the weather modification study of Neyman and Scott by using a model for transforming a point process model of seed particles above Ticinino in Switzerland to a point process of rain drops in clouds blown to Zürich, and estimates the expected delay time from seeding to increased rainfall in Zürich. David's third example is a study of spatial motion of elks fitted with GPS collars in the enclosed Starkey Experimental Forest. An SDE model was employed and the velocity field of elk motion estimated. To what extend does recreational use of the area, i.e. driving ATVs, affect elk behavior? This was studied in an experiment were a driven ATV was tracked and its trajectory introduced as an explanatory variable in the SDE model for the motion of an elk during the experimental period. The elk was significantly affected by the ATV, and the effect is graphically quantified. In his final example David studies the foraging behavior of Hawaiian monk seals fitted with GPS tags, by way of an SDE model cast in potential function form. This and the previous example is akin to Brillinger, Preisler and Ager (2001) discussed above. The potential function was modelled as a linear combination of basic functions, and was estimated by least squares. Synthetic plots, i.e. simulated tracks, from the fitted motion were found not unlike observed tracks where unreasonable satellite recorded position are cleaned up.

Sympathy and admiration for Neyman shines through the paper, but in his modest way David does not say how he was inspired and influenced by Neyman. From similarities in their appetite for applications to substantive sciences and in their great contributions, also to what we used to call mathematical statistics, the influence must have been substantial - or they were similarly gifted and born under the same star.

Lectures on Mathematics in the Life Sciences
Volume 13, 1980

EMPIRICAL MODELLING OF POPULATION TIME SERIES DATA: THE CASE OF AGE AND DENSITY DEPENDENT VITAL RATES

David R. Brillinger, John Guckenheimer,
Peter Guttorp, George Oster[1]

ABSTRACT. The total numbers of births and deaths in a population are given at discrete equispaced time intervals. It is assumed that the birth and death rates depend on age, the population size and possibly time. Further it is assumed that the rates fluctuate randomly from individual to individual. The problem is to estimate average birth and death rates and the age structure of the evolving population. Results are presented for a population of sheep blow-flies maintained under stable conditions for a two year period (361 observations) by A.J. Nicholson.

1. INTRODUCTION. Statistical analyses of population data have generally concentrated on the cases where birth and death rates depend on age alone (see for example Chiang (1968), Gane (1975), Pollard (1973)) or linearly on population size alone (see for example Bartlett (1966), Keiding (1975)). In the first case the ages of the individuals concerned have been assumed known. In the second case the age structure is ignored.

In a variety of practical situations birth and death rates are nonlinear functions of both age and population size. A variety of theoretical and simulated results have been derived for such populations in the deterministic (nonstochastic) case. Oster (1977) presents a review of this case. Supposing $N(t,x)$ to be the number alive in the population at time

---

1980 Mathematics Subject Classification. Primary 60G35, 62M10, 92A15; Secondary 92A17.

[1] Supported by NSF Grant MCS 77-22986, MCS 79-02002.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_19, © Springer Science+Business Media, LLC 2012

t aged x such studies have typically been based on the system of equations

$$\frac{\partial N(t,x)}{\partial t} \quad \frac{\partial N(t,x)}{\partial x} = -\mu(t,x)N(t,x),$$
$$N(t,0) = \int \beta(t,x)\ N(t,x)dx$$

with μ (.) the death rate (force of mortality) per individual and β (.) the corresponding birth rate. Functions (N (t, x)) satisfying the above equations, for plausible cases of μ (.) and β (.), can evidence an extremely broad range of behavior with a rich bifurcation structure as parameters are varied. (See Guckenheimer, Oster and Ipaktchi (1977).) The results that have been obtained are instructive; however their usefulness in practical situations is not clear because, among other things, (i) measurement error is ever present, (ii) individuals differ drastically in propensity to die, (iii) in general only aggregate data(i.e. non age structured) is available, (accurate life tables for populations with nonstationary age distributions are extremely rare in the ecological literature) (iv) adequate criteria for comparing models and data are lacking and (v) the solutions are unstable, so estimation of parameters is virtually impossible.

This paper is concerned with the problem of estimating nonstationary age structures and age-and-density-dependent death rates given aggregate stochastic data. The most detailed set of experiments in the age-and-density-dependent case were undoubtedly carried out by A.J. Nicholson, (Nicholson (1950)). This data has been modelled deterministically by Auslander et al (1974) with a variety of models. These models share many of the apparent qualitative features of the data. Here we begin a more quantitative assessment of the data and immediately find that some of the earlier assumptions are questionable.

2. NICHOLSON'S EXPERIMENTS. During the 1950's the Australian entomologist A.J. Nicholson carried out an extensive series of experiments concerning the population variation of Lucilia cuprina (the sheep blowfly) under various conditions. Nicholson maintained populations of the flies on various diets (some constant, some fluctuating), experiencing

different forms of competition (between larvae and adults, for egg laying space, etc.), and under many other conditions. One list that we have shows 145 major experiments. In this paper we report on the analysis of one cage of Nicholson's L97 experiment, "influence of periodic environmental changes of intrinsic oscillations". This cage was actually a control maintained under constant conditions.

The blowfly's life cycle is made up of a number of stages of varying durations (see Mackerras (1933)). The principal stages and their approximate durations are listed in table 1.

Table 1.

Stage	Duration
egg	12-24 hours
larva	5-10 days
pupa	6- 8 days
immature adult	4 days
mature adult	1-35 days

The durations listed are only meant to be suggestive. Great variations are observed, for example with temperature. The observed time to emergence (from egg to immature adult) for the population studied in this paper varied between 10 and 16 days.

Since the sex ratio is close to 1:1, we shall assume equal numbers of males and females in the models, Mackerras (1933). The development of the reproductive organs in the female is dependent on diet: sufficient protein is required for egg development. Egg laying in sexually mature females occurs in bursts several days apart, with the interval increasing as the female ages. The death rates for the larval and pupal stages are low. Table 3 in the Appendix gives an indication of the dependence of the adult mortality on age.

We turn now to the specific details of Nicholson's L97 cage I experiment. On 19 May 1954 1000 pupae were set up in a perspex box with a balsa wood grid on top of them to retain pupal cases. Food consisted of lump sugar and moistened cotton wool pad. Practically all flies emerged overnight. Adult food consisting of .4 gram ground liver, dried in a desicator, was added to the cage on 20 May. This quantity was added daily. Measurements were first made on 21 May. The basic data recorded were total counts of emerged or dead flies at two day intervals and the dates when the emerged flies had been laid as eggs. The experiment continued until 10 May 1956. Table 3 provides the bidaily data on adult population size, adults emerging, eggs laid and adults dying. Nicholson's final count of the flies in the cage was 6701. (This differs from the figure of 6806 derived by accumulating emergences and deaths.)

Figure 1 presents graphs of the square roots of the basic data.* The adult population is seen to oscillate dramatically with a period of approximately 35-40 days. The population was maintained on a restricted protein diet. The competition for food when the population size was large meant that the females did not receive enough protein to realize their maximum fecundity. Indeed a comparison of adult population with eggs laid, shows that at the beginning of the experiment virtually no eggs were laid when the population size exceeded a certain level. The few eggs laid meant that the population would soon drop sharply. The subsequent generation however, being smaller, faced less intense competition for food and their fecundity increased. An alternation of large and small generations resulted and hence the oscillations evident in the Figure.

From the Figure the egg, emergence and death series also oscillate with periods of 35-40 days, the egg series leading the adult series by 12-14 days; emergences lead adults by 0-2 days and deaths lag adults by 0-4 days. These intervals are in accord with the averages indicated in table 1.

3. INITIAL DATA ANALYSES. Foremost in Nicholson's design of experiment I was the notion that competition was to occur in egg laying only, death rates were to depend on age alone. Figure 2 presents graphs of the birth and death rates and of the proportion of eggs not emerging.

*The square root is graphed to provide more nearly constant variability in the series.

The latter series fluctuates about the level of 10 per cent throughout most of the series (the various spikes occur when the number of eggs is small). However, towards the end there is a definite rise, suggesting that Nicholson's experiment might be collapsing.*

The birth rate series varies immensely with the population series, just as Nicholson arranged. During the last 200 days there is a clear suggestion that the flies are now laying eggs at population sizes which earlier would have inhibited egg laying. Nicholson (1960) regards this as evidence that selection had occurred favoring flies able to lay eggs at higher population levels (i.e. lower protein). (Later in the paper it will be seen that the population is younger at the later stages so that this rise may be caused by younger flies being more fecund. We remark that at the end of this experiment, Nicholson carried out a further experiment and found that these flies could lay eggs with much less protein.)

The simple death rate series is also quite variable. This is expected because of the varying age composition. However, close examination of the upper two graphs of Figure 2 suggests that the death rate is generally higher when the population size is large. Figure 3 is a scatter diagram of death rate and population size. It presents a clear indication that the adult death rate is indeed density dependent —contrary to Nicholson's design. Later we will construct a model of the dependence of the death rate on age and population size.

4. A DETERMINISTIC APPROACH TO AN AGE STRUCTURED POPULATION. We consider the case of discrete time, $t = 0, 1, 2, \ldots$ Let

N_t = the adult population size at time t.

$\underset{\sim}{N}_t$ = the population vector. The entry in row i gives the number of population members aged $i-1$ at time t, $i=1, \ldots, I$.

E_t = the number of entrants (pupae emerging), aged 0, to the adult population in the time period $(t-1, t)$.

$\underset{\sim}{E}_t$ = the entrant vector, having E_t in row 1 and 0 in the other rows.

*Nicholson (1960) observed that the flies left at the end of the experiment laid eggs with a much higher proportion never emerging than the original strain of flies. It may be this that causes the rise.

D_t = the number dying in the time period (t-1, t).

$\underset{\sim}{P}(\underset{\sim}{N})$ = the survival matrix. If the population vector is $\underset{\sim}{N}$, then the entry in row i+1, column i gives the proportion surviving from age i to age i+1. The remaining entries are 0.

The population variation is described by

$$\underset{\sim}{N}_{t+1} = \underset{\sim}{P}(\underset{\sim}{N}_t)\ \underset{\sim}{N}_t + \underset{\sim}{E}_{t+1}, \underset{\sim}{N}_0 = \underset{\sim}{0}. \tag{1}$$

The population size at time t is

$$N_t = \underset{\sim}{1}'\ \underset{\sim}{N}_t \tag{2}$$

with $\underset{\sim}{1}$ the unit vector.

The complete trajectory of the population may be projected from the sequence of entrants, E_t, if the survival matrix $\underset{\sim}{P}(.)$ is given. The death series is given by

$$D_t = N_{t-1} - N_t + E_t. \tag{3}$$

These equations do not represent a complete system for determining future population sizes from current ones. We still need an expression for tomorrow's entrants E_{t+1} in terms of the population history:

$$E_{t+1} = f(\underset{\sim}{N}_t, \underset{\sim}{N}_{t-1}, \ldots). \tag{4}$$

Nicholson's assumptions about his experiments correspond to the hypotheses that $\underset{\sim}{P}(\underset{\sim}{N}_t)$ is a constant matrix and that $f(\underset{\sim}{N}_t)$ is a nonlinear function which tends to zero with increasing population size.

Models of this kind for experiments similar to Nicholson's were studied by Wu (1976). He conducted population experiments with blowflies which attempted to realize a model in which competition was within a single cohort during the larval stage of the insect. He measured, independently, the mortality parameters and the function (4) describing emergences for the blowflies on which his experiments were conducted. Wu carried out simulations and found them to present a similar aspect of periodicity and irregularity as the experiments. Indeed, in these experiments there is good quantitative agreement between the data and the model.

With Nicholson's experiments, estimates of the model parameters must be based upon the population data itself and independent experiments testing the model's assumptions cannot be done with the same flies. Ipaktchi et al (1980) have investigated deterministic models for Nicholson's data. They estimated the model parameters by no criteria other than producing final simulations which gave qualitatively the correct visual appearance. In this paper we attempt to investigate Nicholson's data in a more systematic fashion.

Note that models can produce simulations which appear irregular or "chaotic" despite the fact that they are deterministic (Guckenheimer, et al, 1977). Thus it is a reasonable hypothesis that the aperiodicity of Nicholson's data is primarily a consequence of deterministic causes. The underlying dynamical system described by equations (1) and (4) seems to be one which possesses "sensitivity to initial conditions": individual trajectories diverge from one another with increasing time. Future asymptotic behavior can be predicted only in a statistical sense. Our eventual aim is to assess the validity of this hypothesis, but adequate tests to do so have yet to be developed. We begin here by examining the statistics of Nicholson's data in order to determine whether his assumptions are consistent with the experimental results. Our analysis will show that modifications to the deterministic model should be made to allow for density dependent mortality.

In the remainder of the paper we shall focus upon the relationship between the death series D_t and the population series N_t. This is an important step in reconstructing the age structure $\underset{\sim}{N}_t$ of the population from the experimental data E_t, D_t. Without a model for the mortality of individuals of different ages, one cannot recover an estimate of the trajectory of the population through the phase space of the deterministic model. The aggregate data for N_t must be split into age classes to obtain $\underset{\sim}{N}_t$. As noted above, Nicholson assumed that adults were well supplied with the resources of survival so that per capita mortality should have been independent of population size. This assumption is embodied in a deterministic model with constant matrix $\underset{\sim}{P}$. Let us turn now to an analysis of the data via stochastic models which call into question this independence of $\underset{\sim}{P}$ on $\underset{\sim}{N}_t$.

5. A STOCHASTIC APPROACH. In any real population the age structure vector $\underset{\sim}{N}_{t+1}$ will not be uniquely determined by values of E_{t+1} and $\underset{\sim}{N}_t$, as equations (1) and (4) imply. Rather it is appropriate to replace equation (1) by

$$\underset{\sim}{N}_{t+1} = \underset{\sim}{P}(\underset{\sim}{N}_t)\underset{\sim}{N}_t + \underset{\sim}{E}_{t+1} + \underset{\sim}{\epsilon}_{t+1} \tag{5}$$

with $\underset{\sim}{\epsilon}_{t+1}$ an error variate. Supposing $E(\underset{\sim}{\epsilon}_{t+1}|\underset{\sim}{N}_t) = \underset{\sim}{0}$, and arguing conditionally on the emergence series, equation (1) is replaced by

$$E(\underset{\sim}{N}_{t+1}|\underset{\sim}{N}_t) = \underset{\sim}{P}(\underset{\sim}{N}_t)\underset{\sim}{N}_t + \underset{\sim}{E}_{t+1}.$$

The extent to which the population will follow a trajectory determined by iterating equation (1) will depend on the stochastic variability of $\underset{\sim}{\epsilon}_t$.

Given only the data for total emergences and total adult population, (E_t, N_t), $t = 1, \ldots, T$ we shall now present a method for forecasting the population size and estimating both the age and density dependent survival matrix and the population age-structure vector $\underset{\sim}{N}_t$. We emphasize that the method is not restricted to stationary age distributions.

Suppose that the survival matrix depends on $\underset{\sim}{N}$ only through the population size $N = \underset{\sim}{1}'\underset{\sim}{N}$. Then the model takes the form

$$\underset{\sim}{N}_{t+1} = \underset{\sim}{P}(N_t)\underset{\sim}{N}_t + \underset{\sim}{E}_{t+1} + \underset{\sim}{\epsilon}_{t+1} \tag{6}$$

$$N_t = \underset{\sim}{1}'\underset{\sim}{N}_t. \tag{7}$$

In control systems engineering terminology equation (6) is called the state equation and (7) the observation equation. The methodology of that field suggests an approach to the problems of forecasting and estimation.

Were the matrix $\underset{\sim}{P}(\underset{\sim}{N}_t) = \underset{\sim}{P}_t$ nonrandom and the process $\underset{\sim}{\epsilon}_t$ Gaussian white noise with covariance matrix $\underset{\sim}{V}_t$, then one could use the Kalman-Bucy filter:

$$E\{\underset{\sim}{N}_{t+1}|N_u, u \leq t+1\} = \underset{\sim}{m}_{t+1}$$

$$= \underset{\sim}{P}_t\underset{\sim}{m}_t + \underset{\sim}{E}_{t+1} + \underset{\sim}{A}_t\underset{\sim}{1}(N_{t+1} = E_{t+1} - \underset{\sim}{1}'\underset{\sim}{P}_m)/\underset{\sim}{1}'\underset{\sim}{A}_t\underset{\sim}{1},$$

$$\underset{\sim}{\chi}_{t+1} = \underset{\sim}{P}_t\underset{\sim}{\chi}_t\underset{\sim}{P}'_t - \underset{\sim}{A}_t\underset{\sim}{1}\underset{\sim}{1}'\underset{\sim}{A}'_t/\underset{\sim}{1}'\underset{\sim}{A}_t\underset{\sim}{1}, \tag{8}$$

with

$$\underset{\sim}{m}_0 = \underset{\sim}{0}, \underset{\sim}{\chi}_0 = \underset{\sim}{0}, \underset{\sim}{\chi}_t = \text{var}\{\underset{\sim}{N}_t|N_u, u \leq t\},$$

$$\underset{\sim}{A}_t = \underset{\sim}{V}_t + \underset{\sim}{P}_t\underset{\sim}{\chi}_t\underset{\sim}{P}'_t = \text{var}\{\underset{\sim}{N}_{t+1}|N_u, u \leq t\}. \tag{9}$$

(See Liptser and Shiryayev (1978), page 66.)

The above expressions assume $\underset{\sim}{P}_t$, $\underset{\sim}{V}_t$ to be known. In the case that they are unknown but can be parametrized so that the model is identifiable, the parameters may be estimated by maximizing the log likelihood

$$- 1/2 \sum_{t=2}^{T} \log \sigma_t^2 - 1/2 \sum_{t=2}^{T} (N_t - \underset{\sim}{1}' \underset{\sim}{P} \underset{\sim}{m}_{t-1} - E_t)^2 / \sigma_t^2 + C \quad (10)$$

with C a constant, $\sigma_t^2 = \underset{\sim}{1}' \underset{\sim}{A}_{t-1} \underset{\sim}{1} = \text{var}\{N_t \mid N_u, u \le t-1\}$ subject to the conditions (8), (9). (See for example Gupta and Mehra (1974).)

Theorem 13.4 of Liptser and Shiryayev (1968) shows that the expressions (8), (9) continue to hold when $\underset{\sim}{P}_t = \underset{\sim}{P}(N_t)$; $\underset{\sim}{V}_t = \underset{\sim}{V}(N_t)$. However with respect to the problem under consideration in this paper, a substantial departure from the assumptions is caused by the undoubted variation in $\underset{\sim}{V}(.)$ with $\underset{\sim}{N}_{t-1}$ not just with N_{t-1}. (The departure from a Gaussian distribution is expected to be less important.) Below, a bootstrap procedure will be developed for dealing with this difficulty.

We have assumed that the survival matrix depends only on the population size N. Now, suppose that its functional form is known up to a finite dimensional parameter, θ. Specifically write it as $\underset{\sim}{P}(N; \theta)$.

The updating equation (8) is not appropriate since $\underset{\sim}{V}(.)$ depends on the full population vector $\underset{\sim}{N}$. An intuitively reasonable alternative is

$$\underset{\sim}{m}_{t+1} - \underset{\sim}{E}_{t+1} = \underset{\sim}{P}(N_t; \theta) \underset{\sim}{m}_t (N_{t+1} - E_{t+1}) / (\underset{\sim}{1}' \underset{\sim}{P}(N_t; \theta) \underset{\sim}{m}_t) \quad (11)$$

wherein the projected values are updated proportionately, to yield the measured total N_{t+1}. Equation (11) may be obtained from (8) by choosing $\underset{\sim}{A}_t \underset{\sim}{1}$ proportional to $\underset{\sim}{P}(N_t) \underset{\sim}{m}_t$. (Were the entries of ($\underset{\sim}{N}_{t+1}$ given N_u, $u \le t$) independent Poisson variables then $\underset{\sim}{A}_t \underset{\sim}{1} = \underset{\sim}{P}(N_t) \underset{\sim}{m}_t$.)

Empirical evidence, to be presented later, suggests that it is reasonable to take σ_t in equation (10) proportional to N_{t-1}. Hence maximizing expression (10) comes down to minimizing

$$\sum_{t=2}^{T} (N_t - \underset{\sim}{1}' \underset{\sim}{P}(N_{t-1}; \theta) \underset{\sim}{m}_{t-1})^2 / N_{t-1}^2$$

or equivalently

$$S = \sum_{t=2}^{T} (D_t - \sum_{i=i}^{I} q_{i-1} (N_{t-1}; \theta) m_{i-1, t-1})^2 / N_{t-1}^2 \quad (12)$$

where $q_i(N;\theta)$=Prob {individual aged i, dies aged i given population size N}.

The data fitting procedure we employ below can be described as follows:

1) Evaluate the $\underset{\sim}{m}_t$ via expression (11) employing a trial value of θ.

2) Determine an estimate of θ by minimizing expression (12) using the evaluated $\underset{\sim}{m}_t$.

3) Take the estimate of θ as the new trial value in 1). Iterate until the trial value and the estimate becomes the same.

The values m_t evaluated via expression (11) at the final stage, provide estimates of the age structure of the population throughout the time period of observation. They will be used to fit an age and size dependent fertility function.

6. RESULTS.Two death rate models were fitted to the adult data:

(a) an additive model

$$q_{x,\ N,\ N\_}(\theta) = \alpha_x + \beta N + \gamma N\_ \qquad (13)$$

and (b) a multiplicative model

$$q_{x,N,N\_}(\theta) = 1 - (1-\alpha_x)\ (1-\beta N)\ (1-\gamma N\_). \qquad (14)$$

Here x is age, N is the population size at beginning of the current interval and N_ the population size at the beginning of the preceding interval. The first model was selected because of its simplicity, and the second because it corresponds to independent age and size mortality forces operating on the flies.

The parameter values $\beta, \gamma = 0$ correspond to the case of no density dependence. If $\alpha_x = \alpha$ for all x, then age plays no role in mortality.

Figures 4 and 5 show $q_{x,N,N\_}(\theta)$ for the two models in the case of a small population N, N_ = 1, a large population N, N_= 8000, an increasing population N = 5000, N = 2000 and a decreasing population N = 2000, N_= 5000.

In terms of the criterion (12), the product model (14) provides the better fit, (see Table 2), however the lower three curves of the two figures are almost identical.

The dominant feature is the apparent substantial dependence of death rate on the population size; contradicting Nicholson's desired experimental setup.

Table 2 provides a listing of various of the models that were fit to the data. The S corresponding to "persistence" is defined as

$$\sum_{t=2}^{T} (D_t - D_{t-1})^2 / N_{t-1}^2 \, .$$

If one views the problem as one of developing a predictor for the death series, this S measures the adequacy of predicting by the most recent value. The case of α_x corresponds to (Nicholson's desired) dependence of mortality on age alone. The case $\alpha + \beta N$ corresponds to random age-independent mortality.

Table 2

S for various models

$q_{x, N, N_{-1}}$	S
persistence	13.964
α	7.011
α_x	6.688
$\alpha + \beta N$	5.605
$1-(1-\alpha_x)(1-\beta N)$	3.894
$\alpha_x + \beta N$	3.843
$\alpha_x + \beta N + \gamma N_{-1}$	3.609
$1-(1-\alpha_x)(1-\beta N)(1-\gamma N_{-1})$	3.531
ARMA (1,1)	3.001

10 age intervals

7. DIAGNOSTICS. Figure 6a provides a graph of the observed death series, D_t and the fitted series $\hat{D}_t$ (one-step predictor), based on the product model of the previous section. The two series are generally quite close together. Even split peaks are traced fairly accurately.

Figure 6b plots the values $(D_t - \hat{D}_t)/N_{t-1}$ of weighted residuals between the model and the data. This graph makes it clear that a fair amount of autocorrelation remains in the error. If the model fit the data very well, such autocorrelation would not be present. From the stand-point of prediction there is information remaining in the data that has not been utilized.

Autoregressive moving average models were fit to this residual series and an ARMA (1, 1) scheme was found to fit fairly well. S was further reduced by this procedure as shown in Table 2.

Figures 7 and 8 provide scatter diagrams of the values $(D_t-\hat{D}_t, N_{t-1})$ and $((D_T-\hat{D}_t)/N_{t-1}, N_{t-1})$. The wedging apparent in Figure 7 suggests strong dependence of the residual variance on the population size. This dependence is still present when the residuals are normalized by $\sqrt{N_{t-1}}$. It is not apparent in Figure 8, suggesting that the standard error may be taken as proportional to the population size. This remark is the source of the N_{t-1}^2 term in S of (12). Parenthetically, were the variation in the numbers dying multinomial, then one would have

$$\text{var}\{D_t \mid \underset{\sim}{N}_t\} = \sum_x N_{t,x}\, q_x (1-q_x).$$

These values, when estimated, are considerably smaller than the empirical values $(D_t-\hat{D}_t)^2$. The extra variation in the population may be ascribed to, among other things, the variable sex ratio in the population and the inherent variability between individuals of the population. Taking note of the ages of individuals and the population sizes of the preceding two time periods leaves a fair amount of variation.

8. DISCUSSION. A variety of things become possible once the death probability function q has been estimated. Foremost among these is the construction of the age structure $\hat{N}_{x,t}$ = the estimated number aged x at time t. These values are important, for example,in understanding the dependence of birth intensity on age and population size. This was the principal focus of Nicholson's experimental design in this case.

Figures 9a, b show the $\hat{N}_{x,t}$: initially each of these series oscillates - as successive cohorts march along in time. Eventually the numbers in the older age classes appear to stabilize somewhat. Further, it is clear that in the later stages of the experiment very few insects were surviving to a very old age. Note that the population is not becoming stable.

Figure 10 makes this last remark more apparent. It is a graph of the average age of the population as a function of time. After an initial period of strong oscillations, this function settles down; moreover the population is apparently becoming younger.

It was remarked earlier that the flies were apparently being selected to lay eggs at low protein levels, as the experiment progressed. Figure 10 shows that it is also possible that this change in egg laying behavior might have been due to younger flies being more effective at egg laying.

9. FURTHER WORK. Quite a number of interesting problems remain to be studied. These include:

a) the statistical properties of the estimates,

b) estimation of the birth rate function and other parameters of the complete life cycle,

c) the relevance, if any, of the deterministic dynamic behavior of the mean values,

d) the statistical properties of forecasted population sizes based on the fitted model,

e) the form of the optimal controller (timing, amount, age-dependence),

f) the improvement, if any, resulting from the use of alternate estimates of the age structure (e.g. based on all, not just past data),

g) developing nonparametric estimates of the dependence of the vital rates on age and population size.

ACKNOWLEDGEMENT.
Professor Don McNeil of Macquarrie University, NSW, Australia provided us with Nicholson's basic data, that was used in the study.

APPENDIX

Table 3

EGGS	NONEMERGING EGGS	EMERGING	DEATHS	TOTAL
0.	0.	948.	0.	948.
0.	0.	4.	10.	942.
0.	0.	0.	31.	911.
0.	0.	0.	53.	858.
2149.	121.	0.	57.	801.
4627.	260.	0.	125.	676.
4523.	281.	0.	172.	504.
603C.	458.	0.	107.	397.
2684.	120.	0.	149.	248.
3373.	176.	0.	102.	146.
446.	9.	1763.	108.	1801.
133.	2.	4487.	53.	6235.
17.	1.	1830.	2091.	5974.
56.	2.	7952.	5005.	8921.
58.	10.	1953.	4264.	6610.
C.	0.	2419.	3056.	5973.
6.	1.	1966.	2266.	5673.
25.	3.	132.	1930.	3875.
30.	1.	36.	1550.	2361.
0.	0.	16.	1025.	1352.
0.	0.	85.	211.	1226.
548.	30.	17.	331.	912.
461.	47.	0.	391.	521.
163E.	111.	5.	163.	363.
1524.	66.	41.	175.	229.
2338.	270.	10.	97.	142.
1473.	105.	0.	60.	82.
3287.	346.	494.	34.	542.
1367.	220.	424.	27.	939.
617.	52.	1541.	49.	2431.
936.	76.	1317.	61.	3687.
112.	4.	2016.	1160.	4543.
91.	8.	1496.	1504.	4535.
0.	0.	2898.	1992.	5441.
0.	0.	855.	1884.	4412.
0.	0.	469.	1859.	3022.
1.	1.	925.	1291.	2656.
47.	2.	522.	1211.	1967.
8.	1.	55.	727.	1295.
0.	0.	45.	425.	915.
77.	3.	0.	364.	551.
598.	37.	0.	238.	313.
6814.	513.	0.	146.	167.
1537.	91.	5.	77.	95.
1296.	48.	45.	47.	93.
451.	11.	2.	35.	60.
1863.	185.	47.	39.	68.
2408.	303.	5205.	14.	5259.
358.	29.	1496.	82.	6673.
847.	19.	1957.	3189.	5441.
14.	14.	855.	2309.	3987.
0.	0.	501.	1536.	2952.
62.	5.	1771.	1075.	3648.
4.	2.	1607.	1033.	4222.
39.	2.	674.	1007.	3889.
0.	0.	197.	1791.	2295.
0.	0.	700.	1486.	1509.
107.	0.	0.	581.	928.
172.	15.	16.	205.	739.
828.	69.	43.	216.	566.
1216.	108.	0.	183.	383.
2109.	123.	37.	146.	274.
3460.	251.	0.	82.	192.
2438.	229.	105.	71.	226.
3234.	175.	326.	33.	519.
1574.	130.	729.	24.	1224.
7445.	900.	1072.	60.	2236.
2019.	167.	1747.	165.	3818.
672.	62.	3794.	1404.	6208.
0.	0.	2576.	2788.	5996.
0.	0.	2220.	2427.	5789.
0.	0.	3248.	2385.	6652.
0.	0.	4808.	3521.	7939.
33.	5.	1052.	4123.	4868.

EGGS	NONEMERGING EGGS	EMERGING	DEATHS	TOTAL
26.	5.	1270.	2186.	3952.
32.	6.	95.	1335.	2712.
0.	0.	3.	981.	1734.
321.	10.	0.	510.	1224.
230.	25.	0.	521.	703.
392.	30.	18.	213.	508.
253.	16.	10.	152.	366.
324.	65.	19.	106.	279.
1369.	155.	27.	63.	243.
1252.	94.	120.	20.	343.
977.	92.	435.	17.	761.
1336.	48.	298.	34.	1025.
1450.	155.	247.	51.	1221.
2520.	287.	456.	77.	1600.
3057.	419.	897.	230.	2267.
302.	19.	1459.	436.	3290.
166.	19.	721.	540.	3471.
20.	4.	1258.	1092.	3637.
27.	3.	1045.	979.	3703.
70.	0.	2300.	1127.	4876.
4.	0.	1782.	1294.	5364.
38.	2.	1252.	1726.	4890.
5.	2.	232.	2093.	3029.
48.	3.	53.	1132.	1950.
115.	6.	23.	748.	1225.
364.	17.	50.	199.	1076.
1034.	65.	30.	201.	905.
1295.	110.	39.	172.	772.
2383.	144.	0.	144.	628.
2435.	188.	33.	188.	473.
2535.	104.	113.	47.	539.
1701.	110.	383.	97.	825.
1601.	152.	928.	51.	1702.
580.	62.	1201.	35.	2868.
460.	55.	2175.	570.	4473.
363.	18.	1668.	920.	5221.
0.	0.	2980.	1609.	6592.
106.	2.	1872.	2064.	6400.
80.	0.	1228.	2876.	4752.
0.	0.	461.	1692.	3521.
303.	16.	660.	1462.	2719.
755.	77.	165.	953.	1931.
536.	34.	16.	447.	1500.
802.	96.	106.	524.	1082.
688.	38.	78.	311.	849.
1244.	172.	0.	75.	774.
1662.	145.	247.	157.	864.
5050.	432.	623.	179.	1308.
926.	101.	479.	163.	1624.
3011.	224.	775.	175.	2224.
1461.	134.	678.	479.	2423.
526.	48.	1078.	542.	2959.
107.	10.	1504.	916.	3547.
136.	29.	4771.	1081.	7237.
0.	0.	612.	2631.	5218.
0.	0.	2869.	2776.	5311.
8.	3.	1149.	2187.	4273.
10.	2.	592.	1595.	3270.
418.	44.	224.	1213.	2281.
616.	28.	47.	779.	1549.
1319.	113.	3.	461.	1091.
396.	35.	0.	295.	796.
1636.	90.	3.	189.	610.
1249.	76.	10.	175.	445.
975.	76.	621.	172.	894.
813.	52.	668.	108.	1454.
699.	55.	865.	57.	2262.
1908.	147.	227.	126.	2363.
952.	62.	1724.	240.	3847.
477.	66.	1136.	1107.	3876.
206.	24.	855.	796.	3935.
121.	11.	746.	1202.	3479.
39.	3.	794.	858.	3415.
70.	6.	1466.	1020.	3861.
162.	15.	759.	1049.	3571.
3.	0.	636.	1094.	3113.
170.	16.	172.	966.	2319.
300.	19.	127.	816.	1630.
417.	25.	110.	443.	1297.
623.	49.	50.	486.	861.
1061.	93.	118.	218.	761.
1852.	172.	79.	181.	659.

EGGS	NONEMERGING EGGS	EMERGING	DEATHS	TOTAL
1766.	155.	147.	105.	701.
2681.	313.	143.	82.	762.
2577.	302.	524.	98.	1188.
2333.	293.	694.	104.	1778.
1383.	153.	758.	108.	2428.
5915.	422.	1661.	283.	3806.
1241.	135.	1443.	730.	4519.
38.	5.	2440.	1313.	5646.
108.	5.	1628.	2423.	4851.
10.	2.	2180.	1657.	5374.
2.	2.	1235.	1896.	4713.
65.	5.	4323.	1669.	7367.
128.	11.	2193.	2324.	7236.
6.	4.	806.	2797.	5245.
71.	1.	99.	1708.	3636.
336.	31.	45.	1264.	2417.
343.	23.	0.	1159.	1258.
991.	52.	1.	493.	766.
1709.	174.	61.	348.	479.
766.	51.	115.	192.	402.
749.	62.	2.	156.	248.
6267.	573.	87.	81.	254.
3750.	440.	397.	47.	604.
4393.	559.	783.	41.	1346.
1527.	153.	1057.	61.	2342.
5520.	635.	1161.	175.	3328.
756.	66.	1075.	804.	3599.
1166.	96.	1762.	1280.	4081.
1294.	94.	5589.	2027.	7643.
206.	17.	4573.	4297.	7919.
253.	18.	1412.	3233.	6098.
233.	47.	5092.	4294.	6896.
1024.	76.	1218.	2480.	5634.
185.	11.	834.	1334.	5134.
112.	11.	805.	1751.	4188.
521.	43.	843.	1562.	3469.
287.	26.	311.	1338.	2442.
650.	36.	46.	557.	1931.
1254.	63.	806.	947.	1790.
1977.	145.	431.	499.	1722.
976.	65.	161.	395.	1488.
2880.	234.	165.	237.	1416.
2505.	254.	463.	510.	1369.
1394.	132.	545.	248.	1666.
2233.	199.	1247.	286.	2627.
629.	37.	1888.	675.	3840.
683.	45.	938.	734.	4044.
397.	16.	2693.	1808.	4929.
725.	48.	1828.	1646.	5111.
14.	3.	1649.	3608.	3152.
7.	3.	1906.	596.	4462.
146.	13.	556.	936.	4082.
285.	12.	457.	1513.	3026.
694.	57.	492.	1929.	1589.
558.	31.	775.	289.	2075.
1221.	85.	176.	422.	1829.
901.	70.	16.	457.	1388.
1754.	166.	28.	267.	1149.
2066.	242.	260.	441.	968.
1311.	142.	476.	274.	1170.
452.	24.	611.	316.	1465.
3092.	317.	403.	192.	1676.
3747.	303.	1659.	260.	3075.
541.	31.	1470.	730.	3815.
1606.	115.	1943.	1119.	4639.
884.	69.	1130.	1345.	4424.
675.	44.	550.	2190.	2784.
1037.	74.	4482.	1406.	5860.
1804.	117.	1393.	1472.	5781.
765.	48.	1007.	1891.	4897.
733.	38.	1001.	1978.	3920.
981.	63.	1144.	1229.	3835.
581.	35.	907.	1124.	3618.
311.	34.	557.	1125.	3050.
1313.	119.	1550.	828.	3772.
946.	56.	757.	1012.	3517.
755.	51.	919.	1066.	3350.
951.	143.	814.	1146.	3018.
422.	8.	434.	827.	2625.
330.	16.	358.	571.	2412.
188.	13.	626.	817.	2221.
1643.	99.	1004.	606.	2619.

EGGS	NONEMERGING EGGS	EMERGING	DEATHS	TOTAL
1178.	90.	1126.	542.	3203.
1861.	110.	364.	861.	2706.
1717.	196.	909.	898.	2717.
1031.	90.	323.	865.	2175.
2973.	280.	196.	743.	1628.
2225.	254.	1507.	747.	2388.
1983.	243.	1959.	670.	3677.
1745.	174.	291.	812.	3156.
1321.	82.	2078.	962.	4272.
1942.	130.	1122.	1623.	3771.
433.	33.	2525.	1341.	4955.
2009.	194.	2016.	1387.	5584.
254.	19.	841.	2534.	3891.
248.	9.	1579.	1969.	3501.
2511.	228.	2121.	1186.	4436.
2404.	163.	1367.	1434.	4369.
1652.	138.	814.	1789.	3394.
2657.	186.	1771.	1296.	3869.
4169.	484.	405.	1352.	2922.
2525.	571.	165.	1244.	1843.
1907.	188.	1863.	869.	2837.
1498.	161.	2703.	850.	4690.
1197.	70.	1409.	980.	5119.
3238.	346.	2229.	1509.	5839.
476.	38.	3787.	4237.	5389.
819.	60.	1568.	1964.	4993.
1094.	59.	1234.	1781.	4446.
1238.	99.	2131.	1726.	4851.
1167.	84.	985.	1593.	4243.
2082.	309.	2216.	1839.	4620.
3168.	225.	1424.	1195.	4849.
1645.	252.	600.	1785.	3664.
3892.	642.	391.	1039.	3016.
4603.	535.	1060.	1195.	2881.
2717.	416.	1779.	839.	3821.
2162.	376.	1510.	1031.	4300.
1998.	192.	971.	1103.	4168.
3547.	407.	2668.	1390.	5446.
2825.	361.	2713.	2682.	5477.
4377.	376.	5117.	2015.	8579.
1189.	100.	1907.	2953.	7533.
1460.	155.	2606.	3255.	6884.
279.	28.	1108.	3865.	4127.
188.	29.	3394.	1975.	5546.
1190.	155.	2097.	1330.	6313.
450.	37.	3062.	2725.	6650.
2189.	259.	2320.	2666.	6304.
1276.	103.	1444.	2906.	4842.
2174.	291.	938.	1428.	4352.
2737.	254.	230.	1367.	3215.
2822.	281.	577.	1140.	2652.
2460.	274.	610.	932.	2330.
1793.	199.	1695.	902.	3123.
603.	39.	1226.	394.	3955.
509.	42.	1445.	906.	4494.
1194.	139.	1967.	1681.	4780.
2047.	218.	3093.	2120.	5753.
1601.	117.	2302.	2500.	5555.
631.	77.	1711.	1554.	5712.
1717.	125.	702.	1628.	4786.
1560.	203.	722.	1442.	4066.
2512.	197.	347.	1522.	2891.
2617.	278.	1502.	1123.	3270.
4033.	427.	2083.	949.	4404.
8103.	795.	1001.	1007.	4398.
7991.	745.	841.	1127.	4112.
10546.	1020.	1583.	1294.	4401.
5230.	504.	2725.	1347.	5779.
10857.	944.	2293.	1475.	6597.
3167.	330.	3782.	2288.	8091.
2164.	223.	6819.	3628.	11282.
1405.	114.	7638.	6474.	12446.
389.	21.	9769.	8503.	13712.
0.	0.	4728.	7423.	11017.
156.	21.	10346.	6680.	14683.
2.	0.	2216.	9641.	7258.
676.	36.	1314.	2377.	6195.
652.	47.	2166.	2399.	5962.
2622.	215.	377.	2126.	4213.
763.	105.	34.	1472.	2775.
2109.	155.	90.	1084.	1781.

EGGS	NONEMERGING EGGS	EMERGING	DEATHS	TOTAL
2293.	262.	72.	917.	936.
852.	76.	571.	609.	898.
1926.	114.	565.	303.	1160.
3428.	413.	2260.	262.	3158.
3465.	383.	577.	349.	3386.
3454.	341.	2192.	1031.	4547.
8258.	1011.	1475.	1199.	4823.
1711.	175.	1479.	1332.	4970.
1019.	96.	1624.	1654.	4940.
952.	65.	2206.	1353.	5793.
1767.	238.	4270.	2227.	7836.
1351.	121.	1793.	5172.	4457.
812.	78.	4769.	2325.	6901.
331.	18.	4361.	3071.	8191.
427.	40.	1340.	2765.	6766.
626.	47.	1277.	2878.	5165.
638.	92.	1162.	3408.	2919.
609.	79.	1273.	777.	3415.
1520.	152.	636.	620.	3431.
4004.	505.	630.	899.	3162.
3131.	394.	244.	881.	2525.
2882.	206.	572.	807.	2290.
2613.	370.	357.	692.	1955.
2543.	280.	625.	644.	1936.
4525.	432.	1053.	605.	2384.
4127.	844.	2930.	648.	4666.
1729.	300.	3606.	1053.	7219.
2899.	506.	2587.	1500.	8306.
3383.	778.	2232.	2511.	8027.
3367.	656.	1884.	2901.	7010.
2623.	510.	3451.	2312.	8149.
4005.	600.	3192.	2392.	8949.
3489.	516.	2070.	4914.	6105.
5893.	2704.	2259.	3040.	5324.
5907.	1955.	2610.	2168.	5766.
3671.	1522.	1942.	1494.	6214.
7.	0.	2994.	2201.	7007.
0.	0.	3250.	2103.	8154.
0.	0.	3510.	2615.	9049.
0.	0.	2319.	4485.	6883.
0.	0.	4967.	3747.	8103.
0.	0.	2483.	3783.	6803.

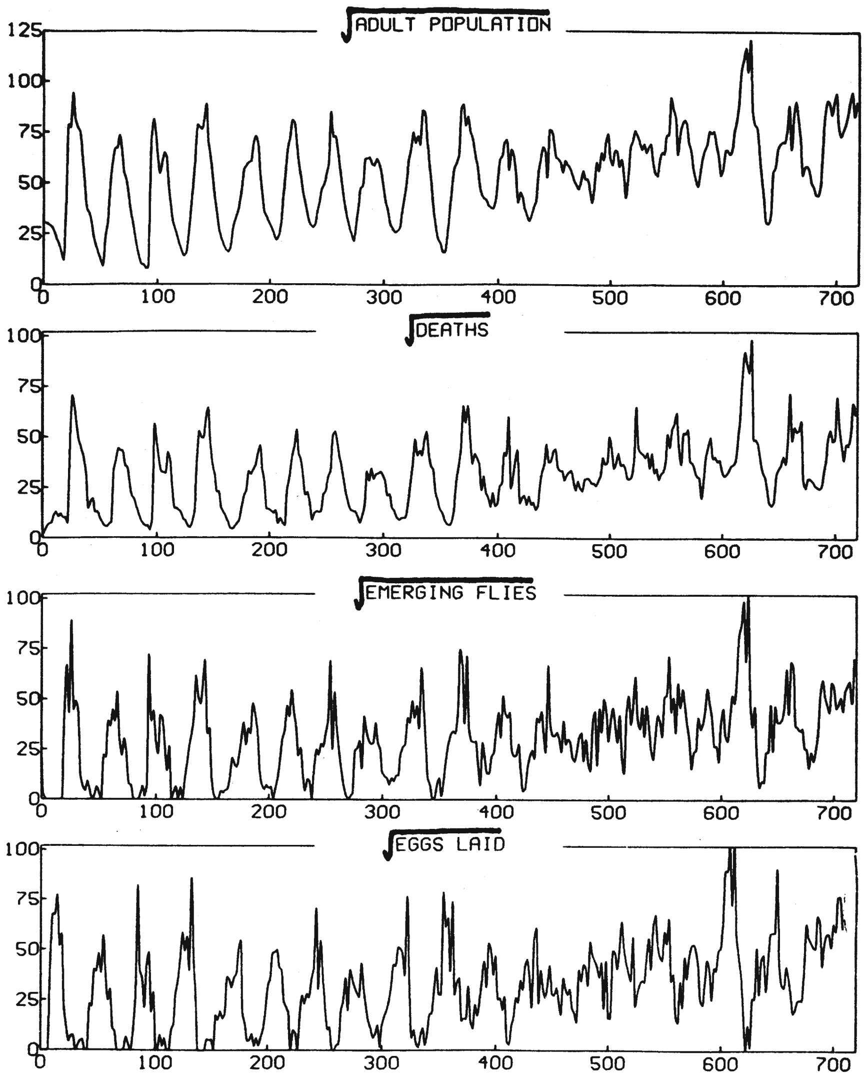

√ADULT POPULATION
125
100
75
50
25
0
0
100
200
300
400
500
600
700
√DEATHS
100
75
50
25
0
0
100
200
300
400
500
600
700
√EMERGING FLIES
100
75
50
25
0
0
100
200
300
400
500
600
700
√EGGS LAID
100
75
50
25
0
0
100
200
300
400
500
600
700

FIGURE 1

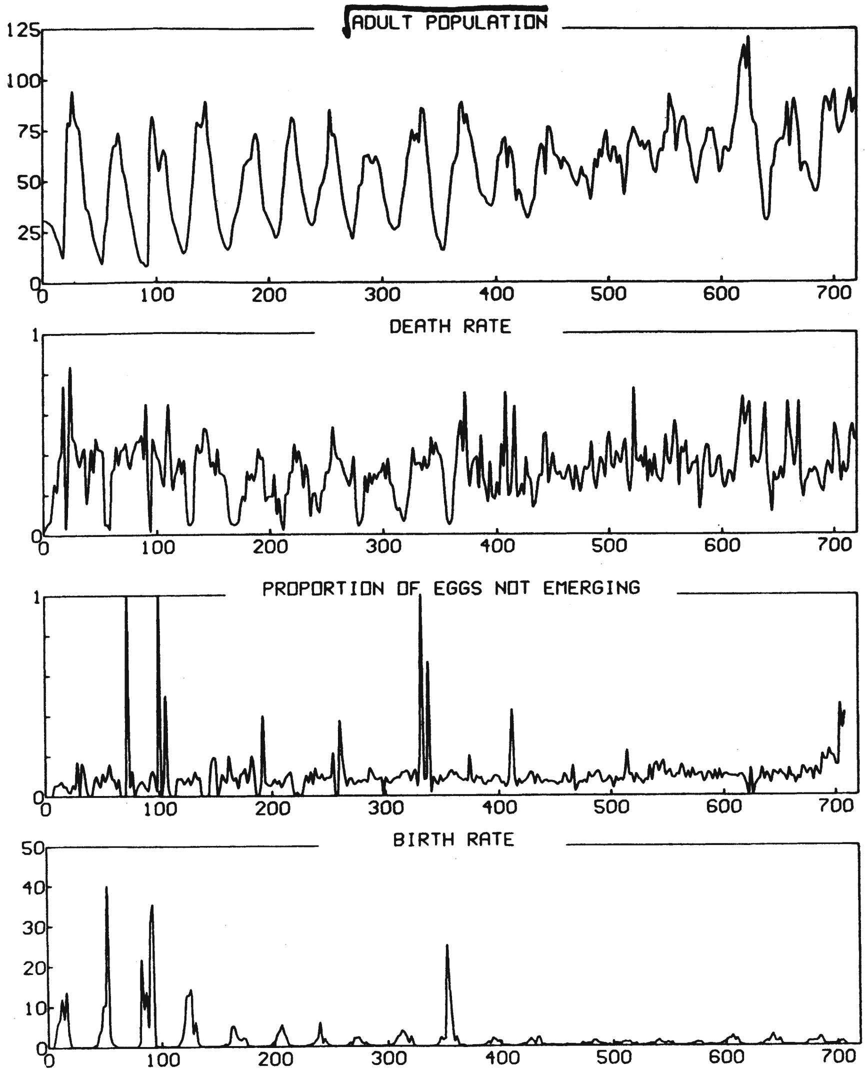

√ADULT POPULATION
125
100
75
50
25
0
0
100
200
300
400
500
600
700
DEATH RATE
1
0
PROPORTION OF EGGS NOT EMERGING
BIRTH RATE
50
40
30
20
10

FIGURE 2

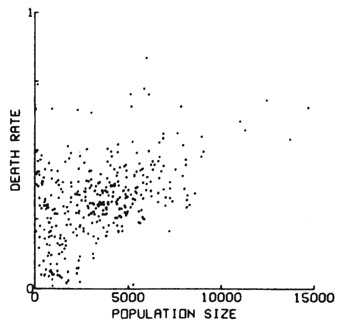


FIGURE 3

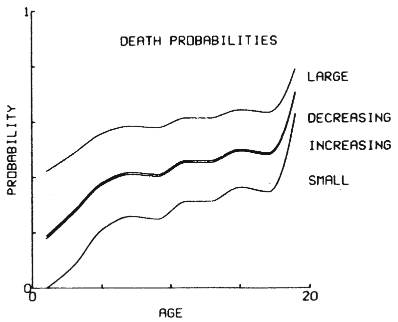


FIGURE 4

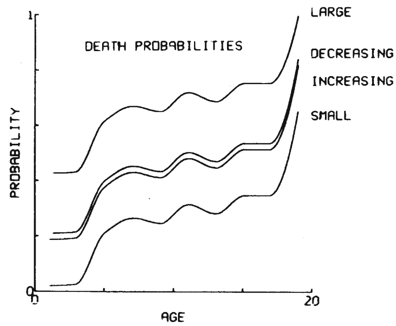


FIGURE 5

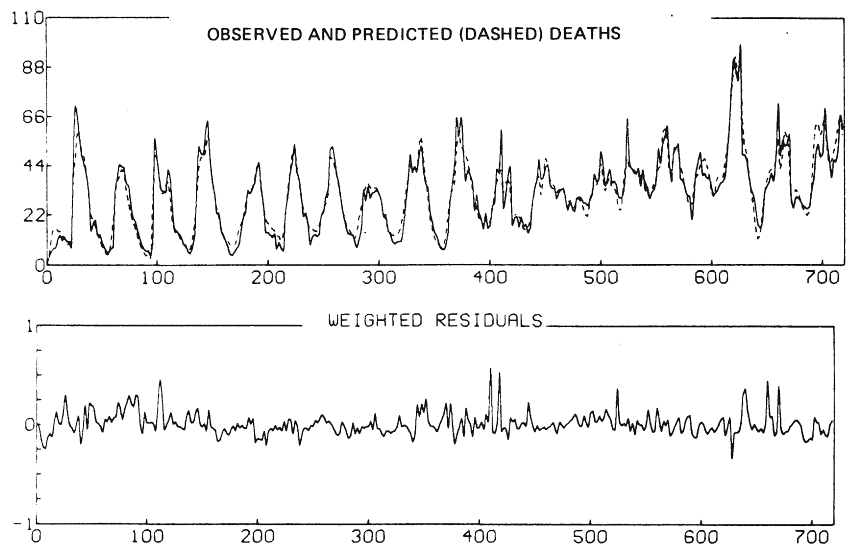


FIGURE 6

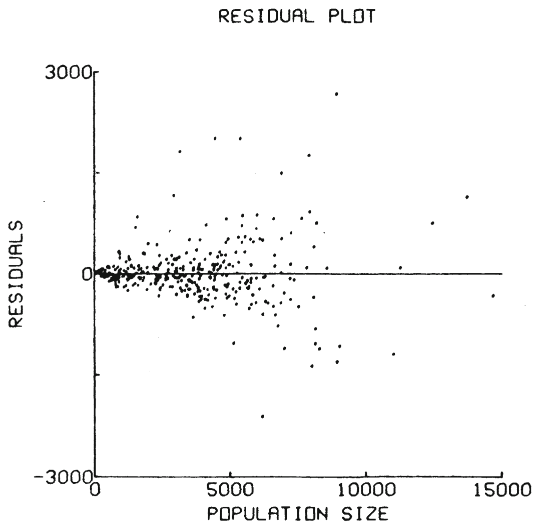


FIGURE 7

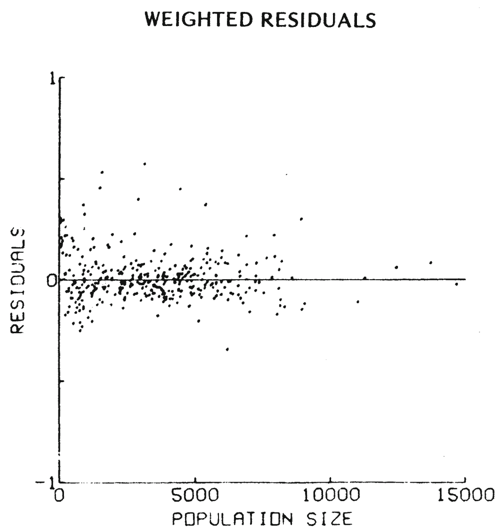


FIGURE 8

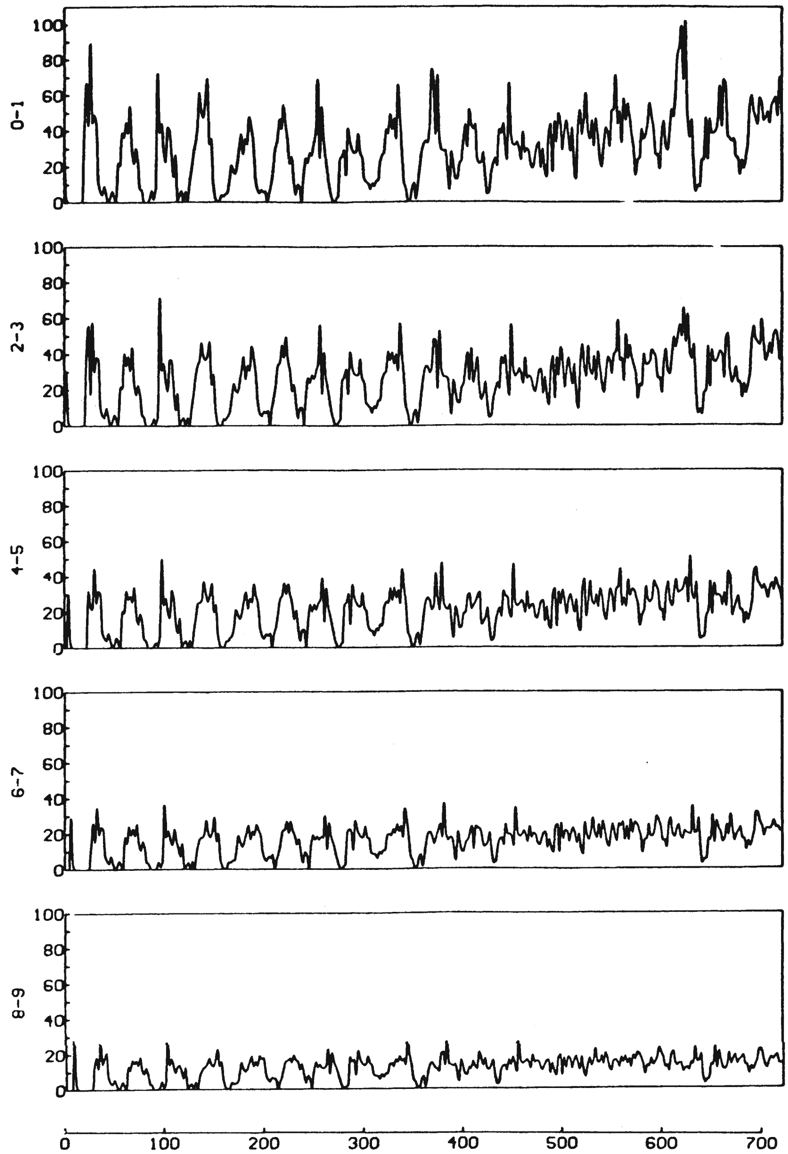


FIGURE 9a

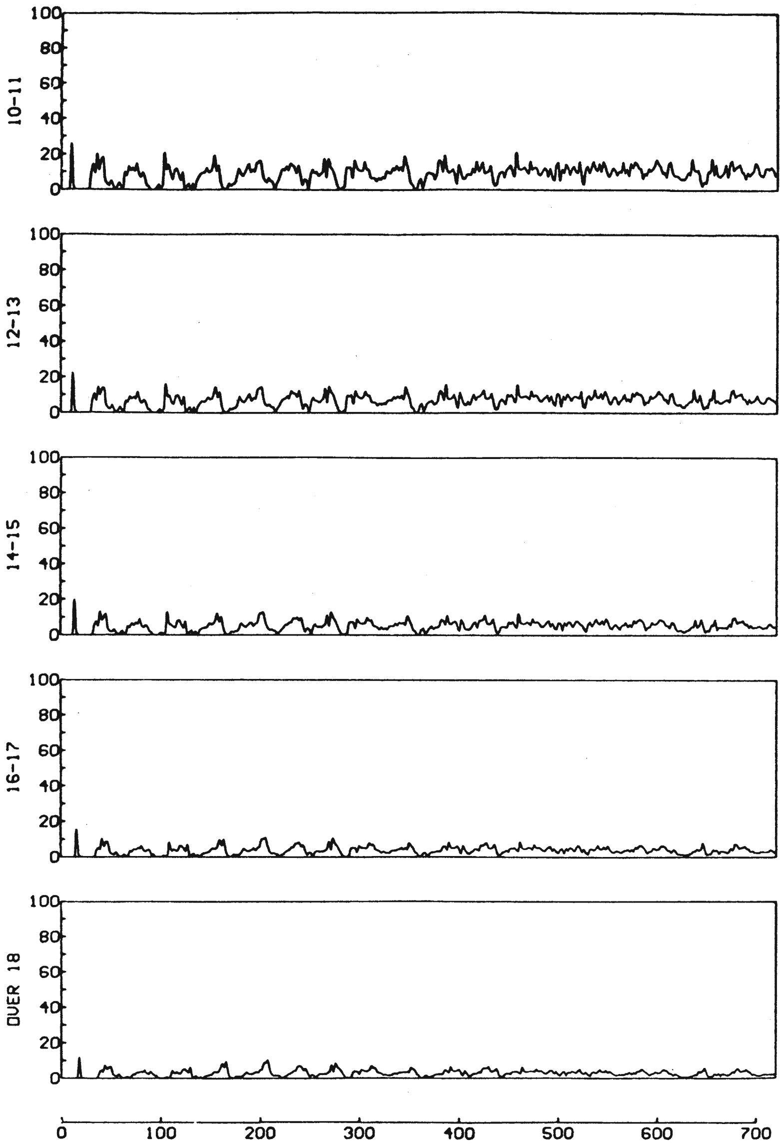

10-11
12-13
14-15
16-17
OVER 18
0
20
40
60
80
100
0
100
200
300
400
500
600
700

FIGURE 9b

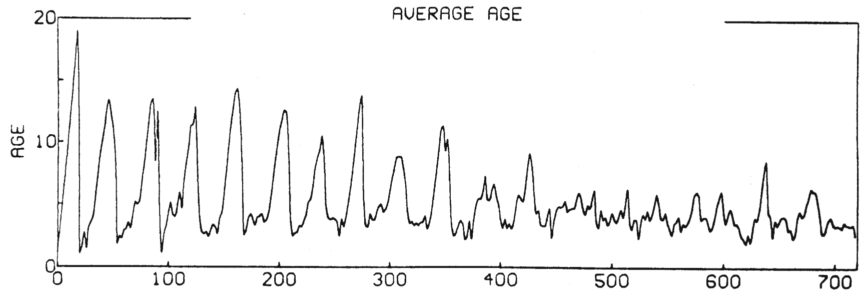


FIGURE 10

BIBLIOGRAPHY

Auslander, D.M., Oster, G.F. and Huffaker, C.B. (1974), "Dynamics of interacting populations", J. Franklin Inst. 297, 345-376.

Bartlett, M.S. (1966), "Stochastic Processes", Cambridge, Cambridge.

Chiang, C. L. (1968), "Introduction to Stochastic Processes in Biostatistics", Wiley, New York.

Gani, J. (1975), "Processes stochastiques de population", Lecture Notes in Math., vol. 480, Springer, New York, pp. 189-293.

Guckenheimer, J., Oster, G. and Ipaktchi, A. (1977), "The dynamics of density dependent population models", Theor. Pop. Biol.

Gupta, N.K. and Mehra, R.K. (1974), "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations", IEEE Trans. Aut. Control AC-19.

Ipaktchi, A., Oster, G.F., and Auslander, D.M. (to appear).

Keiding, N. (1975), "Maximum likelihood estimation in the birth-and-death process", Ann. Stat. 3, 363-372.

Lipster, R.S. and Shiryayev, A.N. (1978), "Statistics of Random Processes. II", Springer, New York.

Mackerras, J. (1933), "Observations on the life-histories, nutritional requirements and fecundity of blowflies", Bull. Ent. Res. 24, 303-362.

Nicholson, A.J. (1957), "The self-adjustment of populations to change", Cold Spring Harbor Symp. Quant. Biol. 22, 153-173.

Nicholson, A.J. (1960), "The role of population dynamics in natural selection", Evolution After Darwin. I, S. Tax (ed.) Univ. Chicago Press, Chicago, pp. 477-520.

Oster, G. (1977), "Lectures in population dynamics" Modern Modeling of Continuum Phenomena, R. DiPrima (ed.), Amer. Math. Soc., Providence, R.I.

Pollard, J.H. (1973), "Mathematical Models for the Growth of Human Populations", Cambridge, Cambridge.

Wu, Y.C. (1976), "An experimental and theoretical study of population cycles of the blowfly, Phaenicia Sericata (Calliphorides), in a Laboratory Ecosystem", Ph. D. Dissertion, University of California, Berkeley.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY, CA. 94720

DIVISION OF NATURAL SCIENCES, UNIVERSITY OF CALIFORNIA, SANTA CRUZ, CA.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY, CA. 94720

DEPARTMENT OF ENTOMOLOGY, UNIVERSITY OF CALIFORNIA, BERKELEY, CA. 94720

Learning a Potential Function From a Trajectory

David R. Brillinger

Abstract—This letter concerns the use of stochastic gradient systems in the modeling of the paths of moving particles and the consequent estimation of a potential function. The work proceeds by setting down a parametric or nonparametric model for the potential function. The method is simple, direct, and flexible, being based on a linear model and the least squares. Explanatories, attractors, and repellors may be included directly. The large sample distribution of the estimated potential function is provided, under specific assumptions. There are direct extensions to updating, sliding window, adaptive, robust, and real-time variants. An example analyzing the path of an elk is presented.

Index Terms—Mobility model, monitoring, potential function, stochastic differential equation, stochastic gradient system.

I. Introduction

Location signals of moving objects, obtained for example by GPS or LORAN, have become common in practice. Typically, one has scattered positions along trajectories of the objects. The questions of how to summarize, how to predict, and how to simulate such movements arise. This happens particularly when a number of paths are involved or the path of an object is a tangle. See Fig. 1, which shows 1571 locations over a period of a month along the track of an elk in Starkey Project in Oregon. (Reference [1] provides the project's website address.)

This letter provides a unified approach for dealing with movement modeling and associated data. The fields in which movement data have arisen include animal tracking [2], [3] and soccer [4]. There are papers developing a statistical potential approach to tracks. These include [2], [3], and references therein. This letter provides some formal background missing in those papers, discussion, and an example.

Let $\mathbf{r}$ denote a point in R^p. (In the mathematical expressions below, all the vectors appearing are column vectors and set in boldface.) A potential function, $V(\mathbf{r})$, is a real-valued function of location. Its use can lead to simpler representations of motion than those based on modeling velocities directly. One can note that in the overdamped case, the equation of motion of a particle in the potential field, $V(\mathbf{r})$, is

$$d\mathbf{r}(t) = -\nabla V(\mathbf{r}(t))dt \qquad (1)$$

having assumed $V(\mathbf{r})$ differentiable and with ∇ denoting the gradient. [The negative sign in (1) is traditional.] The entity $d\mathbf{r}(t)/dt$ is called a vector field. When $p = 2$, the level surfaces

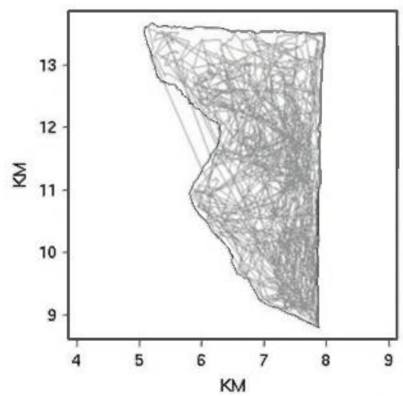


Fig. 1. Path of an elk around the NE pasture of the Starkey Experimental Forest in Oregon. Locations were estimated approximately every two hours and are joined by consecutive straight lines.

of the potential function are conveniently displayed in contour form and its gradient as arrows on a grid (see Figs. 2 and 3).

The estimation method to be presented can be motivated by stochastic gradient systems, that is, systems that can be written in the time invariant case as

$$d\mathbf{r}(t) = -\nabla V(\mathbf{r}(t))dt + \boldsymbol{\sigma}(\mathbf{r}(t))d\mathbf{B}(t) \qquad (2)$$

for some differentiable V with $\mathbf{B}(t)$ a p-dimensional Brownian motion and $\boldsymbol{\sigma}$ a p by p matrix. Expression (2) is a particular case of the stochastic differential equation (SDE)

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t))dt + \boldsymbol{\sigma}(\mathbf{r}(t))d\mathbf{B}(t). \qquad (3)$$

What distinguishes the traditional SDE work from the present study is that the drift term $\boldsymbol{\mu}$ here has the special form $-\nabla V$ for some real-valued function V. It will be seen that the modeling situation is simplified when such a V is assumed to exist.

II. Problem and Approach

The basic problem assumes the model (2) and seeks to learn $V(\mathbf{r})$ given data $(t_i, \mathbf{r}(t_i), i = 1, \cdots, n)$. These data will be viewed as locations at successive times, $\{t_i\}$, of an object moving along a trajectory of the process (2). One seeks both vector field and potential function estimates.

Supposing that $\nabla V(\mathbf{r})$ is a smooth function of $\mathbf{r}$, and that the observation times are close together, one can set down the following approximation to (2):

$$\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i) = -\nabla V(\mathbf{r}(t_i))(t_{i+1} - t_i) + (t_{i+1} - t_i)^{1/2}\boldsymbol{\sigma}\mathbf{Z}_{i+1} \qquad (4)$$

Manuscript received December 12, 2006; revised April 2, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Brian M. Sadler.

The author is with the Statistics Department, University of California Berkeley, Berkeley, CA 94720 USA (e-mail: brill@stat.berkeley.edu).

Digital Object Identifier 10.1109/LSP.2007.900032

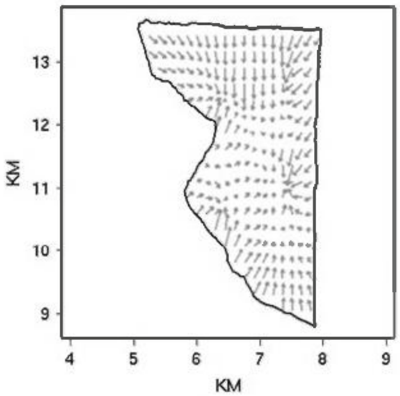


Fig. 2. Estimated vector field for the path of Fig. 1.

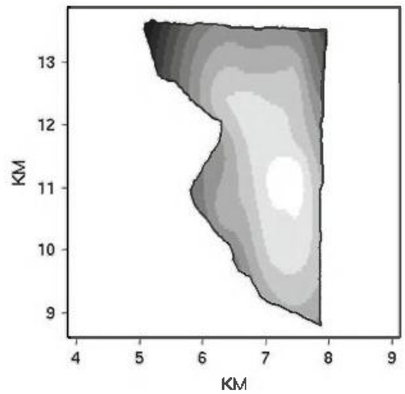


Fig. 3. Estimated potential for the path of Fig. 1. The lighter shading corresponds to smaller values.

for $i = 1, 2, 3, \cdots, n$, with σ a p by p matrix and with the Z_i independent p-dimensional variates having mean 0 and covariance matrix I. The reason for the multiplier $(t_{i+1} - t_i)^{1/2}$ is that for real-valued Brownian, $\text{Var}(dB(t)) = dt$.

The approximation

$$(\mathbf{r}(t_{i+1}) - \mathbf{r}(t_i))/(t_{i+1} - t_i) = \boldsymbol{\mu}(\mathbf{r}(t_i)) + \sigma Z_{i+1}/(t_{i+1} - t_i)^{1/2} \quad (5)$$

for the SDE (3) was employed in [2] and [3] for elk and movement and is employed in the example of this letter. In [2], an early attempt was made at estimating a potential function by numerical integration and simulation. The question was asked whether the vector field, $\boldsymbol{\mu}$, had the form $-\nabla V(\mathbf{r})$. This may be studied by comparing an unrestricted estimate of $\boldsymbol{\mu}$ with one assuming the existence of a potential function. The approach of papers [2] and [3] was informal.

III. Potential Functions

A basic issue is how to describe mathematically a potential function, $V(\mathbf{r})$, $\mathbf{r}$ in R^p. Suppose one exists. For introductory purposes in the development, suppose V is linear in a vector-valued parameter $\boldsymbol{\beta}$. Write $V(\mathbf{r}) = \boldsymbol{\phi}(\mathbf{r})^T\boldsymbol{\beta}$ with $\boldsymbol{\phi}$ an L by 1 vector of functions of known form and $\boldsymbol{\beta}$ an L by 1 unknown parameter. Examples of such a V follow. The gradient of V is the p by 1 vector $\nabla\boldsymbol{\phi}(\mathbf{r})^T\boldsymbol{\beta}$

Example 1: Polynomial expansion.

Consider $V(\mathbf{r}) = \sum \beta_m \mathbf{r}^m$, where $\mathbf{m} = (m_1, \cdots, m_p)$ and $\mathbf{r}^m = x_1^\wedge m_1 \ldots x_p^\wedge m_p$, with $\sum$ over $m_1, \cdots, m_p \geq 0$ and $1 \leq m_1 + \cdots + m_p \leq M$.

One could employ a trigonometric polynomial, a spline function, or a wavelet expansion here. Many functions, V, can be well approximated by taking M large. In practice, one might employ M_n with M_n increasing with n.

Example 2: Node based.

Consider nodal points $\mathbf{u}_l, l = 1, \cdots, L$ in R^p and set $V(\mathbf{r}) = \sum \beta_l K(\mathbf{r} - \mathbf{u}_l)$ for some real-valued differentiable kernel K. As a specific example of K, one has the radial basis thin plate splines [5, pp. 30–34]

$$K(\mathbf{r}) = |\mathbf{r}|^{2q-p}\log|\mathbf{r}| \text{for } p \text{ even}, = |\mathbf{r}|^{2q-p} \text{for } p \text{ odd}. \quad (6)$$

Here q denotes the order of differentiability of K, $2q - p > 0$, and $|\mathbf{r}| = (\mathbf{r}^T\mathbf{r})^{1/2}$. An expression like (6) leads to a smooth representation for V.

Example 3: Attraction and repulsion.

Consider a region A and a point $\mathbf{r}$ outside A. Potential functions can be set down leading to attraction or repulsion from A. Specifically, if one lets $d_A(\mathbf{r})$ denote the minimum distance from the point $\mathbf{r}$ outside A to A and sets $V(\mathbf{r}) = \beta d_A(\mathbf{r})^\alpha$, then for $\alpha > 0$, one has attraction to A and repulsion if $\alpha < 0$. One can reverse attraction and repulsion by changing the sign of d_A. It can be convenient to use $V(\mathbf{r}) = \beta_1 \log d_A(\mathbf{r}) + \beta_2 d_A(\mathbf{r})$ for similar purposes.

The functional forms of Examples 1–3 may be added together to provide other forms.

Reference [6] considers the observed trajectory of a monk seal near the island of Molokai employing the mixed function

$$V(\mathbf{r}) = \gamma_1 x + \gamma_2 y + \gamma_{11} x^2 + \gamma_{12} xy + \gamma_{22} y^2 + C/d(x, y) \quad (7)$$

where $\mathbf{r} = (x, y)^T$ is in R^2 and represents the location of the animal on the ocean surface. The value $d(x, y)$ is the distance from the location $\mathbf{r}$ to the nearest point on the island. The γ's and C are unknown parameters to be estimated. The final term in (7) keeps the seal off of the island. A different monk seal is studied in [7], and a different, now time dependent, potential function is employed

$$V(\mathbf{r}, t) = \alpha \log d(\mathbf{r}, t) + \beta d(\mathbf{r}, t)$$

with $d(\mathbf{r}, t)$ being the animal's distance from an attractor at time t. The attractor switched depending on whether the animal was on an outbound or an inbound journey.

Reference [4] studies the motion of a soccer ball during a very exciting World Cup moment. The potential function used is

$$\alpha \log d(\boldsymbol{r}) + \beta d(\boldsymbol{r}) + \gamma_1 x + \gamma_2 y + \gamma_{11} x^2 + \gamma_{12} xy + \gamma_{22} y^2$$

with $d(\boldsymbol{r})$ the shortest distance to the goalmouth from $\boldsymbol{r} = (x, y)^T$. The first two terms lead to attraction to the goalmouth and the remaining to general motion on the field.

Potential function and vector field estimates are provided in each of the papers just referenced. The function $V(\boldsymbol{r})$ is linear in the parameter, and least squares is employed as the estimation procedure in each case. The function could be nonlinear, and then, nonlinear least squares could be employed. Reference [8] develops asymptotic results pertinent to the nonlinear case. Alternately, the $\{\boldsymbol{Z}_i\}$ in (5) could be non-Gaussian and maximum likelihood estimation employed.

IV. Estimation

The representation (4) with $\boldsymbol{r}$ in R^p and $\nabla V(\boldsymbol{r}) = \nabla \boldsymbol{\phi}(\boldsymbol{r}_i)^T \boldsymbol{\beta}$ will be employed. The values $\boldsymbol{r}(t_i)$ will be written $\boldsymbol{r}_i$. Consider the p by 1 vector $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2}$. Following expression (5), the model has the form

$$(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2} = -\nabla \phi(\boldsymbol{r}_i)^T \boldsymbol{\beta}(t_{i+1} - t_i)^{1/2} + \boldsymbol{\sigma} \boldsymbol{Z}_{i+1})$$

$i = 1, \cdots, n-1$ involving the L by 1 vector $\boldsymbol{\beta}$, the L by p matrix $\nabla \boldsymbol{\varphi}(\boldsymbol{r}_i)$, the p by p matrix $\boldsymbol{\sigma}$, and the p by 1 vector $\boldsymbol{Z}_{i+1}$. Suppose $\boldsymbol{\sigma} = \sigma \boldsymbol{I}$ with σ positive and $\boldsymbol{I}$ the p by p identity matrix. Stack the $n-1$ values $(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)/(t_{i+1} - t_i)^{1/2}, i = 1, \cdots, n-1$ vertically to form the $(n-1)p$ by 1 array $\boldsymbol{Y}_n$. Stack the $n-1$ matrices $-\nabla \boldsymbol{\varphi}(\boldsymbol{r}_i)^T (t_{i+1} - t_i)^{1/2}$ to form the $(n-1)p$ by L matrix $\boldsymbol{X}_n$. Stack the $n-1$ values $\sigma \boldsymbol{Z}_{i+1}$ to form $\boldsymbol{\varepsilon}_n$. Then one has the regression model

$$\boldsymbol{Y}_n = \boldsymbol{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n \quad (8)$$

with the difference from ordinary regression that $\boldsymbol{Y}_n$ and $\boldsymbol{X}_n$ are statistically dependent. Using a generalized inverse, if necessary, one can compute an ordinary least-squares estimate $\boldsymbol{b} = (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n^T \boldsymbol{Y}_n$ of $\boldsymbol{\beta}$, and then, if $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta}$ is estimable, $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b}$ is a reasonable estimate of $V(\boldsymbol{r})$.

Supposing the individual entries of $\boldsymbol{\varepsilon}_n$ to be independent, zero mean, variance σ^2 variates, asymptotic properties of $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b}$ may be obtained from [9, Theorem 3]. The theorem is given in the Appendix.

Let y_j denote the jth row of $\boldsymbol{Y}_n$. Let $\boldsymbol{x}_j^T$ denote the jth row of $\boldsymbol{X}_n^T$. One can compute $s_n^2 = ((n-1)p)^{-1} \sum (\boldsymbol{y}_j - \boldsymbol{x}_j^T \boldsymbol{b})^T (\boldsymbol{y}_j - \boldsymbol{x}_j^T \boldsymbol{b})$ as an estimate of σ^2 and, for example, set down a confidence interval for $\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta}$ using the results of [9]. Specifically, provided lim log $\lambda_{\max}(\boldsymbol{X}_n^T \boldsymbol{X}_n)/n \to 0$ almost surely, one has $s_n \to \sigma$ almost surely and by a Slutsky Theorem

$$(\boldsymbol{\varphi}(\boldsymbol{r})^T (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{\varphi}(\boldsymbol{r}))^{-1/2} \boldsymbol{\varphi}(\boldsymbol{r})^T (\boldsymbol{b} - \boldsymbol{\beta})/s_n \to N(0, 1)$$

with $N(0, 1)$ the standard normal. This leads to the approximate $100(1 - \alpha)\%$ confidence interval

$$\boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{\beta} = \boldsymbol{\varphi}(\boldsymbol{r})^T \boldsymbol{b} \pm z_{\alpha/2} (\boldsymbol{\varphi}(\boldsymbol{r}) (\boldsymbol{X}_n^T \boldsymbol{X}_n)^{-1} \boldsymbol{\varphi}(\boldsymbol{r})^T)^{1/2} s_n$$

where $z_{\alpha/2}$ denotes the $100\alpha/2$ percent point of the standard normal. As mentioned in [9], one could use the F distribution to construct an approximate confidence region for a collection of values $\{\boldsymbol{\phi}(\boldsymbol{r}_k)^T \boldsymbol{\beta}\}$.

V. Example

The Starkey Project is a large area in Oregon set aside to study the interactions of elk, deer, cows, and man sharing an environment [1]. Fig. 1 shows a sampled trajectory of one of the elk in the NE Pasture. There were 1571 GPS locations and times of location obtained with a time interval of approximately two hours between successive locations. It is recognized that the theory connecting the sampled times case to the continuous time case expects the times to be close together. It is still anticipated that the discrete model studied is of interest in its own right and will provide results of practical use.

A potential function $V(\boldsymbol{r})$ was approximated by a thin plate radial basis spline employing the kernel function of (6) with p and $q = 2$, and $L = 36$. The x and y components of the $\boldsymbol{u}_l$ were taken to be the $100m/7, m = 1, \cdots, 6$ percentiles of the standardized x and y values. These values were chosen was for illustrative purposes.

The coefficients β_l were estimated by ordinary least squares employing the model (5) with $\boldsymbol{\sigma} = \sigma \boldsymbol{I}$. The results are provided in Figs. 2 and 3. One sees the confusion of Fig. 1 much reduced. A point of attraction appears near the point (7.5,11.0). When one looks at a topographic plot of elevations, the point of attraction appears to be a valley/canyon of sorts. Fig. 3 provides an image plot of the potential function. Now one sees the point of attraction immediately.

The confusion of Fig. 1 has been referred to. An empirical gradient plot is similarly confused.

VI. Extensions and Conclusion

Various generalizations of the letter's results may be mentioned. One could set down an expansion for V employing wavelet functions. One could consider updating methods for real-time work, e.g., those based on a Kalman filter. One could envisage a potential function as a spatial state variable and the paths of objects determined by the measurement equation. If the potential function is changing slowly, one could consider a sliding window estimate [10]. Estimates that are robust to non-normailty and resistant estimation can be considered. In video analysis, one might consider the model $\boldsymbol{I}(\boldsymbol{r}, t) = \boldsymbol{I}_0(\boldsymbol{r}) + \delta(\boldsymbol{r}(t) - \boldsymbol{r})$ with t indexing the video frames and δ the Dirac delta. The term I_0 represents a stationary background and $\boldsymbol{r}(t)$ the location of an object moving around in the scene [11].

This letter presents an estimation method for handling moving objects. The computations may be implemented by the least-squares algorithm. The model may be viewed as parametric or nonparametric.

APPENDIX

Because of the statistical dependence of the location of the object at time t_i on past locations, one needs special arguments to get the asymptotic distribution. For the simple cases of the letter, results based on martingale arguments are available in [9]. A result is [9, Theorem 3].

Theorem. Consider the regression model $y_j = \boldsymbol{x}_j^T\boldsymbol{\beta}+\varepsilon_j, j = 1, 2, \ldots$ with the $\{\varepsilon_j\}$ martingale differences with respect to an increasing sequence of σ-fields $\{F_N\}$. Suppose that $\sup_n E(\|\varepsilon_N\|^\alpha|F_{N-1}) < \infty$ almost surely for some $\alpha > 2$. Suppose further that $\lim_{n\to\infty} \text{var}(\varepsilon_N|F_{N-1}) = \sigma^2$ almost surely for some nonstochastic σ. Define $\boldsymbol{X}_N = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_N]^T$. Assume that $\boldsymbol{x}_j$ is a F_{j-1} -measurable random variable and that there exists a nonrandom positive definite symmetric L by L matrix $\boldsymbol{B}_N$ for which $\boldsymbol{B}_N^{-1}(\boldsymbol{X}_N^T\boldsymbol{X}_N)^{1/2} \to \boldsymbol{I}$, $\sup_{1\le i\le N}\|\boldsymbol{B}_N^{-1}\boldsymbol{x}_N\| \to 0$ in probability. Then as $N \to \infty$

$$(\boldsymbol{X}_N^T\boldsymbol{X}_N)^{1/2}(\boldsymbol{b}-\boldsymbol{\beta}) \to N(0, \sigma^2\boldsymbol{I})$$

in distribution.

Note that zero-mean independent observations like the successive entries of ε_n of (9) form a martingale difference sequence with respect to the σ-field generated by the preceding locations.

Reference [15] shows that under the further assumption, lim log $\lambda_{\max}$ $(\boldsymbol{X}_n^T\boldsymbol{X}_n)/n \to 0$ almost surely, one has $s_n \to \sigma$ almost surely.

ACKNOWLEDGMENT

The author would like to thanks A. A. Ager and H. K. Preisler for involving him with Starkey data sets and ideas that proved useful.

REFERENCES

[1] Starkey Project. [Online]. Available: http://www.fs.fed.us/pnw/starkey.

[2] D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie, "The use of potential functions in modelling animal movement," in *Data Analysis from Statistical Foundations*, A. K. M. E. Saleh, Ed. Huntington, NY: Nova, pp. 369–386. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/fraser8.pdf, (2001a).

[3] D. R. Brillinger, H. K. Preisler, A. A. Ager, J. Kie, and B. S. Stewart, Modelling Movements of Free-ranging Animals, Univ. California Berkeley Statistics, Tech. Rep. 610. [Online]. Available: http://www.stat.berkeley.edu/~brill/Preprints/610.pdf, (2001b).

[4] D. R. Brillinger, "A potential function approach to the flow of play in soccer," *J. Quant. Anal. Sports* vol. 3, no. 1, 2007. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/jqas.pdf.

[5] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1980.

[6] D. R. Brillinger, B. Stewart, and C. Littnan, "A meandering hylje," in Festschrift for Tarmo Pukkila on His 60th Birthday, E. P. Liski, J. Isotalo, S. Puntanen, and G. P. H. Styan, Eds., Dept. Mathematics, Statistics and Philosophy, Univ. Tampere. [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/hylje1.pdf, 2006a.

[7] D. R. Brillinger, B. Stewart, and C. Littnan, "Three months journeying of a Hawaiian monk seal," *Lecture Notes Statist.* [Online]. Available: http://www.stat.berkeley.edu/~brill/Papers/monk33.pdf 2006b.

[8] T. L. Lai, "Asymptotic properties of nonlinear least squares estimates in stochastic regression models," *Ann. Statist.*, vol. 10, pp. 1917–1930, 1994.

[9] T. L. Lai and C. Z. Wei, "Least squares estimation in stochastic regression models with applications to identification and control of dynamic systems," *Ann. Statist.*, vol. 10, pp. 154–166, 1982.

[10] M. Belge and E. L. Miller, "A sliding window RLS-like adaptive algorithm for filtering alpha-stable noise," *IEEE Signal Process. Lett.*, vol. 7, no. 4, pp. 86–89, Apr. 2000.

[11] V. Zlokolica, A. Pizurica, and W. Philips, "Noise estimation for video processing based on spatio-temporal gradients," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 337–340, Jun. 2006.

A Potential Function Approach to the Flow of Play in Soccer*

David R. Brillinger

Abstract

There is a growing literature on the statistical analysis of data from association-football/soccer games, seasons or groups of seasons. In contrast this paper is concerned with a single play, that is a sequence of successful passes. The play studied contained 25 passes and ended in a goal for Argentina in World Cup 2006. One question addressed is how to describe analytically the spatial-temporal movement of such a particular sequence of passes.

The basic data are points in the plane, successively joined by straight lines. The resulting figure represents the trajectory of the moving soccer ball. The approach of this study is to develop a useful potential function, a concept arising from physics and engineering. In particular the potential function leads to a regression model that may be fit directly by linear least squares.

The resulting potential function may be used for simple description, summary, comparison, simulation, prediction, model appraisal, bootstrapping, and employed for estimating quantities of interest. The purpose illustrated here is to simulate play in a game where the ball goes back and forth between two teams each having their own potential function.

KEYWORDS: Argentina, association football, exploratory data analysis, potential function, regression model, Serbia-Montenegro, simulation, soccer, vector field, World Cup 2006, 25-pass play.

*I wish to thank Tatyana Shepova of Online Media Technologies Limited for providing additional detail on the play, beyond those in the standard World 3D Cup 2006 Player package. I also wish to thank the Referees and the Editor for various pithy remarks. The work was supported by the NSF Grant DMS-200504162.

1. Introduction

The 2006 World Cup included some grand moments. One of the most spectacular was the 25-pass scoring play of Argentina in the Serbia-Montenegro (S-M) game on 16 June. The shot that ended the play was a goal scored by Cambiasso, but some 8 players worked hard to get the ball into position for his shot. The play has been described as: "one of the all time great World Cup goals", "the play of the tournament", "a joy forever", "a glorious goal", "mesmerizing", and "a string of pearls". Not long after the game, sketches of the path of the ball started to appear in newspapers (e.g. *Expressen*, Sweden) and magazines (e.g. *Cambio*, Columbia). Purposes of this paper are to develop an analytic description and model for such a play and to explore its uses, e.g. for simulation of plays where the ball changes sides.

The play began in the Argentine half of the field with Maxi passing back to Heinze. The sequence of players involved then was: Mascherano, Riquelme, Maxi, Sorin, Maxi, Sorin, Mascherano, Riquelme, Ayala, Maxi, Mascherano, Maxi, Sorin, Maxi, Cambiasso, Riquelme, Mascherano, Sorin, Saviola, Riquelme, Saviola, Cambiasso, Crespo, Cambiasso with Cambiasso scoring. Maxi was involved 6 times, while Requilme, Masherano and Sorin contributed 4 passes each. Videos of the goal may be found at YouTube, see YouTube (2006a, 2006b).

Figure 1 below provides the estimated locations of where the passes initiated during the play. (How the estimation was carried out will be described below.) The locations are denoted by small circles. Straight lines join them in order of time. The track is meant to represent the path of the ball being played about the field as seen from above. One notes that the ball generally moved towards the S-M goal with passes going off in many directions and some back passes being made. There are no very short passes, the shortest being 5.6m .

There is a growing literature on the statistical modeling of aspects of soccer matches. One highly quoted study was performed by Reep and Benjamin (1968). They investigated data on goal scoring and lengths of passing sequences from 3213 games. They summarize the counts of successful passes by the negative binomial distribution and, for example, conclude that "it takes 10 shots to score 1 goal." Hughes, M. and Franks, I. (2005) describe the Reep and Benjamin paper as a "landmark", but in a discussion of the 'long-ball game' versus 'direct play' complain about various of Reep and Benjamin's conclusions and their impact.

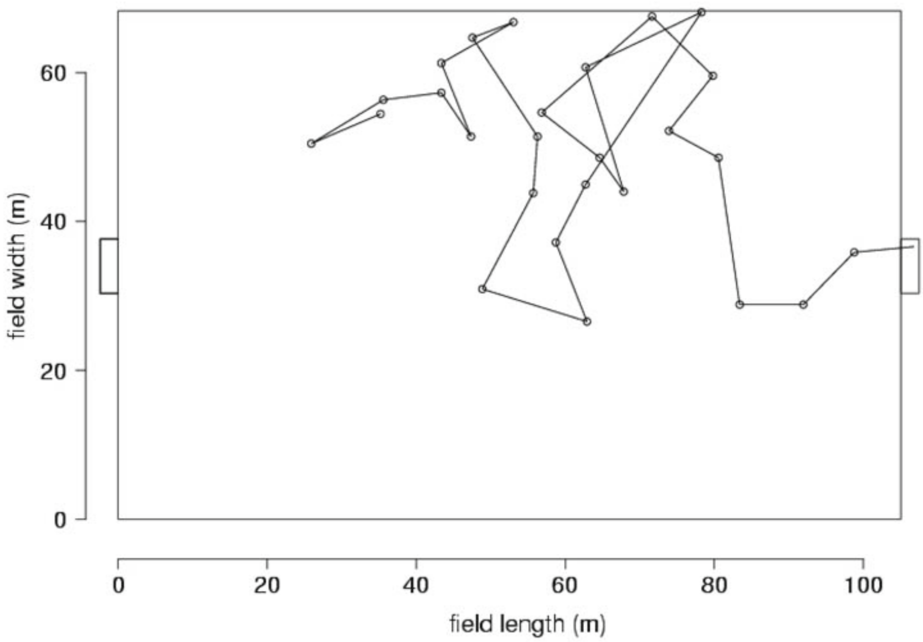


Figure 1. The trajectory of the play. The ball ends in the S-M goal represented by a box on the right hand side of the figure. The circles represent the positions of players initiating passes.

Lee (1997) fits Poisson models for the number of goals scored in games for the 1995-96 season to assess the strengths of various teams. Karlis and Ntzoufras (2003) fit bivariate Poissons to the pairs, *(X,Y)*, of goals scored where *X* is the number of goals scored by one team and *Y* by the other. Hirotsu and Wright have a succession of papers.

One paper, Hirotsu and Wright (2002), modeled the progress of play in a game as a continuous time Markov process with four states. The four states were: each team is in possession of the ball, and when each team scores a goal. Another paper, Hirotsu and Wright (2006) applies game theory to develop effective strategies.

Brillinger (2006a) viewed the results of games as ordinal (win, tie, loss) and fit a model for such data to the Norwegian League results for the 2003 season. A setup of quite a different type altogether is that of Kozlov et al. (1993). They consider an abstract version of a soccer match. The field is infinite with the usual two goals. The path of the ball is planar Brownian motion. They consider the variance of the number of goals and discuss its dependence on the width of the goal for example.

It appears that most of the existing published papers study whole games, tournaments, seasons, or groups of seasons. The initial purpose of this paper was to study that one play in that one game, but the purpose went on to include using the results to develop a flexible model including changes of possession, variable play lengths

There is concern about focusing on a highly unusual play, on an outlier.. It was unusual, it was highly exciting, it contained an unusual number of passes, and it lead to an important goal. Now in statistical data analyses an outlier is to be noted and studied. The analysis should split with a part dropping or weighting down the outlier, and a part looking into it. In DeVeaux et al (2006) one can read, page 534,

"An analysis of the nonoutlying points, along with a separate discussion of the outliers, is often more informative, and can reveal important aspects of the data."

Briefly, there are things to be learned by analyzing outliers.

The path in Figure 1 may be viewed as a realization of a stochastic process described by the time t_i at which the i-th pass was initiated and $(x(t_i),y(t_i))$ the location where the pass was started on the field for $i = 1,...I$. A statistical question is how to describe such a trajectory, that is one involving points connected by straight lines.

The approach employed here involves potential functions motivated by classical mechanics and advanced calculus. It lets one describe instantaneous velocity at an arbitrat place and time. Where will the particle head next and at what speed? This method has proven helpful in describing the motion of a broad variety of objects. Books discussing potential functions include Taylor (2005) and Stewart (1995) while Brillinger et al. (2001) the potential function approach was proposed to describe the motion of elk in a large reserve. Brillinger et al. (2006b) fits a potential function, like the one of this paper, to the motion of a Hawaiian Monk seal.

The potential plus statistical model approach allows simulation of future paths.

Take the fitted potential of the play, symmetrize about the middle, use for each team (different) ends of the field. Show, with additional data, can simulate the flow in a game with the ball changing sides.

After this Introduction the sections of the paper are: The Data, Some Formal Background, Results, Further Developments, Limitations of the Study, Discussion and Summary.

2. The Data

Consideration turns to how the data were obtained and to providing some elementary descriptive statistics. The Argentina Serbia-Montenegro game was played at Gelsenkirchen, Germany. Their field is 105 by 68 m. The data of the play will be denoted by $(t_i, x(t_i), y(t_i))$ where $(x(t),y(t))$ denotes the position of the ball on the field at time t and the t_i are taken to be the times at which the passes were initiated, $i = 1,...,25$. There is also t_{26}, the time that Cambiasso shot. Figure 1 shows the discrete location points where the passes were initiated joined by straight lines to approximate the movement of the ball.

Estimation of the $(t_i, x(t_i), y(t_i))$ was done in two parts. Both used the computer program World 3D Cup 2006, Ascensio System Limited (2006). First, to obtain the $(x(t_i),y(t_i))$, screen dumps were made at the moments of pass initiation. The desired coordinates were then read off using the Windows program Paint. Next the times, t_i, of initiation of the passes were estimated by running the program again, and again, stopping it at the times of pass initiation. The video moved at 25 frames per second so the times could be estimated to .04 of a second.

Figures 2, 3 and 4 present some elementary descriptive statistics concerning the play. The cumulative count panel of Figure 2 may be used to assess whether the rate of passing is changing as a function of time passed. The dashed line joining the first and last points is useful in doing this. It suggests that the rate of passing was increasing towards the final moves of the play. One also sees a long time gap in the middle. This gap corresponds to the longest pass of the play, one from Maxi to Sorin.

The right hand panel provides the stem-and-leaf of the lengths of the various passes. One sees there is an apparent minimum length and a long tail to the distribution. That there is a minimum length to the passes is also apparent in Figure 1. The shortest pass, estimated from the data, was 5.6 m. The distribution also has a long tail with the longest pass 27.9 m . Figure 3 is a stem-and-leaf display of the times between successive paths. It will prove useful in the simulations later. One sees an outlier of 8.2 seconds and that most of the times are less than 3 seconds.

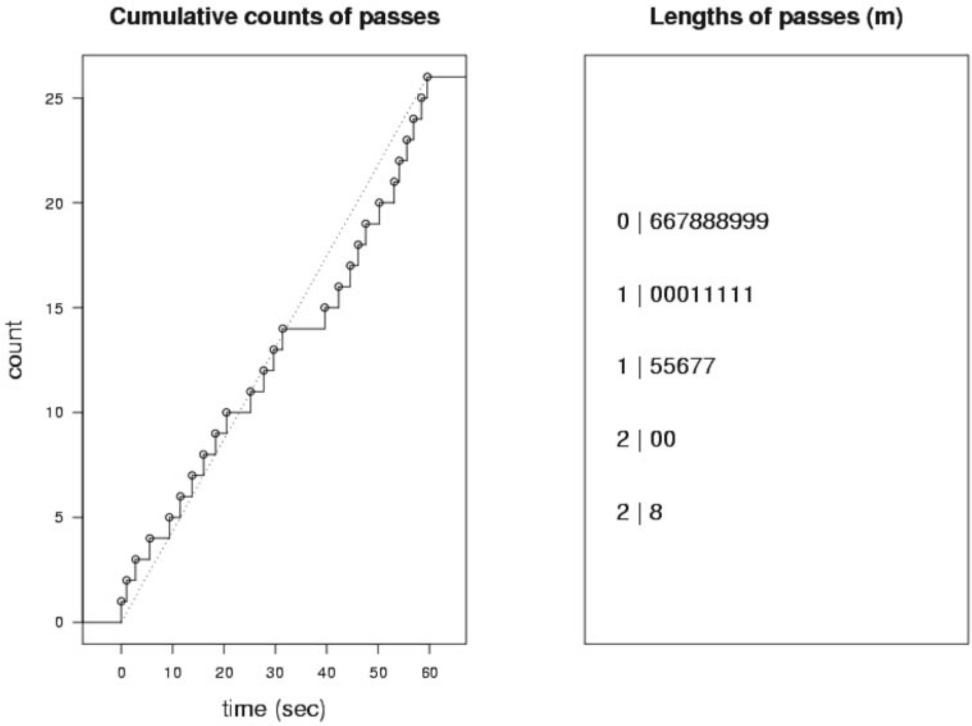


Figure 2. Cumulative counts of passes as a function of time and a stem-and-leaf of the lengths of the passes of the play in meters.

Consideration next turns to the speed of motion of the ball. Figure 4 shows the estimated speed between the time points t_{i+1} and t_i , computed as

$$\sqrt{\{(x(t_{i+1})-x(t_i))^2+(y(t_{i+1})-y(t_i))^2\}}/(t_{i+1}- t_i)$$

In the left panel, speed is graphed as a function of time from start of the play and in the right as a function of distance from the S-M goal. One sees an apparent slowing after the start of the play followed by a a general speeding up of the flow of the ball as time progresses. There is a related speeding up as the ball gets closer to the S-M goal. Lowess lines, Chambers et al. (1983), have been added to the two plots. The speed ranges from 1.87 m/sec to 15.56 m/sec. The outlier, 15.56 m/sec, at the top of both panels, is the long pass from Saviola to Cambiosso near the end of the play.

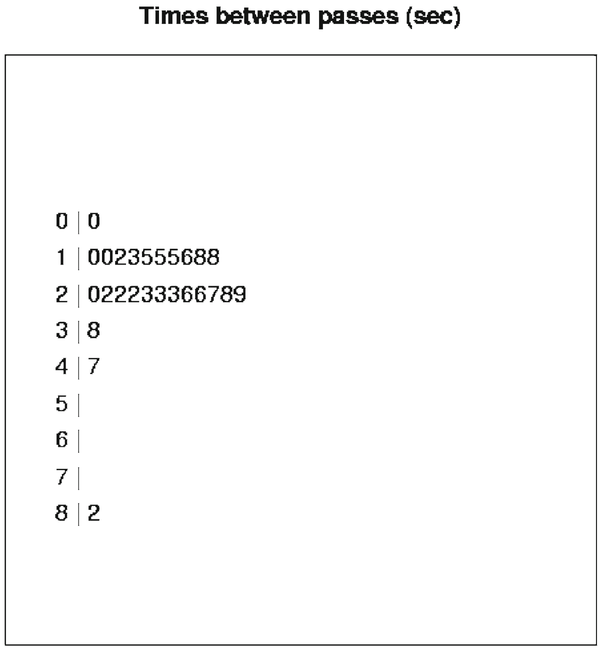


Figure 3. Times between successive passes in seconds.

3. Some Formal Background

Potential functions have long been employed in physics and engineering to describe the motion of particles. These are scalar-valued functions whose dips correspond to points of attraction and peaks correspond to points of repulsion. In the papers, Brillinger et al. (2001), Brillinger (2006b) it was been shown that setting down a potential function allows a consequent setting down of a regression model. The approach there was motivated by consideration of stochastic differential equations.

The classical potential function of Newtonian gravity is given by $-1/|\boldsymbol{r}|$ with $\boldsymbol{r} = (x,y)$, $\boldsymbol{r}$ denoting a 2D row vector and $|\boldsymbol{r}| = \sqrt{(x^2+y^2)}$. This particular potential goes to negative infinity as $|\boldsymbol{r}|$ goes to 0. A particle moving in its field will be attracted to the origin, (0,0). A class of potential functions, of which the Newtonian is a member is provided by $\pm|\boldsymbol{r}|^{\theta}$.

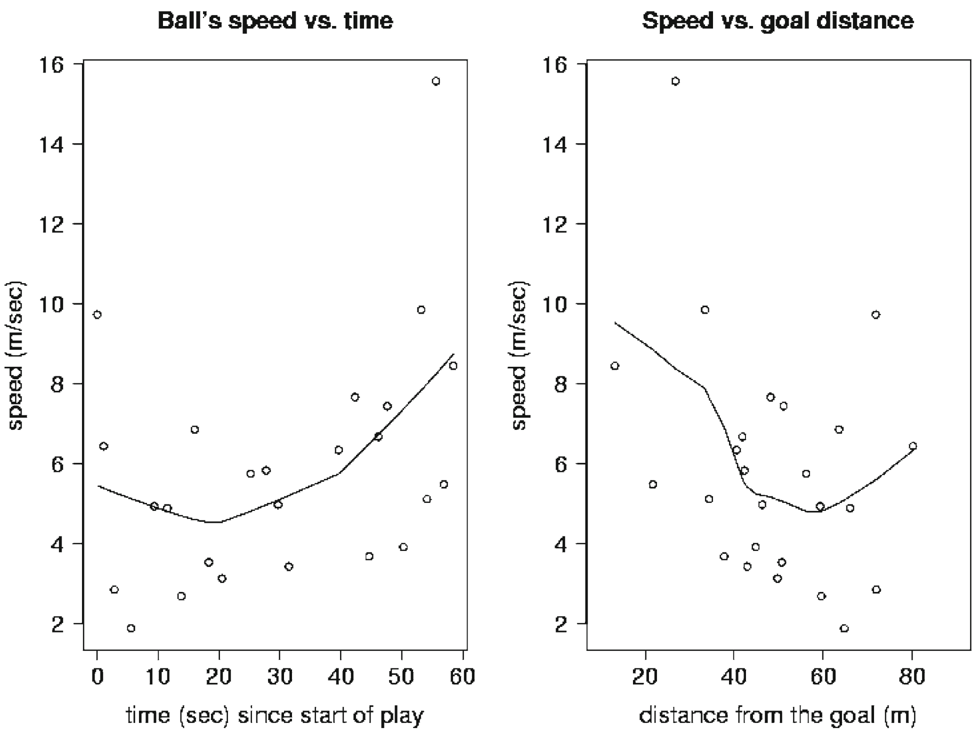


Figure 4. Speed (m/sec) vs. time of play from the start time (left panel) and versus distance from the nearest point of S-M's goalmouth (right panel). Lowess lines have been added in each case.

However the potential function that will be employed in this work includes the term

$$\alpha \, log \, |r| + \beta |r| \qquad (1)$$

for r real-valued. With α positive it will have a point of attraction at 0. It is motivated by formulas on bird navigation in D. G. Kendall (1974). The function is graphed in Figure 5 for the particular values $\alpha = 176.21$ and $\beta = -11.13$. These values were estimated from the data at hand. In the figure one sees the function falling off slowly at first, and then rapidly as one moves from left to right. The "o" on the right is the point of attraction at 105m.

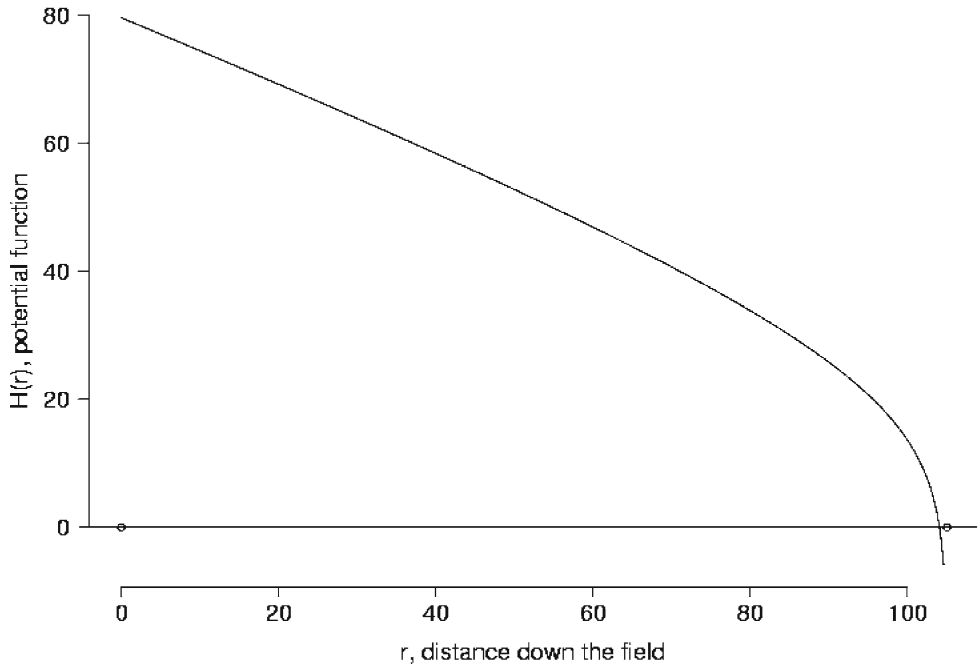


Figure 5. A plot of the estimated Kendall potential function, (1), along the length of the soccer field. (Field ends are indicated by the o's.)

The form (1) may by made more flexible by adding a general quadratic term, specifically by writing

$$H(\boldsymbol{r}) = \alpha \log |\boldsymbol{r}| + \beta |\boldsymbol{r}| + \gamma_1 x + \gamma_2 y + \gamma_{11} x^2 + \gamma_{12} xy + \gamma_{22} y^2 \qquad (2)$$

with $\boldsymbol{r} = (x,y)$.This potential is linear in the parameters considerably simplifying its fitting. In the modeling of the play the S-M goalmouth, as seen in Figure 1, is taken as the attractor.

The negative gradient of a potential function H, namely $\mu = -\nabla H = -(\partial H/\partial x, \partial H/\partial x)$, at location $\boldsymbol{r}$ gives the velocity of the object at $\boldsymbol{r}$. Specifically it gives information on how speed and direction vary with location. It is useful to show velocity on a grid as arrows, the so-called vector field. This is another way to show the flow of motion. Vector fields are discussed in Stewart (1995) for example.

Consideration now turns to setting up a specific model for this type of data. Let $\boldsymbol{r}(t)$ denote the location $(x(t),y(t))$ and consider the model

$$\boldsymbol{r}(t_{i+1})-\boldsymbol{r}(t_i) = \mu(\boldsymbol{r}(t_i))\ (t_{i+1} - t_i) + \eta_i$$

for the changes of position with μ the gradient of a specified parametric potential function, η_i a stochastic noise and the t_i the times of pass initiation.

In the fitting and conceptualization it proves more convenient to write

$$(\boldsymbol{r}(t_{i+1}) - \boldsymbol{r}(t_i))/(t_{i+1} - t_i) = \mu(\boldsymbol{r}(t_i)) + \varepsilon_i \qquad (3)$$

This expression makes it clear that μ has the interpretation as average velocity at $\boldsymbol{r}$. Regarding these noise values it will be assumed, for the moment, that the x- and y-components are independent normals with mean 0 and variance σ^2. Residual plots will be employed to assess whether the variance does not appear constant. Later, expression (3) will be used for simulation purposes.

As expression (2) is linear in the parameters, so too will be its gradient with the implication that simple least squares may be used to estimate the parameters. The variance may be estimated by the standardized sum of the squared residuals of the fit of (3).

4. Results

Figure 6 below provides the estimated potential function, $\hat{H}$, employing expressions (2) and (3). The redder/darker the colors the lower the value of $\hat{H}$. The S-M goalmouth is a line of attraction. The distance $|\boldsymbol{r}|$ of (1) and (2) is taken to be shortest distance to the goal mouth from the location (x,y).

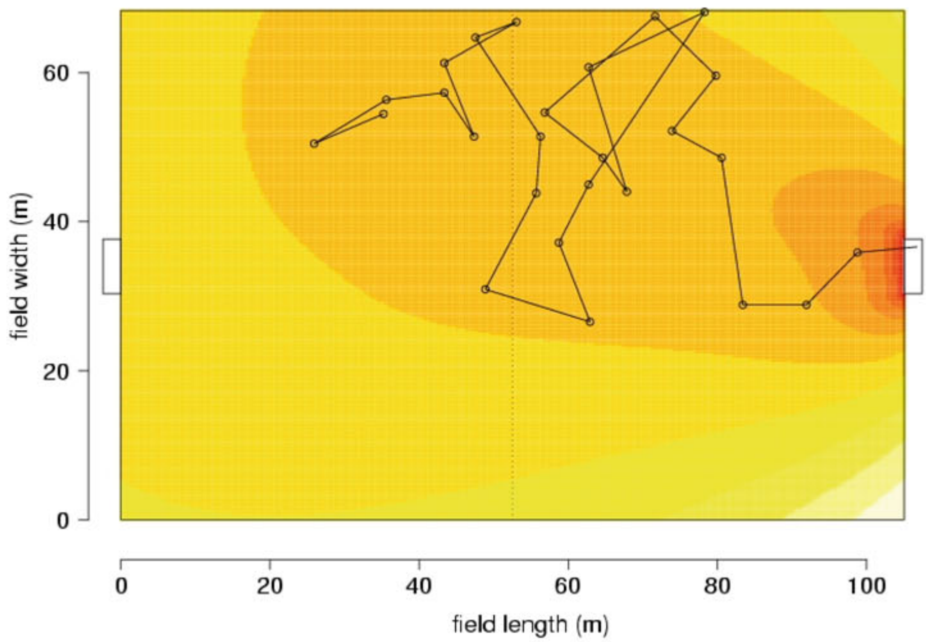


Figure 6. The redder/darker the color is, the smaller the value of the potential estimate $\hat{H}$. The ball's track has been superposed. The dashed vertical line is midfield.

The potential function (2) was fit by linear regression with the total number of observations 50 and 7 linear parameters. The resulting estimate is shown in Figure 6. One sees a funneling/descending valley leading downwards toward the goalmouth. The actual path of the play is also shown. The potential estimate appears quite consistent with the data. The estimate of σ for the model (3) is $\hat{\sigma}$ = 4.27 m/s.

Besides comparing the estimated function with the data, as in Figure 6, various residual plots were prepared, e.g. residual versus time difference and versus its square root. There was no evidence that weights were needed in the model fitting. Stem and leaf displays of residuals were also examined. The variability of the *x*- and *y*-noise component residuals is comparable supporting the assumption of a common σ; however there was some skewing to the left. This ultimately led to a change of noise distribution in the simulations of Section 5. The independence assumption of the *x*- and *y*- components was examined via estimating the correlation coefficient and the value obtained was -.026 . The method of simulation will be employed in the next section to assess the model fit.

The estimated vector field is shown in Figure 7. It provides information at the selected locations on how the average speed and direction of the ball vary with location in the field. One sees average movement from the left hand side towards the S-M goal with a speeding up as one gets closer. The value being an average, the estimated speed (arrow's length) is small where there is back-and-forth motion in a particular area, for example in the middle of the top half of the field. One does see the to-ing and fro-ing in Figure 1.

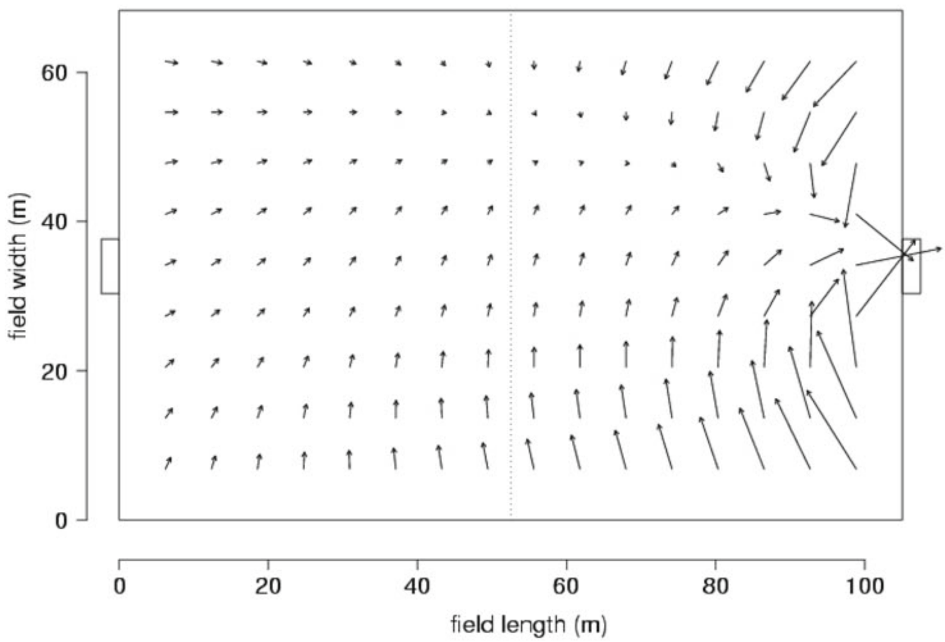


Figure 7. The arrows represent the negative gradient, $\hat{\mu} = -\nabla \hat{H}$. for the locations *(x,y)* the arrows' directions they provide the estimated average direction and their lengths the average speed of the motion.

5. Further Developments

It would be nice to be able to fit potential functions to the data for other segments of some game. Once the data are available a way to do the fitting is clear, but the difficulty is obtaining the data. Still one can use the potential function, estimated above, for other purposes as follows.

Concerning the play and the estimated potential function Michelle Morgan, women's soccer coach at Amherst College, remarked that its development was highly unusual in coming down just one side of the field. This behavior by the players leads to the particular shape of the estimated potential function as shown in Figure 6. To obtain another potential function of interest one can symmetrize that estimate by reflection and addition. The result obtained is Figure 8.

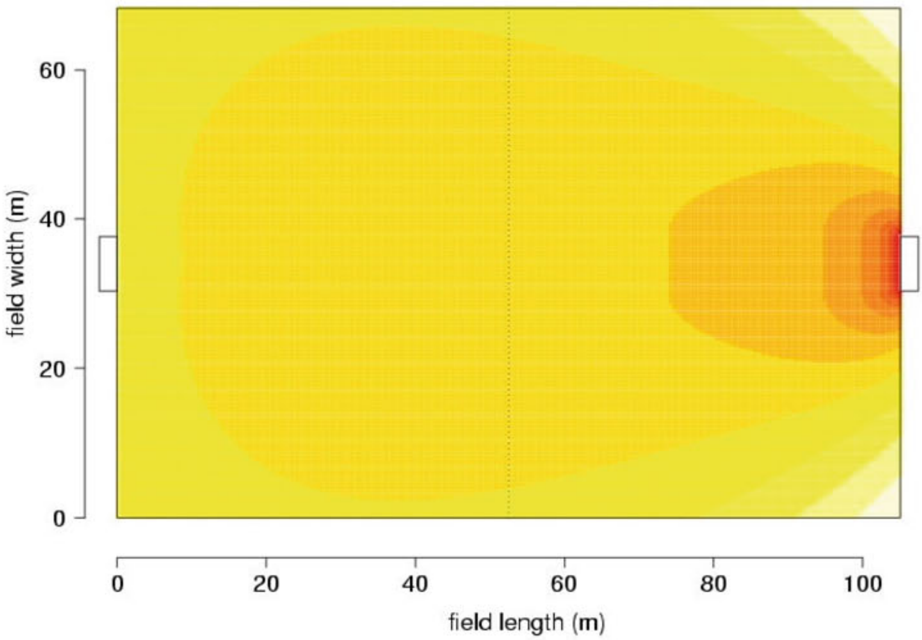


Figure 8. The result of symmetrizing the potential function of Figure 6 by reflecting about the horizontal center line and adding the two.

It would be nice to be able to include the behavior of the other team in the study. The opponent would have a different potential function, one pointing to the Argentinean goal. It would come into use when the ball changed sides. A potential function to consider may be obtained by symmetrizing Figure 8 about midfield line.

To bring in an opposing team one needs information on the relative lengths of the various possessions. Figure 2 in Hughes and Franks (2005) may be used to estimate the probabilities of various lengths. Their data are for the 1990 and 1994

World Cups. Table 1, adapted from their Figure 2, gives estimates of individual probabilities up to 8 passes and the proportion remaining.

Table 1. The top line is the length of possession. The second line provides estimates of the proportion that have the indicated possession length.

0	1	2	3	4	5	6	7	8	>8
.398	.194	.123	.086	.062	.040	.031	.022	.015	.028

The next table, based on Table 1, provides estimates of the corresponding conditional probabilities of the play ending at that possession number as a function of the number of passes already completed.

Table 2. The top line refers to the number of passes so far completed. The second line provides estimates of the probability that the possession ends.

0	1	2	3	4	5	6	7	8
.398	.323	.303	.304	.313	.295	.323	.333	.357

The numbers have the interpretation of being estimates of the probability the player will lose the ball after the indicated number of complete passes. This will be needed in the simulations to be developed. The values of the table are surprisingly constant around .3 .

Also needed are times between successive passes. These were provided in the stem-and-leaf of Figure 2 for the 25 pass play.

6. Simulation

One gains further understanding of a phenomenon when one can produce plausible simulations. A further objective of this work then is to be able to simulate plays similar to the Argentinean one, as well as other types of plays that might occur in the course of a game.

The estimation procedure employed was made specific by an assumption of normal noise in model (3). This meant that all the distances between successive locations were possible. However as the stem-and-leaf in Figure 2 shows, and common knowledge suggests, there typically appears to be some noticeable minimum length pass in plays. Further the definition of a pass in this paper includes the dribbling by the player receiving the ball. Dribbling adds to both the time and the distance. For this reason, in the simulations presented, a minimum pass length was employed. A few initial simulations had shown that if the length was allowed to be very short, synthetic runs did not resemble soccer plays well. A minimum length of 5m is consequently employed in the simulations.

The field's boundaries are also an issue. Plays can end: by the ball going out of bounds, by a goal, by the ball going to an opponent, and by a referee's whistle. There are various formal ways of dealing with boundaries when simulating realizations. The case of continuous time is reviewed in Brillinger (2003). In the present case random numbers leading to passes of length less than 5m will be rejected, as will those of passes longer than 5m that go out of bounds. One effect of this is that the noise distribution now depends on the field position in contrast to that of the model (3) which assumed common variance normal errors.

The specific steps of the simulation procedure employed are:

1.The estimates $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}$ are obtained by running a standard least squares program employing the model (3).

2. The differences between the pass times employed, the $t^*_{i+1} - t^*_i$, are sampled from those of the Argentinean play, i.e. from those of Figure 3.

3. The starting field location of a simulation run is taken to be $\hat{\boldsymbol{r}}(t_1) = \boldsymbol{r}(t_1)$.

4. A tentative value generated is

$$\hat{\boldsymbol{r}}(t^*_{i+1}) = \hat{\boldsymbol{r}}(t^*_i) + \hat{\mu}(t^*_i)\,(t^*_{i+1} - t^*_i) + \hat{\sigma}\,\boldsymbol{Z}_{i+1}(t^*_{i+1} - t^*_i)$$

$i = 1, \ldots$ with the $\boldsymbol{Z}$'s independent bivariate standard normals.

If the pass generated was less than 5m, or "the ball" goes out of bounds, then that iteration is ignored and a new $\boldsymbol{Z}$ generated.

5. A negative binomial variate is generated to determine how many passes are in a sequence. (In the simulations the probability is .2 .)

6. At this point one switches to the mirror image potential function.

7. One continues until the ball changes hands a second time.

7. Further Results

The results for the first six simulation runs are provided in Figure 9. In the panels blue represents Argentina in possession of the ball and red Serbia-Montenegro. A simulation ended with either the ball changing sides a second time or a goal.

As indicated at step 5 above the simulations took the conditional probability of a pass being incomplete as .2, despite .3 or so being suggested by Table 2. This was to give the plays of Figure 9 greater length for illustrative purposes.

Panel 1 shows Argentina losing possession just as they enter the S-M half followed by an S-M shot. Panel 2 shows Argentina losing possession directly, and then an S-M shot. Panel 3 shows Argentina losing possession immediately followed by S-M doing the same. Panel 5 shows a representation like that of the actual goal with Argentina keeping the ball and shooting at the S-M goal..

In the simulations and figure one could have had the simulation run with the ball changing side more than once, but then the figures would have become cluttered. A video is called for when the ball changes hands more than twice.

8. Limitations of the Study

It would be remiss not to mention some of the limitations of this study. To begin, just one play is studied. The reasons for this are twofold: the excitement of the particular play and the unavailability, just now, of other data to study. As more data become available further empirical studies may be carried out directly. In the meantime the results of the 25 play analysis can be employed to generate other potential functions, as in Section 6. Also of soccer know how can be used to set down and study other potential functions. Further it can be noted that the play studied did cover much of the field and thereby contains information on the behavior of an attacking team over a substantial part of the defending teams half.

In drawing conclusions, one needs to remember that one is dealing with shots on goal, not actual goals. Shot on goals include both goals and the ball going over the crossbar. One could include goals in the simulation by having a shot become a goal with some probability, for example the 1/10 of Reep and Benjamin (1968). Also it needs to be remembered that the dribbling after receipt of the ball is included in the definition of a pass.

There is measurement error, due to the discretization of the Ascensio representation. This could be studied in some detail, but it does not seem that the conclusions of the paper, for example Figures 6 and 7, are likely to change a great deal.

Lastly it is to be noted that formal models are but mimics of actuality. Real people are involved. There is diving, professional fouling, anticipation, and delays of the game. In the work of the paper such effects go into the error term of expression (2).

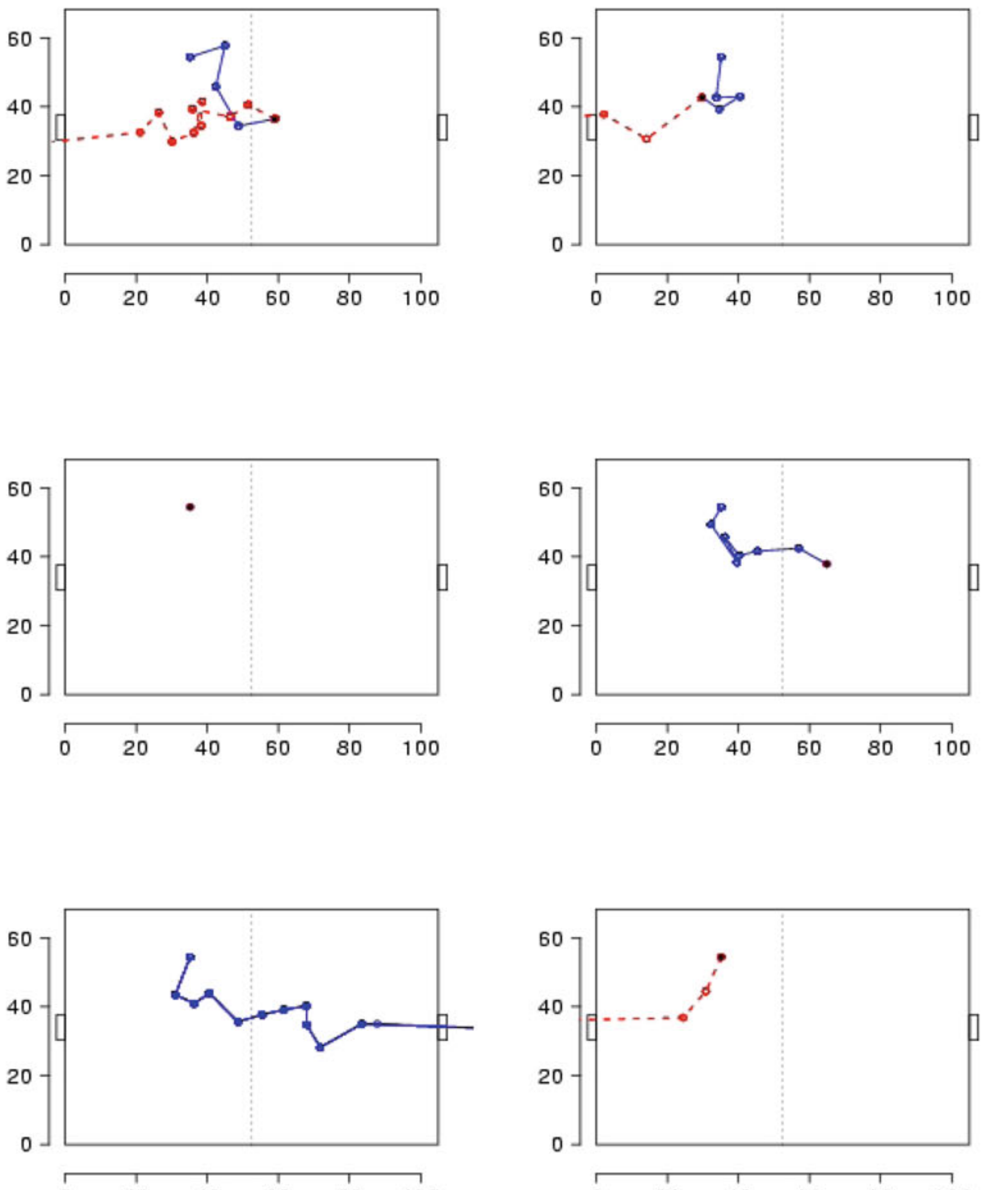


Figure 9. In the panels blue represents Argentina in possession of the ball and red-dashed Serbia-Montenegro. The lines going through the goal are shots on goal. The solid circle corresponds to the ball changing hands.

9. Discussion and Summary

In this paper soccer has been viewed as a dynamical system with its details such as: regions of attraction, boundaries, and repulsion. The work of this paper suggests that there are analytic and conceptual advantages to employing a potential function in the description and simulation of the motion of a soccer ball. The potential is scalar-valued making it simpler to set down an expression for the instantaneous velocity as a function of location (x,y). Substantive knowledge may be employed to set down a form for a potential function and to interpret one that has been estimated. The method provides a flexible approach that is direct to invoke when other data sets come along.

So much is known about the particular sport of soccer that it provides a useful test case for the potential function approach in other sports' contexts. The approach could lead to comparative studies, classifications of plays, and further empirical studies. The approach has further led to a viable method for simulation, e.g. for bootstrapping and model appraisal. Details include: i) as an analytic formula, (2), is available for the potential function plays may be followed for any position (x,y) on the field, and ii) the potential function may change with time as in the simulations of Figure 9.

One would like interpretations of the results. One can ask what might the potential function and vector field be used for and represent? Simulation use has already been emphasized. Perhaps computer-based training might be introduced to teach strategy. Perhaps the function can be used to summarize history, tactics, and even players' intuition.

Following its use in ice hockey, Thomas (2006), possession time might be another important explanatory in broader soccer studies. The possession time of the Argentinean goal was 59.6 sec. This particular goal had substantial elements of both patience and speed.

References

Ascensio System Limited (2006) World 3D Cup 2006 Player. Available at http:www.footballsoftpro.com

David R. Brillinger, Haiganoush K. Preisler, Alan A. Ager, and John Kie (2001) "The Use of Potential Functions in Modelling Animal Movement" *Data Analysis From Statistical Foundations,* 369-386. Available at: http://www.stat.berkeley.edu/~brill/papers.html

David R. Brillinger (2003). "Simulating Constrained Animal Motion Using Stochastic Differential Equations." *Probability, Statistics and their Applications* (Eds. K. Athreya, M. Majumdar, M. Puri and E. Waymire) Institute of Mathematical Statistics, Beachwood. Available at: http://www.stat.berkeley.edu/~brill/papers.html

David R. Brillinger (2006a) "Modelling Some Norwegian Soccer Data" *Doksum Festschrift*, forthcoming. Available at: http://www.stat.berkeley.edu/~brill/papers.html

David R. Brillinger (2006b) "A Meandering *Hylje*", *Festschrift for Tarmo Pukkila on His 60th Birthday* (Eds. E. P. Liski, J. Isotalo, S. Puntanen, and G.P.H. Styan) Dept. of Mathematics, Statistics and Philosophy, Univ. of Tampere, Finland. Available at: http://www.stat.berkeley.edu/~brill/Papers/hylje1.pdf

John M. Chambers, William S. Cleveland, Beat Kleiner and Paul A. Tukey (1983). *Graphical Methods For Data Analysis*. Duxbury, Boston.

Richard D. De Veaux, Paul F. Velleman and David E. Bock (2006). *Intro Stats*. Pearson, Boston.

D. Karlis, and J. Ntzoufras (2003) "Analysis of Sports Data Using Bivariate Poisson Models" *The Statistician* 52, 381-393.

Nobuyoshi Hirotsu and Mike B. Wright (2002) "Using a Markov process model of an Association Football Match to determine the Optimal Timing of Substitution and Tactical Decisions" *Journal of the Operational Research Society*: Vol. 53: No. 1, 88-96.

Nobuyoshi Hirotsu and Mike B. Wright (2006) "Modeling Tactical Changes of Formation in Association Football as a Zero-Sum Game", *Journal of Quantitative Analysis in Sports*: Vol. 2: No. 2, Article 4. Available at: http://www.bepress.com/jqas/vol2/iss2/4

M. Hughes and I. Franks (2005). "Analysis of Passing Sequences, Shots and Goals in Soccer". *Journal of Sports Sciences* 23, 509-514.

David G. Kendall (1974). "Pole-seeking Brownian Motion and Bird Navigation." *Journal of the Royal Statistical Society B*: Vol. 36, 365-417.

S. Kozlov, J. Pitman and M. Yor (1993) "Wiener Football". *Probability Theory and its Applications* 40, 530-533.

Alan J. Lee (1997). "Modelling Scores in the Premier League: Is Manchester United Really the Best?" *Chance*: Vol. 10, 15-19.

C. Reep and B. Benjamin (1968) "Skill and chance in association football" *Journal of the Royal Statistical Society A*, 131 581- 585

James Stewart (1995) *Calculus: Early Transcendentals*, Third Edition. Brooks Cole, Pacific Grove.

John R. Taylor (2005) *Classical Mechanics* University Science, Sausalito.

Andrew C. Thomas (2006) "The Impact of Puck Possession and Location on Ice Hockey Strategy," *Journal of Quantitative Analysis in Sports*: Vol. 2: No.1 Article 6. Available at: http://www.bepress.com/jqas/vol2/iss1/6

YouTube (2006a). Video of the play. Available at http://youtube.com/watch?v=2uC5mUFye50&search=world%20cup%20006

YouTube (2006b). Video of the play. Available at: http://www.youtube.com/watch?v=EoL8Lr9LVM4

THE USE OF POTENTIAL FUNCTIONS IN MODELLING ANIMAL MOVEMENT

D. R. BRILLINGER, H. K. PREISLER, A. A. AGER, J. G. KIE

Department of Statistics
University of California
Berkeley, CA, 94720-3860

Pacific Southwest Research Station
USDA Forest Service
Albany, CA 94710

Pacific Northwest Research Station
USDA Forest Service
La Grande, OR 97850

Pacific Northwest Research Station
USDA Forest Service
La Grande, OR 97850

SUMMARY

Potential functions are a physical science concept often used in modelling the motion of particles and planets. In the work of this paper potential function based

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_22, © Springer Science+Business Media, LLC 2012

models are considered for the movement of free-ranging elk in a large, fenced experimental forest. Equations of motion are set down and the parameters involved are estimated nonparametrically. The question of whether a potential function is plausible for describing the elk motion is considered. The conclusion is that it is not possible to reject this hypothesis for the data set and estimates considered.

Key Words and Phrases: animal movement, diffusion models, elk, force field, nonparametric regression, potential functions, stochastic differential equations, telemetry data.

1. INTRODUCTION

The problem of interest is the description of the movement of elk, *Cervus elaphus*, in a large free-ranging environment. Models of animal movement are becoming important tools in the study of a variety of ecological problems, especially habitat selection, animal migration and dispersal in heterogeneous landscapes. Specific questions that wildlife biologists have include: How to allocate forage amongst competing species? What is the effect of vehicular traffic? Is change taking place? What is the sequence of habitat use? The physical and biological mechanisms that regulate such movements are clearly complex.

The data available are the locations of M elk, labelled by $m = 1, ..., M$, recorded at times, $t_{mk}, k = 1, ..., K_m$. More specifically the data consist of the locations $\mathbf{r}_{mk} = (X(t_{mk}), Y(t_{mk}))$, corresponding to the UTM (Universal Transverse Mercator) coordinates of the k-th time measurement of the m-th elk. Explanatory variables describing vegetation, topography, and other habitat features (e.g., distance to road, distance to water) known to influence elk movement, are also available.

The approach developed in this work is to assume that the animals are moving in a potential field, $H(\mathbf{r}, t)$, that controls their direction and speed of motion. The potential field may have points, lines or regions of attraction or repulsion and may include barriers. The barriers may represent actual physical constructions (e.g., fences or be natural). Stochastic differential equations (SDEs) are used to include variability in the model such as attractors and repellors not in the potential H. The estimated SDEs may be used to produce estimates of other parameters, eg. speed, to predict spatial and temporal patterns of animal distribution and habitat preferences, to simulate trajectories and to study the directionality of the movement, amongst other possibilities. Later in the paper simulations of the trajectories will be used to estimate the potential function.

The paper begins with a description of both deterministic and stochastic methods for describing the paths followed by particles under the influence of a potential field. Next the experiment in which the elk data were collected is described. Section 4 provides details of the statistical methods employed in the problem. Section 5 presents the results obtained. A key examination of the assumption that a potential function exists is a comparison of second-order partial derivatives taken in the two possible orders, separately for daytime and nighttime data. The final section reviews some of the merits and limitations of employing the potential function to model animal movement.

References describing models for animal movement include: [6, 9, 10, 18, 27]. Reference [18] sets down deterministic differential equations (DDEs) for density functions describing the expected pattern of space use by coyotes being influenced by the accumulation and decay of scent marks, also described by DDEs. This is to be contrasted with the approach in [4, 20] where stochastic equations were set down describing the individual realizations or trajectories, for elephant seals migrating and female bark beetles responding to male pheromones emitted from a point source, respectively.

2. SOME MATHEMATICS OF MOVING PARTICLES

Both deterministic and stochastic approaches are available for describing the trajectories of moving particles.

2.1 *Deterministic case.*

Motion in Newtonian dynamics has often been described by a potential function, $H(\mathbf{r}, t)$, see [19]. Here $\mathbf{r} = (x, y)$ is location and t is time. The equation of motion takes the form

$$d\mathbf{r}(t) = \mathbf{v}(t)dt$$

$$d\mathbf{v}(t) = -\beta\mathbf{v}(t)dt - \beta\nabla H(\mathbf{r}(t), t)dt$$

with $\mathbf{r}(t)$ the particle's location at time t, $\mathbf{v}(t)$ the particle's velocity and $-\beta\nabla H$ the external force field acting on the particle, β being the coefficient of friction, [19]. Here $\nabla = (\partial/\partial x, \ \partial/\partial y)$ is the gradient operator. The function H is seen to control the particle's direction and velocity. For example $H(\mathbf{r}) = |\mathbf{r} - \mathbf{a}|^2$ corresponds is a point of attraction at $\mathbf{a}$ and $H(\mathbf{r}) = 1/|\mathbf{r} - \mathbf{a}|^2$ is a potential function with a point of repulsion at $\mathbf{a}$.

In the case that β is large, the equations are approximately

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r}(t), t)dt \tag{2.1}$$

and only the location, $\mathbf{r}(t)$, at time t is involved.

There exists considerable mathematical development in the time stationary case. A force field, $\mathbf{F}$, may be given and the question arises whether there exists a real-valued function H, such that $\mathbf{F} = \nabla H$. When it does the field is called *conservative*. Such a field then has the property that line integrals

$$\int_C \mathbf{F} \cdot d\mathbf{r}$$

depend only on the initial and terminal points of the curve C, see [26], and $\cdot$ refers to the fact that a line integral is involved.

In this case the function H may be obtained from its partial derivatives, $\mathbf{F} = (H_x, H_y)$, [25, 26]. Specifically for motion in an open connected region the potential function may be obtained, up to an additive constant, as

$$H(x, y) = \int_{(a,b)}^{(x,y)} \mathbf{F} \cdot d\mathbf{r} \qquad (2.2)$$

where (a, b) is a point in the region. When a potential function exists, the path of the line integral taken from the starting point (a, b) to the terminal point (x, y) will not affect the final result. The function H may also be estimated, given H_x, H_y via simulation experiments as described below.

If $\mathbf{F}$ has components H_x, H_y, then a necessary condition for the existence of a corresponding potential function is that

$$\frac{\partial}{\partial y} H_x = \frac{\partial}{\partial x} H_y \qquad (2.3)$$

[25, 26]. In the case that the region is simply connected, this condition is also sufficient.

2.2 *Stochastic case.*

A pertinent probabilistic concept for dynamic situations is a stochastic differential equation (SDE), see [3, 16]. Such equations lead to Markov processes and take the form

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t), t)dt + \boldsymbol{\Sigma}(\mathbf{r}, t)d\mathbf{B}(t) \qquad (2.4)$$

with $\boldsymbol{\mu}$ the drift parameter, $\boldsymbol{\Sigma}$ the variance or diffusion parameter and $\mathbf{B}$ bivariate Brownian motion. Here $\mathbf{r}$, $\boldsymbol{\mu}$, $\mathbf{B}$ are vectors while $\boldsymbol{\Sigma}$ is a matrix.

The parameters have the interpretations

$$E\{d\mathbf{r}(t)|\mathcal{H}_{t}\} = \boldsymbol{\mu}(\mathbf{r}(t),t)\,dt$$

$$var\{d\mathbf{r}(t)|\mathcal{H}_{t}\} = \boldsymbol{\Sigma}(\mathbf{r}(t),t)\,dt$$

with $\mathcal{H}_t$ representing the time history of the process. Since the process is Markov, these conditional parameters depend only on the previous position, as indicated.

Many properties are known concerning solutions of SDEs, for example in the present context when H does not depend on t and $\boldsymbol{\Sigma} = \sigma_0^2\mathbf{I}$, there may be an invariant density

$$\pi(\mathbf{r}) = c\ exp\{-2H(\mathbf{r})/\sigma_0^2\} \tag{2.5}$$

representing the longrun density of locations the particle visits, [3]. Thus, by modelling movements, population distributions may be estimated. At the same time given $\boldsymbol{\mu} = (-H_{\boldsymbol{x}}, -H_{\boldsymbol{y}})$ and a σ_0, realizations of the process (2.4) may be generated, from which the density $\pi(\mathbf{r})$ may be estimated from the realizations and then (2.5) inverted to obtain an estimate of H.

There may be barriers restraining the motion. Also the stimulus, here represented by $\boldsymbol{\Sigma}(\mathbf{r},t)\,d\mathbf{B}(t)$, may have periodic properties in t.

A particular case of an SDE is provided by the mean-reverting *Ornstein – Uhlenbeck* (O-U) process where

$$\boldsymbol{\mu}(\mathbf{r},t) = \mathbf{A}(\mathbf{a} - \mathbf{r}(t))$$

$$\boldsymbol{\Sigma}(\mathbf{r},t) = \boldsymbol{\Sigma}$$

and the mean is $\mathbf{a}$. The papers [9, 10] propose the O-U process as a model for animal motion and develop maximum likelihod estimates of the parameters. The O-U process becomes the *random walk* when $\mathbf{A} = \mathbf{0}$, i.e., when the drift term, $\boldsymbol{\mu}(\mathbf{r}, t)$, is $\mathbf{0}$.

If $\mathbf{A}$ is symmetric, the potential function corresponding to an O-U process is

$$H(\mathbf{r}, t) = (\mathbf{a} - \mathbf{r})^{\tau} \mathbf{A}(\mathbf{a} - \mathbf{r})/2$$

Its invariant distribution is multivariate normal, $N(\mathbf{a}, \mathbf{\Psi})$, where

$$\mathbf{\Psi} = \int_0^{\infty} e^{-\mathbf{A}u} \mathbf{\Sigma}\mathbf{\Sigma}^{\tau} e^{-\mathbf{A}u} du$$

see [3], p. 597. If $\mathbf{\Sigma} = \sigma_0^2 \mathbf{I}$, then $\mathbf{\Psi} = \sigma_0^2 \mathbf{A}^{-1}/2$.

The situation of a particle being affected by the force field of a potential function is conveniently visualized by picturing a ball rolling around in the interior of a perspective plot of the potential function. Some simulations are provided below.

To derive simulated paths on can proceed as follows. Consider a one-dimensional process $dx(t) = \mu(x, t)dt + \sigma(x, t)dB(t)$. Suppose that at time t the particle is at location $x(t) = x$. Now for the location at time $t + dt$ take

$$x(t + dt) = x \pm \sigma(x, t)\sqrt{dt} \quad with\ prob \quad \frac{1}{2} \pm \frac{\mu(x, t)}{2\sigma(x, t)}\sqrt{dt}$$

See [17, 22]. In the bivariate case one generates x and y processes.

Figure 1 presents examples of such simulations in the case of the process

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r})dt + d\mathbf{B}(t)$$

and two particular potential functions. In the first example $H(\mathbf{r}, t) = \mathbf{r}^{\tau}\mathbf{r}$, i.e. the process is Ornstein-Uhlenbeck reverting to the origin. The trajectory is seen to meander around the origin and one can imagine a ball rolling around in the interior of the parabaloid in the left column of the Figure 1.

In the second example a mound has been added at the origin. Now the trajectory is seen to circle around the mound staying in the groove of the bottom of H.

Keeping in mind these examples one can visualize the motion of particles given particular potential functions.

2.3 *Random potential/environment.*

The discussion above provides a means of interpreting the drift term of a bivariate SDE. It is also important to have an understanding of what phenomena can lead to the variance/diffusion term.

Suppose that at time t there are other particles and that they are at random positions $\mathbf{r}_j(t)$. These particles might be attracted towards each other following the existence of a potential function

$$H(\mathbf{r},t) \;=\; \alpha(\mathbf{r})\sum_{j=1}^{J}|\mathbf{r} \;-\; \mathbf{r}_j(t)|^2$$

for some pertinent function $\alpha(.)$. Following equation (2.1)

$$d\mathbf{r}(t) \;=\; -\nabla H(\mathbf{r},t)dt$$

with $\nabla H(\mathbf{r},t)$ approximately normal for large J via some Central Limit Theorem. One has, approximately, an SDE such as (2.4) with no drift term.

The concept of other particles in the field might be used to portray the attraction among elk traveling together in a herd for example. Conversely, it could be used to portray repulsion between two different species of animals where, because of social interactions, individuals of one species are avoiding individuals of the other species.

3. THE EXPERIMENT

The main study area at Starkey Experimental Forest and Range consists of 7,762 ha in the Blue Mountains of northeastern Oregon [23]. It was enclosed with a gameproof fence in 1988 and radio-telemetry studies were initiated. Each spring a sample of the resident population of elk and mule deer (*Odocoileus hemionus*) are fitted with collars containing Loran-C receivers. A sample of the domestic cattle herd brought to Starkey Forest each summer is also fitted with collars. The collars are instructed at regular intervals to intercept Loran-C broadcsts and relay these signals to a central receiver. Locations are then computed from the Loran-C time delay. They have a mean error about 50 m [12]. The telemetry system attempts to locate some animal every 20 seconds, and thus cycles through approximately 190 collared elk, deer, and cattle in about 60-65 minutes. The study area is also managed for a variety of public uses such as recreation, hunting, forest management, cattle grazing, and other activities. An extensive database was built describing vegetation, topography, and location of roads, streams and other features relevant to the study of elk [23]. The data used in the work of this paper were collected from the analyses in 1994 and involve 53 female elk. Observations were omitted from the analysis for 30 days when hunting of elk by rifle occurred in the forest, and also when time intervals between successive locations were greater than 1.5 hours. This was done in an attempt to make the situation more uniform and reduce the difficulties of interpreting widely spaced observations. Figure 2 illustrates the successive movements for two typical elk during 1994. Two small game-proof exclosures within the study area are shown in white. Elk 43 is seen to spend much of its time below the larger fenced off area on the right. The trajectory plotted is a sequence of straight line segments and jagged. This discreteness results from the fact that location estimates are available but every 1-4 hours.

Turning to Elk 42, it is seen to spend most of the time in the northern part of the forest. The implications of the time sampling are particularly apparent in this case in the upper right corner. It is not that the elk is jumping the fence, rather the locations are at time points an hour or so apart. Elk 42 does stay within the Starkey Forest (at least as far as is known).

Figure 3 shows separately the daytime and nighttime locations visited by all 53 elk, but restricting the points plotted to those less than 1.5 hours apart and excluding the days with hunting.

The points plotted have been jittered to make their apparent density clearer. A variety of heavily used and and also sparsely used regions may be seen. When a detailed map is consulted it can be seen that some of these regions relate to the locations of roads and other habitat features. This circumstance will be addressed in later research. There is also an apparent difference between day and night distributions, which is no surprise because the animals forage at dawn and dusk and rest in the daytime.

4. THE STATISTICAL METHODS USED

Kernel methods, [14], may be employed to form an estimate of the longrun density of elk locations. Estimates take the form

$$\hat{\pi}(\mathbf{r}) \;=\; \sum_{m,k} K(\mathbf{r} \;-\; \mathbf{r}(t_{mk})) \;/\; \sum_{m,k} 1 \tag{4.1}$$

for some kernel function $K(.)$. Such an estimate will be employed later in the paper, together with the realtion (2.5), to obtain an estimated of the (assumed to exist) potential function.

Turning to the SDE (2.4) its solution may be approximated by

$$(\mathbf{r}(t_{l+1}) - \mathbf{r}(t_l))/(t_{l+1} - t_l) \;\approx\; \boldsymbol{\mu}(\mathbf{r}(t_l), t_l) \;+\; \boldsymbol{\Sigma}(\mathbf{r}(t_l), t_l)\mathbf{Z}/\sqrt{t_{l+1} - t_l}$$

$l = 1, 2, \ldots$ with $t_1 < t_2 < t_3 < \ldots$ sampling times and with $\mathbf{Z}$ a bivariate standard normal. In terms of the individual components of $\mathbf{r}$ one can write

$$\frac{\Delta X(t)}{\Delta t} \;=\; \mu_1(X,Y) \;+\; noise$$

$$\frac{\Delta Y(t)}{\Delta t} \;=\; \mu_2(X,Y) \;+\; noise$$

further assuming time invariance. If the drift functions, $\mu_1,\ \mu_2$, are smooth, one has a nonparametric regression problem. The functions μ_1, μ_2 may be estimated via loess(.), [7], or by a kernel method, [14].

Acting as if H exists, from estimates of μ_1, μ_2 one has an estimate of H's gradient $(\hat{H}_x, \hat{H}_y) \;=\; -\,(\hat{\mu}_1, \hat{\mu}_2)$. The function H itself may then be estimated following (2.2), specifically one could employ

$$\sum_i \hat{H}_x(x_i, y_i)\triangle x_i \;+\; \sum_i \hat{H}_y(x_i, y_i)\triangle y_i$$

for some path of points $(x_i, y_i),\ i = 0, 1, 2, \ldots$ from (a, b) to (x, y) staying within the region having taken some starting point (a, b) in the region, i.e. standardized the estimate by $\hat{H}(a, b) \;=\; 0$. Depending on the character of the region complex paths may be needed. This is the case for the region of this paper.

References to inferential methods for diffusion processes include: [1, 2, 5, 8, 13, 15, 24].

5. RESULTS

The results of the model fitting and assessment are provided in Figures 4-7.

Following expression (2.5), and under assumptions leading to its existence, the potential function may be estimated up to an additive constant by

$$-log\ \hat{\pi}(\mathbf{r}) \tag{5.1}$$

with $\hat{\pi}(.)$ the density estimate, usin gthe kernel estimate. The results are given in Figure 5 separately for days and nights. The hotspots of Figures 3 go over into the depressions, i.e. coldspots, of Figure 4.

Figure 5 provides $\hat{H}$ as estimated by simulation following the method described in Section 2.2 having used loess(.) to estimate the gradient of H. In the estimate provided the starting point was taken to be the center point of the region. The value of σ_0 was 20 pixel units, to be sure the trajectory roamed around the region widely. The points of the trajectory were picked to remain within the outer boundary of Starkey by resampling an increment if it led to a point outside of the region.

In both Figures 4 and 5 the hotspots (lighter areas) are in the north of Starkey in daytime and in the south in nightime. One sees the main attractors. The extent of agreement relates in part to the question of whether a potential function actually exists, for this assumption underlies the computations leading to Figure 5. If a potential function does not exist then one needs a different method of estimating $\pi(\mathbf{r})$ because one cannot simply integrate up. The question of statistical uncertainty will be addressed below.

Expression (2.3) suggests one way to address the question of the existence of a potential function. Figure 6 takes $\hat{H}_x$, $\hat{H}_y$ and further computes $\Delta_y \hat{H}_x$ and $\Delta_x \hat{H}_y$. (Here Δ_x, Δ_y are the x and y difference operators.) There is some agreement. The daytime plots are on the same grey scale, as are the nighttime plots.

It is clear that some discussion of sampling uncertainty is needed in order to make plausible inferences. In the work the jackknife, [11], was employed to examine the hypothesis of the existence of a potential function. In its implimentation 50 of the elk tracks were used, 5 tracks were dropped each time in the evaluations of the 10 pseudo estimates.

Given estimates of the variances of the quantities graphed in Figure 6, they may be compared by taking the difference and dividing by the estimate of the standard deviation of the difference, point by point and separately for day and night. Figure 7 graphs the locations where the absolute values of t-statistics obtained exceed the 95 percent point of the Student-t distribution with 9 degrees of freedom. There are not a lot, the proportions of point exceedances are .036 and .026, for day and night respectively to be contrasted with the nominal .05 . One doesn't notice much structure in where the exeedances are located.

The conclusion of the analysis is that with the data set and estimates considered, it is not possible to reject the hypothesis that a potential function exists that may be used to describe the motion of the elk.

6. DISCUSSION AND SUMMARY

A basic advantage of working with a potential function, H, is that H is scalar-valued, as opposed to the bivariate $\boldsymbol{\mu}$ of (2.4). That is one has to model but a single real-valued function. The function can include individual effects, eg. attraction, repulsion, barriers, and this will be done in future work. The estimates computed here are nonparametric. In practice the results obtained can be expected to sometimes suggest particular explanatories to include in parametric forms.

A disadvantage of the work is that such an H may not exist. In such a case one needs an alternate method of estimating $\pi(.)$ given estimates of the parameters of the SDE. The concept of potential comes from the "much simpler" physical sciences. The motion of complex biological entities is surely many times more complicated than that of a falling ball, for example. A further difficulty arises in the drawing of conclusions. An elk's locations are available at successive time points, but they are 1-4 hours apart. The elk can be many different places between the times at which locations are estimated. This complication showed itself in Figure 2, where the track plotted would suggest that Elk 42 jumped some fences.

The assumption of a potential function led to the setting down of a stochastic differential equation for a diffusion process. Such an SDE assumption was needed both in motivating the estimates computed and in estimating the potential function itself. But diffusion processes are Markov, whereas more realistic equations would involve time lags and the process therefore not be Markov.

Some related results are presented in [21]. In current work the SDE approach is being further developed as a convenient way to include covariates in the models.

ACKNOWLEDGEMENT

This paper is dedicated to Don Fraser. But for DASF one of us (DRB) would almost certainly not be a statistician. Don's force field gently deflected DRB from the road towards actuarial science. The research we present concerns the movement of wildlife. From his cottages by Lake Temagami and Georgian Bay, Don has certainly watched lots of wildlife in motion. He has surely constructed some effective models for movement, albethey involving loons and nonanalytic.

Moe Hirsch and George Oster made some comments that helped this research along. The Referee's comments improved the discussion and presentation.

This research is supported in part by the National Science Foundation Grant DMS-9704739.

REFERENCES

[1] Banon, G. and Nguyen, H. T. (1981). Recursive estimation in diffusion model. *SIAM J. Control and Optimization* **19** 676-685.

[2] Bertrand, P. (1998). Comparison de l'erreur quadratique moyenne intégrée pour différents estimateurs du coefficient de diffusion d'un processus. *C. R. Acad. Sci. Paris* **327** 399-404.

[3] Bhattacharya, R. N. and Waymire, E. (1990). *Stochastic Processes with Applications.* Wiley, New York.

[4] Brillinger, D. R. and Stewart, B. S. (1998). Elephant seal movements: modelling migration. *Canadian J. Statistics* **26** 431-443.

[5] Burgière, P. (1993). Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus diffusion multidimensionelle. *Annales de l'Institut Henri Poincaré, Section B, Calcul des Probabilities et Statistique* **29** 357-389.

[6] Clark, J. D., Dunn, J. E., Smith, K. G. (1993). A multivariate model of female

black bear habitat use for geographic information system. *Jour. Wild. Manage.* **57** 519-526.

[7] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1992). Local regression models. Pp. 309-376 in *Statistical Models in S* (Eds. J. M. Chambers and T. J. Hastie). Pacific Grove, Wadsworth.

[8] Dohnal, G. (1987). On estimating the diffusion coefficient *Journal of Applied Probability* **24** 105-114.

[9] Dunn, J. E. and Brisbin, I. L. (1985). Characterization of the multivariate Ornstein-Uhlenbeck diffusion process in the context of home range analysis. *Statistical Theory and Data Analysis.* K. Matusita [ed.]. Elsevier Science Publishers B.V. North-Holland. 181-205.

[10] Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics* **33** 85-101.

[11] Efron,B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

[12] Findholdt, S. L., Johnson, B. K., Bryant, L. D. and Thomas, J. W. (1996). Corrections for position bias of a LORAN-C radio-telemetry system using DGPS. *Northwest Science* **70** 273-280.

[13] Genon-Catalot, V., Laredo, C., Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics* **19** 317-335.

[14] Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Linear Models.* Chapman and Hall, London.

[15] Heyde, C.C. (1994). A quasi-likelihood approach to estimating parameters in diffusion-type processes. Essays in Honour of Takacs. *Journal of Applied Probability*, **31A** 283-290.

[16] Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes.* Academic, New York.

[17] Kloeden, P. E. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations.* Springer, New York.

[18] Moorcroft, P. R., Lewis, M. A. and Crabtree, R. L. (1999). Home range analysis using a mechanistic home range model. *Ecology* **80** 1656-1665.

[19] Nelson, E. (1967). *Dynamical Theories of Brownian Motion.* Princeton U. Press, Princeton.

[20] Preisler, H. K. and Akers, P. (1995). Autoregressive-type models for the analysis of bark beetle tracks. *Biometrics* **51** 259-267.

[21] Preisler, H. K., Brillinger, D. R., Ager, A. A. and Kie, J. G. (1999). Analysis of animal movement using telemetry and GIS data. *Proc. ASA Section on Statistics and the Environment.*

[22] Prohorov, Yu. V. and Rozanov, Yu. A. (1969). *Probability Theory.* Springer-Verlag, New York.

[23] Rowland, M. M., Bryant, L. D., Johnson, B. K., Noyes, J. H., Wisdom, M. J. and Thomas, J. W. (1997). *The Starkey Project: History, Facilities, and Data Collection Methods for Ungulate Research.* Technical Report PNW-GTR-396, Forest Service, USDA.

[24] Sorensen, M. (1997). Estimating functions for discretely observed diffusions: a review. Pp. 305-326 in *Selected Proceedings of the Symposium on Estimating Functions.* Eds. Basawa, I. V, Godambe, V. P. and Taylor, R. L. Vol. 32 Lecture Notes. Institute of Mathematical Statistics, Hayward.

[25] Spivak, M. (1965). *Calculus on Manifolds.* Benjamin, New York.

[26] Stewart, J. (1991). *Calculus, Early Transcendentals.* Brooks/Cole, Pacific Grove.

[27] Turchin, P. (1998). *Quantitative Analysis of Movement.* Sinauer, Sunderland, MA.

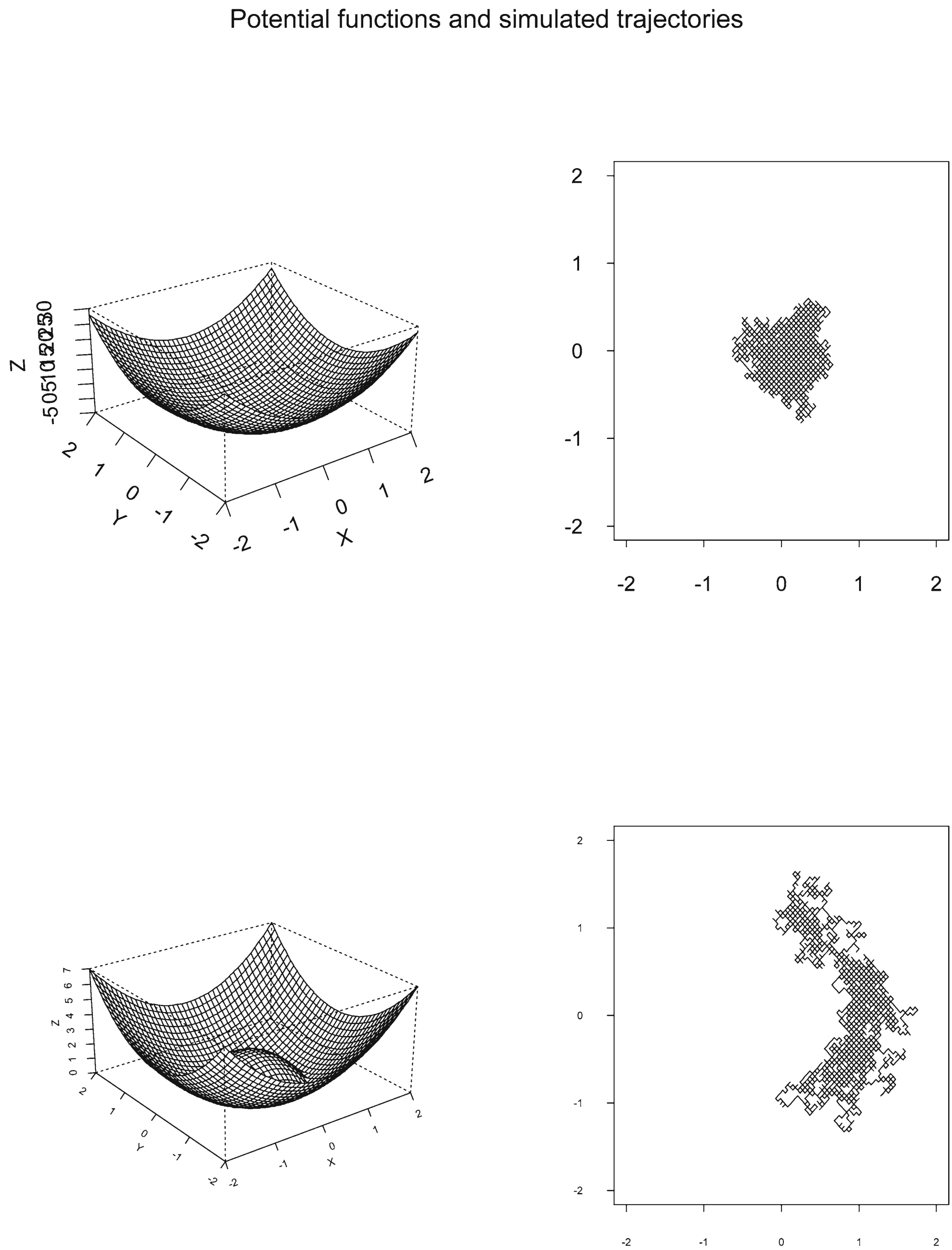


FIG. 1. *Examples of trajectories corresponding to the potential functions of the lefthand column.*

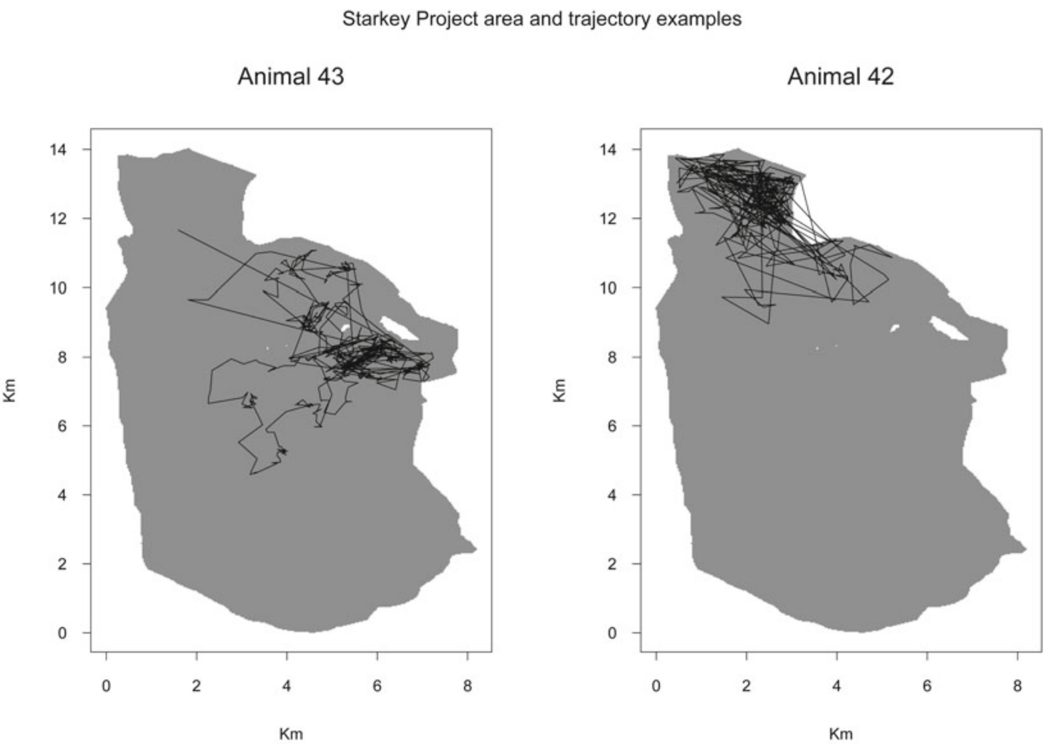


FIG. 2. *Points along the trajectories of two elk.*

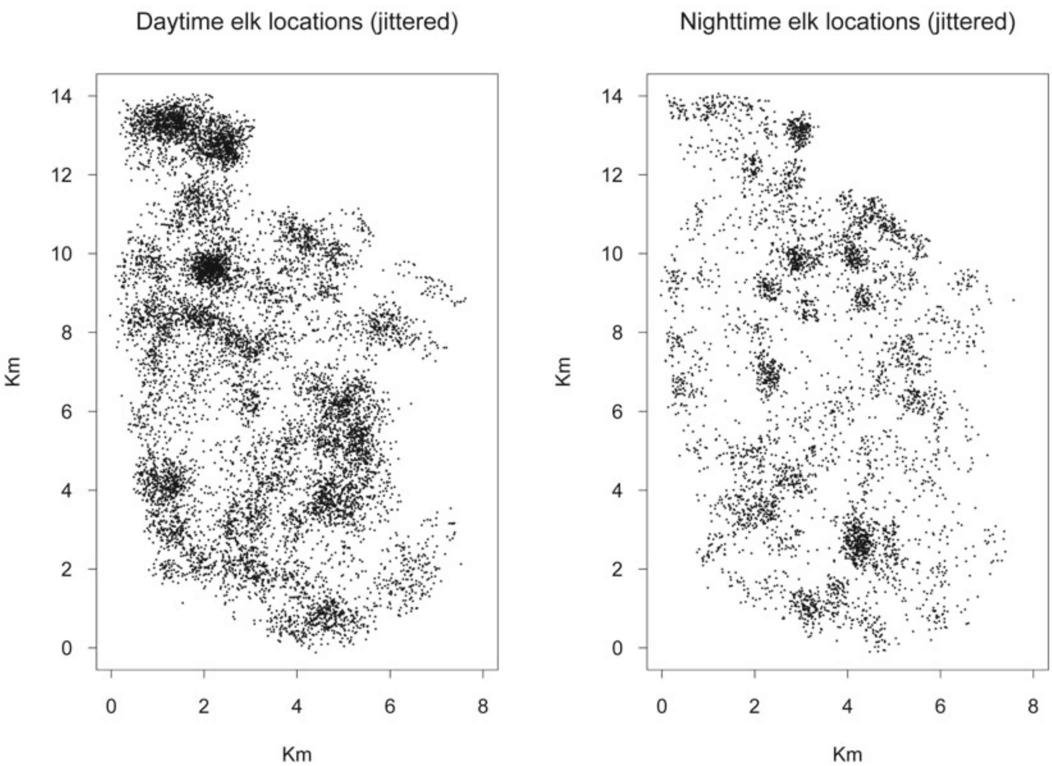


FIG. 3. *Locations visited by all 53 collared elk.*

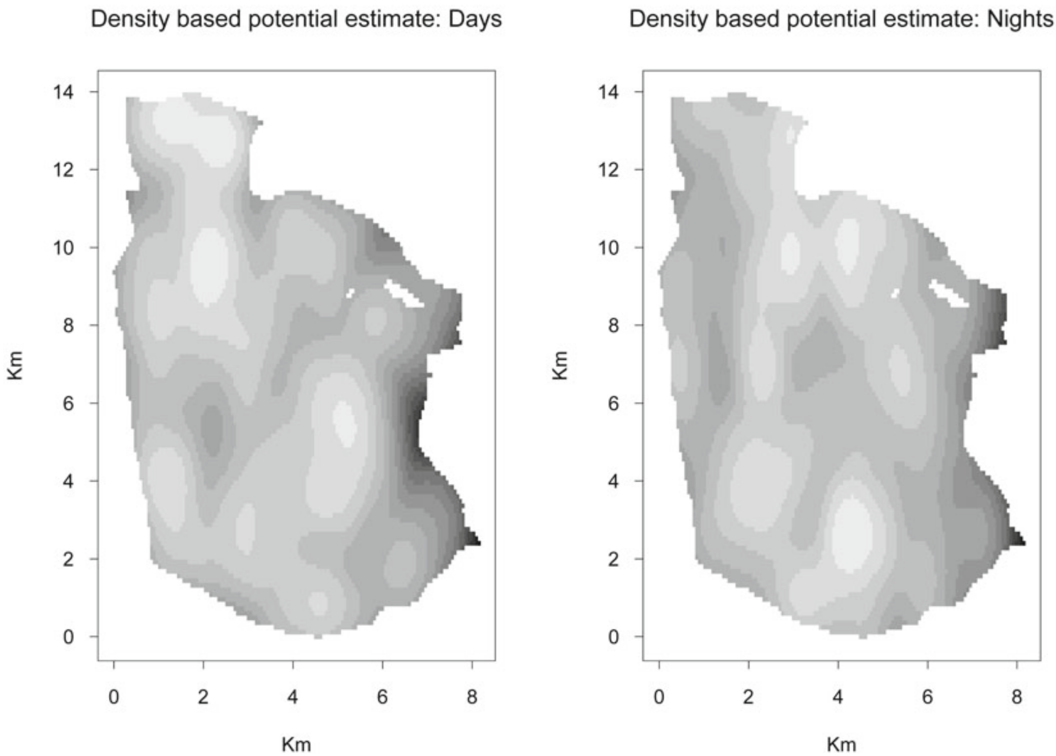


FIG. 4. *Potential function estimate based on density estimate.*

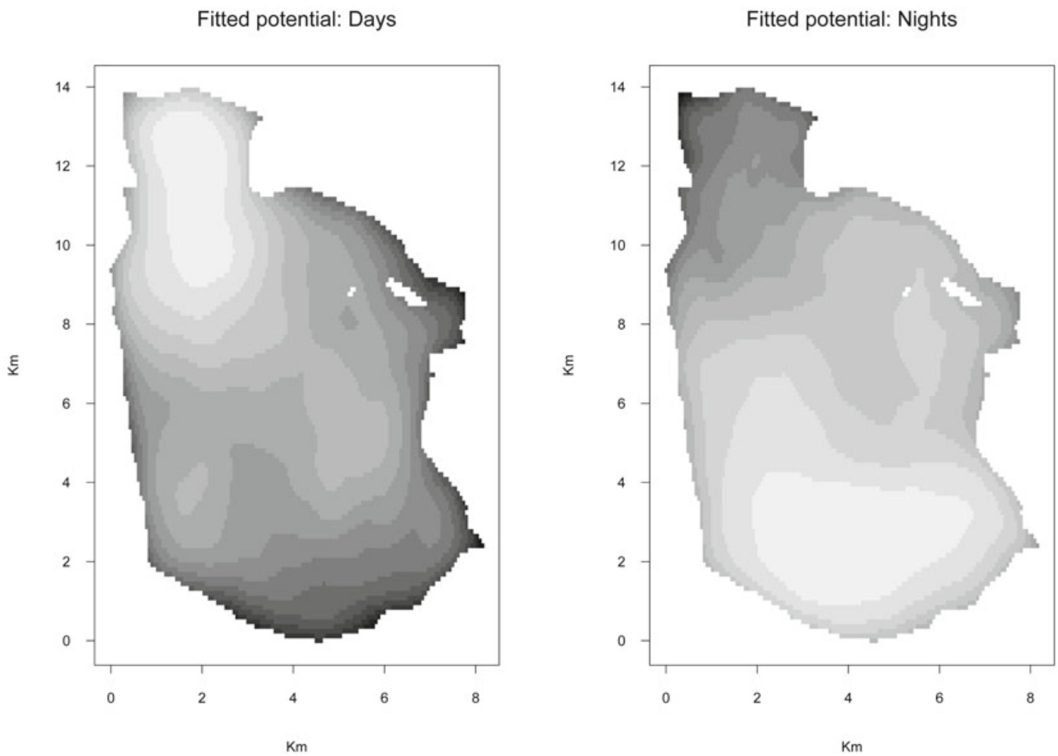


FIG. 5. *Potential function estimate obtained by simulating realizations using $\hat{H}_x, \hat{H}_y$.*

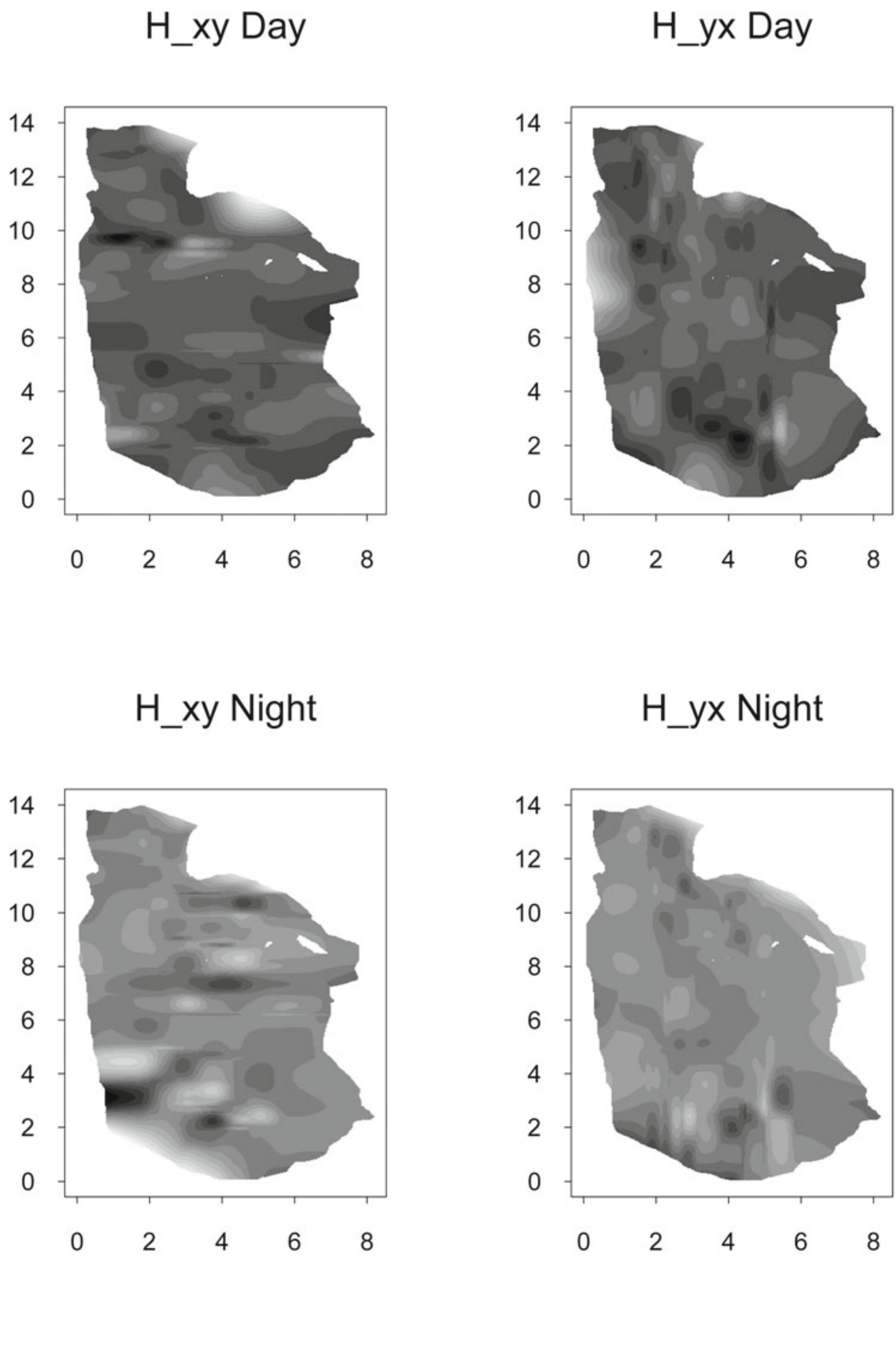


FIG. 6. *Estimates of the second-order mixed partial derivatives.*

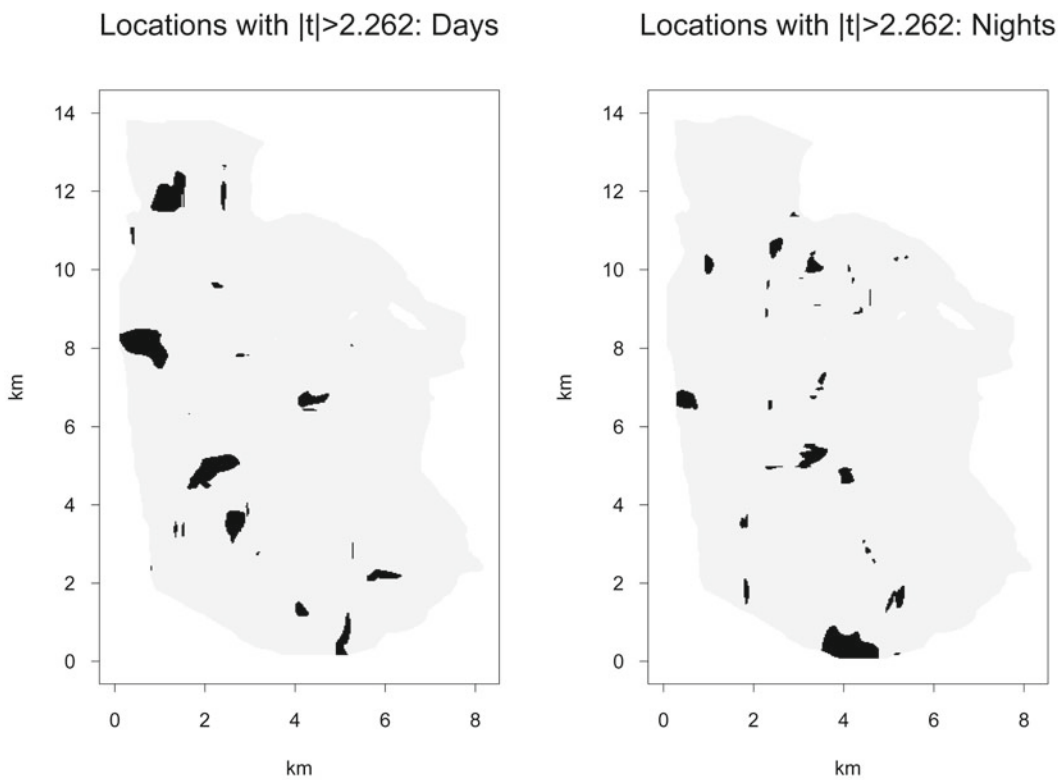


FIG. 7. *Locations where the t-statistic exceeds the 95 percent null value*

The Canadian Journal of Statistics
Vol. 26, No. 3, 1998, Pages 431–443
La Revue Canadienne de Statistique

Elephant-seal movements: Modelling migration*

David R. BRILLINGER and Brent S. STEWART

University of California, Berkeley (D.R.B., B.R.S.) and
Hubbs-Sea World Research Institute (B.R.S.)

Key words and phrases: Animal migration, diffusion process, drift, elephant seal, great-circle route, likelihood, spherical motion, stochastic differential equation.
AMS 1991 subject classifications: 62M05, 62M10, 62M30, 62B99.

ABSTRACT

Elephant seals migrate over vast areas of the eastern North Pacific Ocean between rookeries in southern California and distant northern foraging areas. Several models of particle movement were evaluated and a model for great-circle motion found to give reasonable results for the movement of an adult female. This model takes specific account of the fact that the movement is on the surface of a sphere and that the animal is apparently heading toward a particular destination. The parameters of the motion were estimated. Such a great-circle path of migration may imply that these seals have the ability to assess their position with respect to some global or celestial cues, allowing them to continually adjust their course and achieve the most direct geodesic route between origin and destination of migration. But the navigational mechanism actually used by these seals to accomplish such feats is as yet unknown.

RÉSUMÉ

Deux fois par année, les éléphants de mer entreprennent de longues migrations au nord de l'océan Pacifique. Plusieurs sont porteurs d'instruments qui enregistrent la profondeur et l'intensité lumineuse à intervalles réguliers. Ces instruments sont ensuite récupérés et permettent de faire plusieurs estimations, par exemple les positions à mi-journée. Dans cet exposé on s'interessera à la modélisation des itinéraires de surface des animaux à l'aide d'équations différentielles stochas-tiqeus. Les distances sont suffisament importantes pour être incluses dans le modèle la nature sphérique de la surface terrestre. Une question intéressante est de déterminer si les itinéraires sont des grands cercles de la sphère terrestre.

1. INTRODUCTION

Many marine mammals travel great distances each year between breeding and calving areas and seasonally productive foraging areas. Northern elephant seals (*Mirounga angustirostris*), for example, are exceptional migrators. They spend most of each year at sea and range over vast areas of the eastern North Pacific Ocean during double annual migrations between California rookeries and distant northern foraging areas (Stewart and DeLong 1995, Stewart 1996). Similarly, southern elephant seals (*Mirounga lemina*)

*This research was supported by Office of Naval Research Grant N00014-94-1-0042 and National Science Foundation Grants DMS-9625774, 97-04739 (D.R.B.). Elephant-seal dive data were collected in previous studies with partial support of a contract of B.S.S. from the Space and Missile Command, U.S. Department of the Air Force. Part of the material was presented in the 1997 President's Invited Talk, Annual Meeting of the Statistical Society of Canada, Fredericton, New Brunswick.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_23, © Springer Science+Business Media, LLC 2012

range over vast areas of the Southern Ocean (e.g., McConnell and Fedak 1996, Bester and Pansegrouw 1992). Although the navigational mechanisms involved in these remarkable migrational feats are as yet unknown, an initial step of describing the migratory trajectories with various formal models may help to develop testable hypotheses. One possibility is that the seals follow great-circle paths. If so, this would imply that they are able to assess their position relative to some astronomical or global magnetic background and constantly make course corrections, as do oceangoing ships when navigating, to achieve the shortest geodesic distance provided by such a route. Elephant seals dive and forage continuously while migrating. Such behaviour could pull them away from the direct route from origin to destination, and it could be modelled as stochastic fluctuation. Here the fit of the great-circle model of particle movement is evaluated for a particular northern elephant seal, in part to examine the hypothesis that such animals can migrate along great-circle paths.

The top graph of Figure 1 presents the surface track for one seal during the post-breeding-season migration. This figure led to speculation that the seals would sometimes follow a great-circle path. A great-circle path is indicated in the bottom graph for reference.

The paper first mentions some of the work of previous authors on the stochastic modelling of particle tracks. Then some material concerning stochastic differential equations is recorded. Section 3 concerns the motion of a particle on the sphere for the case of the particle heading towards a particular destination. Section 4 focuses on the problem of estimating the parameters of the spherical motion. The next section reviews the data and data-collection procedures. Section 6 describes the analysis and presents results, the principal one being an examination of the hypothesis that the motion is a great circle. The statistical analysis presented involves a rotation of the spherical coordinates so that the destination is the North Pole, followed by a search for systematic departure of longitude changes from noise of mean 0. Section 7 provides some introductory remarks on dealing with measurement error. Finally there is discussion, an appendix on rotating spherical coordinates, and an appendix presenting the data.

2. MODELS FOR PARTICLE MOVEMENT

Various authors have employed random-walk models for animal movement. Some particular cases follow. Okubo (1980) devotes a chapter to the topic. Kareiva and Shigesada (1983) studied butterflies and caterpillars. A general reference is Levin (1986). McCulloch and Cain (1989) studied swallowtails, butterflies and goldenrod. Dunn and Gipson (1977) modelled deer movements, assuming that such data were generated by a multivariate Ornstein-Uhlenbeck diffusion process (see also Dunn and Brisbin 1985). Moore (1985) and Zwiers (1985) modelled iceberg movements as vector ARIMA processes. Some authors (e.g., Hadeler *et al.* 1980, Niwa 1996), sought to describe annual movements by variants of Newton's equations of motion, with Niwa evaluating fish movements. Preisler and Akers (1995) employed an autoregressive scheme to model the heading of a bark beetle attracted towards a source. Malik *et al.* (1994) investigated the motion of microtubules. Oceanographers have studied drifting-buoy movements; see Brink *et al.* (1991). Wagner (1986) and Wehrhahn *et al.* (1982) studied the motion of one fly pursuing another. Bril (1995), in studying hurricane tracks, considered a state-space model with a randomly varying drift.

An original term for "stochastic process" is "trajectory", so it is interesting to be returning to the roots of the subject. Stochastic differential equations (SDEs) are a powerful tool for conceptualizing processes and investigating trajectories. These equations have

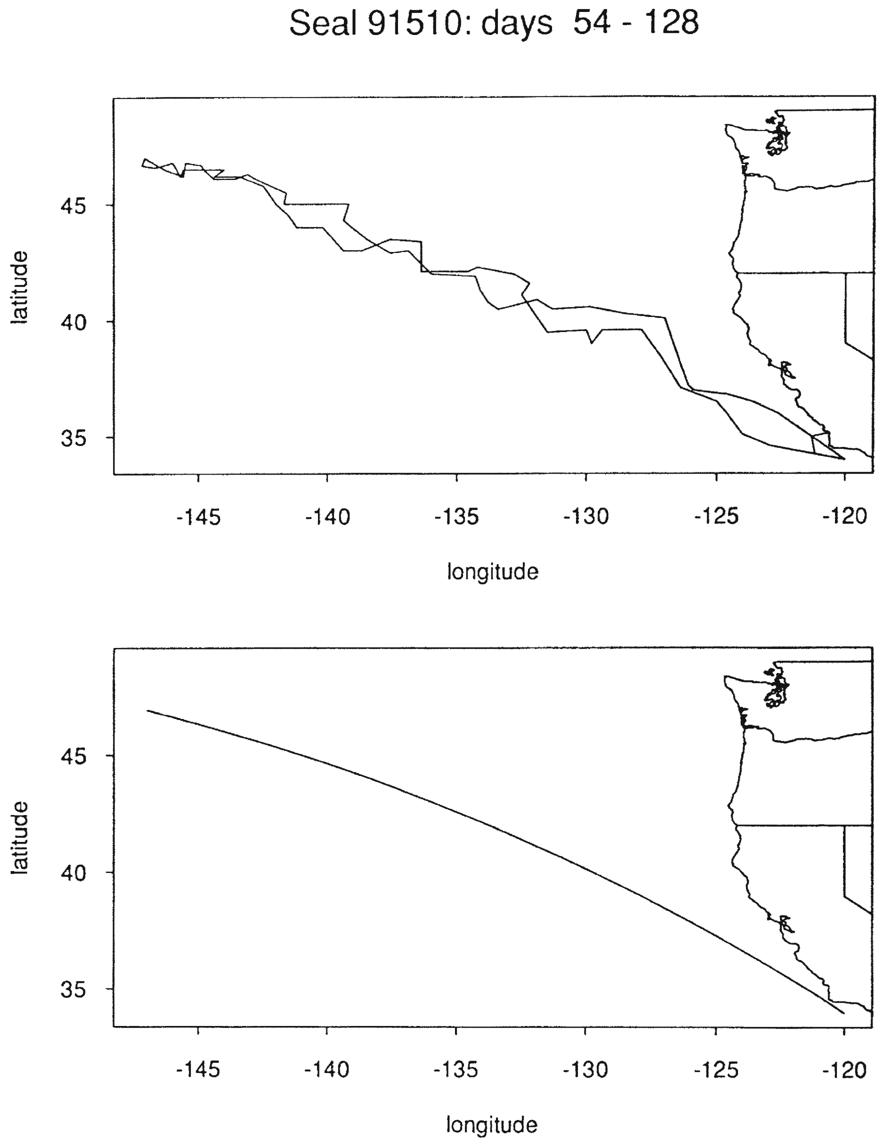


FIGURE 1: The top graph is the track of one seal heading from an island off Santa Barbara to a region in the Northwest Pacific and return. The bottom graph is a great-circle route, for reference.

some surprising properties. Their solutions, when continuous and Markov, are referred to as diffusion processes.

By way of introduction, consider representing a random walk in the plane by a bivariate Brownian. Letting (X_t, Y_t) represent a particle's location at time t, the SDE for the motion may be written

$$dX_t = \sigma\, dU_t, \tag{1}$$

$$dY_t = \sigma\, dV_t \tag{2}$$

with $\{U_t\}$ and $\{V_t\}$ independent standard univariate Brownians, i.e., Gaussian processes with mean 0 and covariance function $\min\{s,t\}$. Suppose one changes to polar coordinates, $r_t = \sqrt{X_t^2 + Y_t^2}, \phi_t = \text{atan}\,(Y_t, X_t)$; then the SDEs (1), (2) become

$$dr_t = \frac{\sigma^2}{2r_t}\,dt + \sigma\,dU_t, \tag{3}$$

$$d\phi_t = \frac{\sigma}{r_t}\,dV_t \tag{4}$$

via Itô's lemma (see Karlin and Taylor 1981, Bhattacharya and Waymire 1990, Oksendal 1995). The appearance of the drift term $\sigma^2/2r_t$ in (3) is perhaps surprising. This term dominates the behaviour of r_t near the origin, pushing the particle away. In the case of ϕ_t the change is highly variable when the particle is near the origin. The process $\{r_t\}$ is known as the two-dimensional Bessel process.

Now consider motion in $\mathbb{R}^3$. One important process, the Langevin or Ornstein-Uhlenbeck, is defined by the SDE

$$d\dot{\mathbf{X}}_t = -\beta\dot{\mathbf{X}}_t\,dt + \Gamma\,d\mathbf{B}_t \tag{5}$$

with $\mathbf{X}$ representing location, $\dot{\mathbf{X}}$ representing velocity, $-\beta\dot{\mathbf{X}}$ representing dynamical friction, Γ a 3-by-3 matrix and $\mathbf{B}_t$ Brownian motion in $\mathbb{R}^3$; see Chandrasekhar (1943) for example. It may be that the particle is moving in a force field, in which case a term $K(\mathbf{X}_t, t)\,dt$ is added to the right-hand side of (5).

What distinguishes the present work is that the particle is supposed to be heading for a specific destination. Kendall (1974) considered the case of a Brownian motion on the plane with an "attractive" polar drift. He worked with polar coordinates (r, ϕ) centered at the target center. The particle, in his case a bird, started at location $(D, 0)$. In a time interval of length dt it moved a distance $\delta\,dt$ towards the target, then was subject to random Gaussian disturbance, of amount $\sigma\,dU_t$ towards the target and amount $\sigma\,dV_t$ at right angles to the path. Here U_t and V_t are independent Brownians with variance σ^2. In Itô form the motion may be described by

$$dr_t = \left(\frac{\sigma^2}{2r_t} - \delta\right)dt + \sigma\,dU_t, \tag{6}$$

$$d\phi_t = \frac{\sigma}{r_t}\,dV_t. \tag{7}$$

These equations reduce to (3), (4) when $\delta = 0$.

For basic material on diffusion processes see Karlin and Taylor (1981), Bhattacharya and Waymire (1990) or Oksendal (1995). Papers and books on inferential aspects of diffusion processes include Basawa and Rao (1980), Burgière (1993), Dohnal (1987), Genon-Catalot *et al.* (1992), Heyde (1994).

3. DIFFUSION ON A SPHERE

The description of a particle moving randomly on the surface of a sphere has been considered by a number of authors, beginning with Perrin (1928). The infinitesimal generator and transition density for spherical Brownian motion were given in Yosida (1949). Following directly from the infinitesimal generator are the Itô SDEs

$$d\theta_t = \frac{\sigma^2}{2\tan\theta_t}dt + \sigma\,dU_t,$$

$$d\phi_t = \frac{\sigma}{\sin\theta_t}\,dV_t.$$

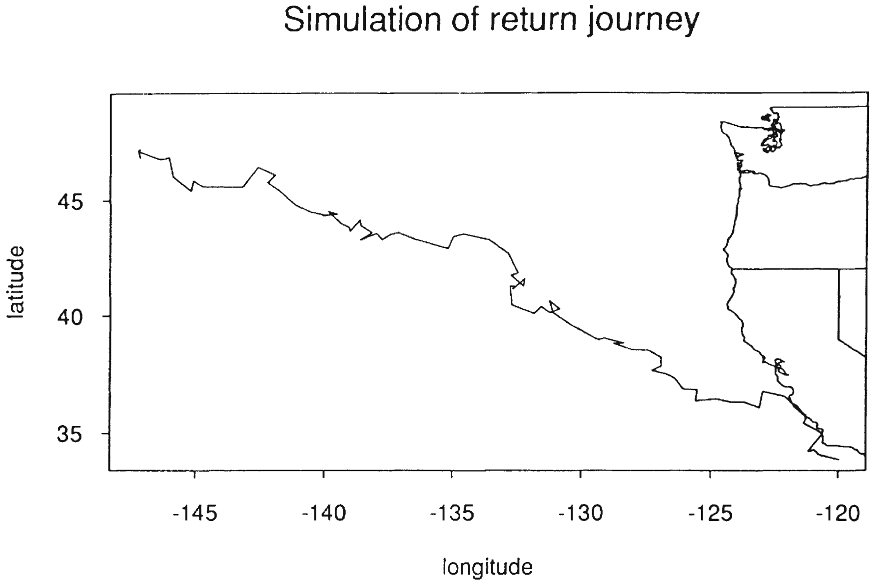


FIGURE 2: A simulation of the process (8, 9) for a seal heading back to the Channel Islands.

Suppose that a particle on the sphere is migrating directly towards the North Pole at speed δ and subject to Brownian disturbances. (The North Pole is taken for convenience.) In analogy with the model of Kendal (1974), the following Itô differential equations are set down in Brillinger (1997):

$$d\theta_t = \left(\frac{\sigma^2}{2\tan\theta_t} - \delta\right) dt + \sigma\, dU_t, \tag{8}$$

$$d\phi_t = \frac{\sigma}{\sin\theta_t} dV_t, \tag{9}$$

so long as $\theta_t \neq 0$ and with ϕ_t defined mod 2π. It will be supposed that the particle does not start at $\theta = 0$ or π. The latitude, θ_t, is analogous to r_t of (6, 7). If one considers a sphere of infinite radius, the planar and spherical formulations coincide.

Because distances around a constant latitude decrease with increasing latitude, the $1/\sin\theta$ term appears in (9). Figure 2 presents a simulation of the process (8), (9) meant to represent a return trip of a seal to the Channel Islands off southern California. The standard error σ here has been taken to be 0.005 rad.

Consider, for example, the expected travel time for the process (8), (9). Suppose the particle starts at $\cos\theta = x$ and heads to $\cos\theta = d$, $1 > d > x > -1$. In Brillinger (1997) it is shown that the expected travel time is given by

$$\int_x^d \frac{2}{\sigma^2} \int_{-1}^{y} \exp\left(-\frac{2\delta}{\sigma^2}\cos^{-1} z\right) dz \exp\left(\frac{2\delta}{\sigma^2}\cos^{-1} y\right) \frac{1}{1-y^2}\, dy, \tag{10}$$

which may be evaluated in specific cases.

4. ESTIMATION

Following Brillinger (1997), the log likelihood ratio of the process, relative to that of the case $\delta = 0$, is

$$\frac{1}{\sigma^2}\left\{(-\delta)\int_0^T d\theta_s - \frac{1}{2}\int_0^T \left(-\frac{2\delta\sigma^2}{\tan\theta_s} + \delta^2\right) ds\right\} \tag{11}$$

In the case that σ is known this leads to the maximum-likelihood estimate

$$\hat{\delta} = \frac{1}{T}\left((\theta_0 - \theta_T) + \sigma^2 \int_0^T \frac{1}{\tan\theta_s}\, ds\right). \tag{12}$$

Because the particle reaches the region of its destination eventually, this estimate becomes unreasonable in practice if $T \longrightarrow \infty$.

One can actually obtain an exact estimate of σ^2; specifically, it is the case that

$$\sum_i (\tilde{\phi}_{t_{i+1}} - \tilde{\phi}_{t_i})^2 \xrightarrow{P} \sigma^2 \int_0^T \frac{1}{\sin^2\theta_s}\, ds. \tag{13}$$

Here $\{t_i\}$ is a partition of the interval that gets finer under the limiting process. The stated result is conditional on the given (continuous) realization of $\theta_s, 0 \leq s \leq T$, and it is assumed that there exists $\epsilon > 0$ such that $|\sin\theta_s| \geq \epsilon$. The curve $\tilde{\phi}_t$ refers to a continuous curve obtained from the curve ϕ_t either by patching together continuous segments or by reflecting ϕ_t whenever it reaches the barriers $\phi = 0, \pi$. (It is assumed that $0 < \phi_0 < 2\pi$.)

In practice the data will be available at discrete time points and the above likelihood ratio (11) is not available. However, with a model such as (14)–(15) below, describing the position of the particle's successive time steps, one can set down the likelihood function and obtain estimates of the parameters. An approximate approach is to do what a ship's navigator has done traditionally. Specifically, at the start of a day, based on a ship's position, the navigator determines the heading of the great-circle course. That heading is followed for the whole day. The next day the navigator determines the ship's new position, then the great-circle course based on that position. The new heading is followed for that day. Unless the ship is heading due north or south, during its travels it will be pulled off the great circle route, but with the course revisions the destination is approached. This method leads to approximating the desired conditional density by a succession of planar motions with different headings.

A discrete approximation to the model (8), (9) is provided by

$$\theta_{t+1} - \theta_t = \frac{\sigma^2}{2\tan\theta_t} - \delta + \sigma\epsilon_{t+1}, \tag{14}$$

$$\phi_{t+1} - \phi_t = \frac{\sigma}{\sin\theta_t}\eta_{t+1}, \tag{15}$$

$t = 0, 1, 2, \ldots$, with the errors independent standard white noise processes and the ϵ_t, η_t independent normal with mean 0 and variance 1. One notes that the conditional expected value of θ_{t+1} given the past is $-\delta + \sigma^2/(2\tan\theta_t)$ and that the conditional variances of the increments are σ^2 and $\sigma^2/(\sin^2\theta_t)$ respectively. Estimates of the parameters may be derived by the method of moments or by maximizing the likelihood. In this discrete case an "exact" estimate of σ^2 is not available. Then minus twice the log likelihood is

$$2T\log\sigma^2 + \frac{1}{\sigma^2}\sum(\sin^2\theta_t)(\phi_{t+1} - \phi_t)^2 + \frac{1}{\sigma^2}\sum\left(\theta_{t+1} - \theta_t + \delta - \frac{\sigma^2}{2\tan\theta_t}\right)^2,$$

which may be minimized to obtain estimates of δ and σ. Such estimates will be presented for the data of Figure 1.

5. THE DATA

The data studied in the present work are from the postbreeding migration of an adult elephant-seal female (*Mirounga angustirostris*). This species breeds on offshore islands and at a few mainland sites along the coasts of California and Baja California (Stewart and Huber 1993, Stewart *et al.* 1994, Stewart 1996). Adults are ashore briefly in winter to breed, and again in spring (females) or summer (males) to molt, but they spend the remainder of the year, 8–10 months, at sea foraging. They make two precise, long-distance migrations each year between islands in southern California and offshore foraging locations in the mid North Pacific, in the Gulf of Alaska and along the Aleutian Islands, covering 18,000 to 20,000 km (surface movements alone) during the double migrations (Stewart and DeLong 1995). The navigational mechanisms employed by these superlative migrators are, as yet, unknown.

The data on diving and movements studied were obtained by a microprocessor-controlled event recorder which was harmlessly glued to a seal's hair (e.g., Stewart and DeLong 1995, Bengtson and Stewart 1992, Stewart *et al.* 1989). The instrument was attached at the end of the breeding season and then recovered when the animal returned to land several months later.

An estimate of daily location was computed from measurements of ambient daylight made and stored in the recording instruments. Briefly, estimates of sunrise, sunset, and local apparent noon were made from those data, and then latitude and longitude were computed [see DeLong *et al.* (1992) and Stewart and DeLong (1995) for description of methods]. The error varies with season and latitude.

The movement data for the journey of the seal studied in our work are given in Appendix B. It is to be noted that days 85 and 111 are missing. This was handled in this preliminary study by simply using the average of the adjacent values. Brillinger and Stewart (1996) carry out some frequency-domain studies of the series of depth values recorded during this particular migration, and Brillinger and Stewart (1997) develop typical shapes for individual dives and study their temporal occurrence.

6. RESULTS OF SOME ANALYSES

To begin, consider the path of the top graph of Figure 1. Figure 3 provides a corresponding smoothed path. This smooth path was determined via the procedure loess of Cleveland *et al.* (1990). One notes the bowing of the route typical of great-circle travel. The variability represented in Figure 1 represents both foraging movements and measurement error for location.

For the next analysis it is necessary to take note of the fact that the seal's positions are given in latitude and longitude with the destination not the North Pole as was assumed the model in (8), (9). Appendix A indicates the formulae for the necessary change of coordinates to make the data correspond to the North Pole model. The rotated coordinates are denoted by $\tilde{\phi}_t$ and $\tilde{\theta}_t$.

The model (14), (15) was fitted to the outbound and inbound daily positions, merged appropriately, by minimizing the minus twice the log likelihood (16). The estimates obtained are

$$\hat{\delta} = -0.0113 \text{ rad/day} = -72.0 \text{ km/day},$$
$$\hat{\sigma} = 0.00805 \text{ rad/day} = 51.3 \text{ km/day}. \tag{16}$$

The estimated standard error of $\hat{\delta}$ is 0.0011.

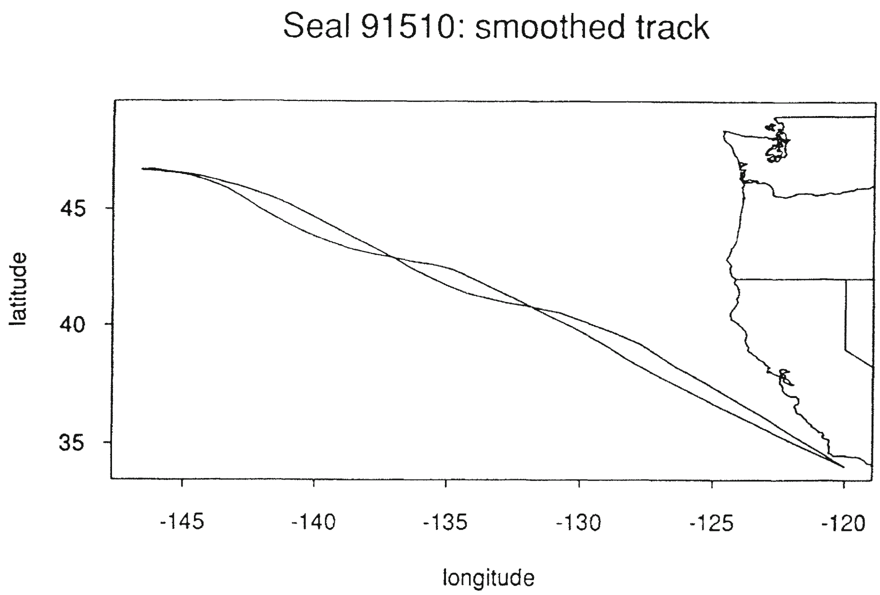


FIGURE 3: Smoothed track of seal 91510.

Figure 4 plots the values

$$\frac{(\sin \tilde{\theta}_t)(\tilde{\phi}_{t+1} - \tilde{\phi}_t)}{\hat{\sigma}} \tag{17}$$

for $t = 0, 1, \ldots$, separately for the outbound and inbound trips. This plot provides a means to examine the hypothesis of a great-circle route. For the great-circle case the points plotted should simply fluctuate about 0. A smoothed loess line has been added to each figure to provide an estimate of some systematic route. Also graphed are ± 2-standard-error levels placed about 0. One does not see evidence against the great-circle hypothesis.

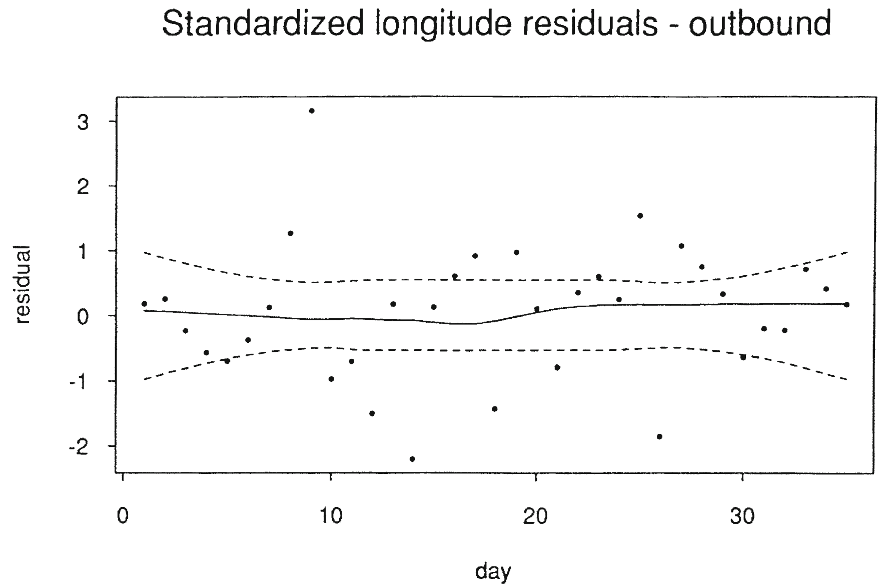
In these computations the procedure adopted is to act as if the uncertainty in the destinations is negligible. The seal appears to have the location of its rookery specifically in mind when it begins the return movement, so the assumption is certainly reasonable then. In the case of the outbound trip the destination was taken as the average of the extreme points in the Northwest.

7. MEASUREMENT NOISE

A difficulty is the presence of measurement noise. It and the foraging variability are confounded in the above analysis. One way to take note of measurement error is to set down the additional equations

$$\begin{aligned} \theta_t' &= \theta_t + \epsilon_t', \\ \phi_t' &= \phi_t + \eta_t'/\sin \theta_t' \end{aligned} \tag{18}$$

with (θ_t', ϕ_t') now representing the available data and supposing ϵ_t', η_t' noise. If these last are assumed independent normals with mean 0 and variances τ^2, then, amongst other procedures, a Kalman-filter-type analysis may be employed to develop a full likelihood and corresponding estimates. The results of this analysis are presented in Brillinger (1998). The Kalman filter is employed with wildlife data in Anderson-Sprecher (1994) and Anderson-Sprecher and Ledolter (1991).



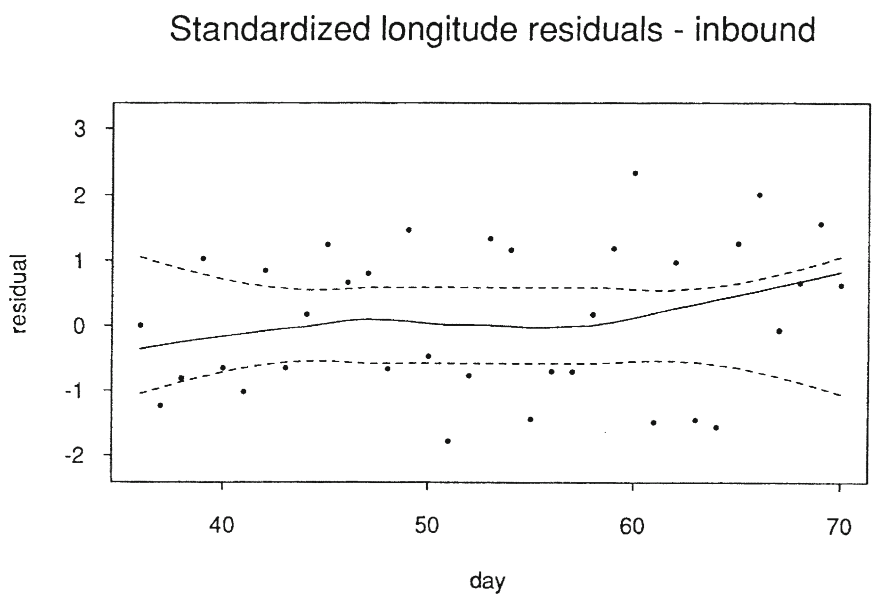


FIGURE 4: The scaled longitude differences of (17) with a smoothed line as produced by loess. The dashed lines are ±2 standard error limits about 0.

8. DISCUSSION

Future work will incorporate explanatory variables in the model, will employ a recursive filter, will better handle the missing values and will analyze other data sets.

The great-circle path hypothesis was not contradicted by the immigration of one northern elephant seal female. The results suggest that a great-circle path model is a possible navigational strategy in this species. They also suggest that the seals have a destination in mind when departing from an origin (i.e., terrestrial rookery or haulout and pelagic foraging area) and that they are able to continually adjust course en route to achieve the most direct route. Further, they imply that natural selection has favoured the development of neural and sensory mechanisms that permit great-circle navigation. However, the sensory clues actually used are as yet unknown, although several have been suggested and studied to various extents in a variety of other animal taxa (e.g., Able 1996, Dingle 1996, Dittman and Quin 1996, Lohmann and Lohmann 1996, Wehner *et al.* 1996, Weindler *et al.* 1996, Wiltschko and Wiltschko 1996).

Navigation by learned reference to geophysical characteristics would seem to play only a minor role, as elephant seals are generally far from coastlines and in areas of great water depth and little submarine features during most of their migrations. But the fit of a great-circle model suggests that some kind of compass may be central to the seals' rather precise migration and navigational performances, allowing them to continually determine the appropriate direction of each subsequent movement to keep en route to the shortest distance between origin and destination. Celestial navigation may be involved to some extent, but the brief and sporadic appearance of migrating seals at the sea surface, where such clues could be assessed, and their propensity to travel mostly at great depths, where such cues are obscured, would argue that it is not a primary mechanism. Large-scale magnetic field orientation may be the most plausible of potential compasses. But the rather precise navigation of the seals may also imply either the existence of a cognitive map to apply the compass to or perhaps simply remarkable fidelity to vectors and assessment of distance travelled, independent of any map. At present, no such mechanism of magnetic sensory ability or cognitive mapping is known for elephant seals. However, knowledge of the ecological and physiological conditions under which northern elephant seals find their way while migrating and foraging, which have come to be known recently (e.g., Stewart and DeLong 1995, Stewart 1996), coupled with the descriptive theoretical model of navigational strategy developed here, can help focus questions properly on navigational and orientational mechanisms in this and other long-distance, deep-dwelling ocean migrators.

APPENDIX A. THE CHANGE OF COORDINATES

A transformation $(\phi, \theta) \longrightarrow (\tilde{\phi}, \tilde{\theta})$ is constructed. Suppose that the sphere is rotated so that the particular point (Φ, Θ) becomes the North Pole $(0, 0)$, and the great circle (ϕ, θ) to (Φ, θ) becomes the great circle $(0, 0)$ to $(0, \tilde{\theta})$. The required change of variables may be derived to be

$$\cos \tilde{\theta} = \cos \Theta \cos \theta + \sin \Theta \sin \theta \cos (\phi - \Phi),$$

$$\tan \tilde{\phi} = \frac{\sin \theta \sin (\phi - \Phi)}{\cos \Phi \sin \theta \cos (\phi - \Phi) - \sin \Theta \cos \theta}.$$

Retaining the signs of the numerator and denominator in the last expression will lead to an appropriate choice of quadrant for the transformed longitude.

APPENDIX B. NUMERICAL DATA

The data are shown in Table 1.

ACKNOWLEDGEMENT

We thank Haiganoush Preisler and the two referees for some very helpful comments.

TABLE 1

Day	Latitude (N)	Longitude (W)	Day	Latitude (N)	Longitude (W)
54	34.0	120.0	92	46.5	146.3
55	35.0	121.3	93	47.0	147.1
56	36.0	122.6	94	46.7	147.2
57	36.5	123.6	95	46.6	146.7
58	36.8	124.6	96	46.8	146.0
59	36.9	125.4	97	46.2	145.7
60	37.0	125.9	98	46.5	145.6
61	37.2	126.1	99	46.5	144.0
62	38.1	126.4	100	46.2	144.4
63	40.1	127.0	101	46.2	143.4
64	40.3	128.5	102	45.8	142.5
65	40.6	129.9	103	45.0	142.0
66	40.5	131.3	104	44.5	141.5
67	40.9	131.9	105	44.0	141.2
68	40.5	133.4	106	44.0	140.2
69	40.8	133.8	107	43.0	139.4
70	41.3	134.1	108	43.0	138.7
71	41.9	134.3	109	43.5	137.6
72	42.0	136.0	110	43.4	136.4
73	42.9	136.8	112	42.1	136.4
74	43.0	136.9	113	42.1	134.6
75	42.9	137.6	114	42.3	134.2
76	43.5	138.5	115	42.0	132.8
77	44.1	139.2	116	41.6	132.2
78	44.3	139.4	117	41.1	132.5
79	45.0	139.2	118	39.5	131.5
80	45.0	141.7	119	39.6	130.0
81	45.5	141.6	120	39.0	129.8
82	46.1	142.8	121	39.6	129.4
83	46.3	143.1	122	38.5	127.2
84	46.1	143.6	123	37.1	126.4
86	46.1	144.4	124	36.5	125.0
87	46.5	144.8	125	36.0	124.6
88	46.7	144.9	126	35.0	124.0
89	46.8	145.5	127	34.6	122.9
90	46.8	145.5	128	34.0	120.0
91	46.2	145.6			

REFERENCES

Able, K.P. (1996). The debate over olfactory navigation by homing pigeons. *J. Exp. Biol.*, 199, 121–124.

Anderson-Sprecher, R. (1954). Robust estimates of wildlife location using telemetry data. *Biometrics*, 50, 406–416.

Anderson-Sprecher, R., and Ledolter, J. (1991). State-space analysis of wildlife telemetry data. *J. Amer. Statis. Assoc.*, 86, 596–602.

Basawa, I.V., and Rao, B.L.S.P. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, New York.

Bengtson, J.L., and Stewart, B.S. (1992). Diving and haulout patterns of crabeater seals in the Weddell Sea, Antarctica during March 1986. *Polar Biol.*, 12, 635–644.

Bester, M.N., and Pansegrouw, H.M. (1992). Ranging behaviour of southern elephant seal cows from Marion Island. *South African J. Sci.*, 88, 574–575.

Bhattacharya, R.N., and Waymire, E. (1990). *Stochastic Processes with Applications*. Wiley, New York.

Bril, G. (1995). Forecasting hurricane tracks using the Kalman filter. *Environmetrics*, 6, 7–16.

Brillinger, D.R. (1997). A particle migrating randomly on a sphere. *J. Theoret. Probab.*, 10, 429–443.

Brillinger, D.R. (1998). Some examples of random process environmental data analysis. *Handbook of Statistics, Vol. 17*. North-Holland, Amsterdam. To appear.

Brillinger, D.R., and Stewart, B.S. (1996). Elephant seal movements: some frequency based studies. *Brazilian J. Probab. Statist.*, 10, 15–33.

Brillinger, D.R., and Stewart, B.S. (1997). Elephant seal movements: Dive types and their sequences. *Modelling Longitudinal and Spatially Correlated Data* (T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohe, W.G. Warren, R.D. Wolfinger, *eds.*), Lecture Notes in Statistics 122, Springer-Verlag, New York, 275–288.

Brink, K.H., Beardsley, R.C., Niiler, P.P., Abbott, M., Huyer, A., Ramp, S., Stanton, T., and Stuart, D. (1991). Statistical properties of near-surface flow in the California coastal transition zone. *J. Geophys. Res.* 96, 14693–14706.

Burgière, P. (1993). Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus de diffusion multidimensionelle. *Ann. Inst. H. Poincaré Probab. Statist.*, 29, 357–389.

Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. *Rev. Modern Phys.*, 15, 1–89.

Cleveland, W.S., Grosse, E., and Shyu, W.M. (1990). Local regression models. *Statistical Models in S* (J.M. Chambers and T.J. Hastie, *eds.*), Wadsworth, Pacific Grove, Calif., 300–376.

DeLong, R.L., Stewart, B.S., and Hill, R.D. (1992). Documenting migrations of northern elephant seals using day length. *Marine Mammal Sci.* 8, 155–159.

Dingle, H. (1996). *Migration: The Biology of Life on the Move*. Oxford Univ. Press, Oxford.

Dittman, A.H., and Quin, T.P. (1996). Homing in Pacific salmon: Mechanisms and ecological basis. *J. Exp. Biol.* 199, 83–91.

Dohnal, G. (1987). On estimating the diffusion coefficient. *J. Appl. Probab.*, 24, 105–114.

Dunn, J.E., and Brisbin, I.L. (1985). Characterization of the multivariate Ornstein-Uhlenbeck diffusion process in the context of home range analysis. *Statistical Theory and Data Analysis* (K. Matusita, *ed.*). Elsevier, Amsterdam, 181–205.

Dunn, J.E., and Gipson, P.S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, 33, 85–101.

Genon-Catalot, V., Laredo, C., and Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.* 19, 317–335.

Hadeler, K.P., de Mottoni, P., and Schumaker, K. (1980). Dynamic models for animal orientation. *J. Math. Biol.* 10, 307–332.

Heyde, C.C. (1994). A quasi-likelihood approach to estimating parameters in diffusion-type processes. *J. Appl. Probab.*, 31A, 283–290.

Kareiva, P.M., and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia (Berlin)*, 56, 234–238.

Karlin, S., and Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York.

Kendall, D.G. (1974). Pole-seeking Brownian motion and bird navigation. *J. Roy. Statist. Soc. Ser. B*, 36, 365–417.

Levin, S.A. (1986). Random walk models of movement and their implications. *Mathematical Ecology* (T.G. Hallam and S. Levin, *eds.*). Springer-Verlag, Berlin, 149–154.

Lohmann, K.J., and Lohmann, C.M.F. (1996). Orientation and open-sea navigation in sea turtles. *J. Exp. Biol.* 199, 73–81.

Malik, F., Brillinger, D., and Vale, R.D. (1994). High-resolution tracking of microtubule motility driven by a single kinesin motor. *Proc. Natl. Acad. Sci. U.S.A.* 91, 4584–4588.

McConnell, B.J., and Fedak, M.A. (1996). Movements in southern elephant seals. *Canad. J. Zool.* 74, 1485–1496.

McCulloch, C.E., and Cain, M.L. (1989). Analyzing discrete movement data as a correlated random walk. *Ecology*, 7, 383–388.

Moore, M. (1985). Modelling iceberg motion: A multiple time series approach. *Canad. J. Statist.* 13, 88–93.

Niwa, H. (1996). Newtonian dynamical approach to fish schooling. *J. Theoret. Biol.*, 181, 47–63.

Oksendal, B. (1995). *Stochastic Differential Equations*. Fourth Edition. Springer-Verlag, New York.

Okubo, A. (1980). *Diffusion and Ecological Problems: Mathematical Models*. Springer-Verlag, New York.

Perrin, M.F. (1928). Mouvement brownien de rotation. *Ann. École Norm. Sup.*, 45, 1–51.

Preisler, H.K., and Akers, R.P. (1995). Autoregressive-type models for the analysis of bark beetle tracks. *Biometrics*, 51, 259–267.

Stewart, B.S. (1996). Uncommon commuters. *Natur. Hist.*, 105, 58–63.

Stewart, B.S., and DeLong, R.L. (1995). Double migrations of the northern elephant seal. *J. Mammalogy* 76, 196–205.

Stewart, B.S., and Huber, H.R. (1993). *Mirounga angustirostris. Mammalian Species*, 449, 1–10.

Stewart, B.S., Leatherwood, S., Yochem, P. K., and Heide-Jorgensen, M.P. (1989). Satellite telemetry of locations and dive durations of a free-ranging harbor seal (*Phoca vitulina richardsi*) in the Southern California Bight. *Marine Mammal Sci*, 5, 361–375.

Stewart, B.S., Yochem, P.K., Huber, H.R., DeLong, R.L., Jameson, R.J., Sydeman, W., Allen, S.G., and LeBoeuf, B.J. (1994). History and present status of the northern elephant seal population. *Elephant Seals: Population Ecology, Behaviour and Physiology* (B.J. LeBoeuf and R.M. Laws, *eds.*), University of California Press, Los Angeles, 29–48.

Wagner, H. (1986). Flight performance and visual control of flight of the free flying housefly. *Philos. Trans. Roy. Soc. London Ser. B*, 312, 581–595.

Wehner, R., Michel, B., and Antonsen, P. (1996). Visual navigation in insects: Coupling of egocentric and geocentric information. *J. Exp. Biol.* 199, 129–140.

Wehrhahn, C., Poggio, T., and Bulthoff, H. (1982). Tracking and chasing in houseflies. *Biol. Cybernet.* 45, 123–130.

Weindler, P., Wiltschko, R., and Wiltschko, W. (1996). Magnetic information affects the stellar orientation of young birds. *Nature*, 383, 158–160.

Wiltschko, W., and Wiltschko, R. (1996). Magnetic orientation in birds. *J. Exp. Biol*, 199, 29–38.

Yosida, K. (1949). Brownian motion on the surface of the 3-sphere. *Ann. Math. Statis.*, 20, 292–296.

Zwiers, F.W. (1985). Estimating the probability of collision between an iceberg and a fixed marine structure. *Canad. J. Statist.*, 13, 94–105.

Received 11 December 1996
Revised 21 June 1997
Accepted 10 July 1997

Department of Statistics
University of California
Berkeley, California
U.S.A. 94720

email: brill@stat.berkeley.edu

Boalt Hall School of Law
University of California
Berkeley, California
U.S.A.

ENVIRONMETRICS, VOL. 8, 269–281 (1997)

RANDOM PROCESS METHODS AND ENVIRONMENTAL DATA: THE 1996 HUNTER LECTURE*

DAVID R. BRILLINGER†
Statistics Department, University of California, Berkeley, CA 94720, USA

SUMMARY

Random processes are basic to the study of environmental data, particularly data in time and space. This work presents three data analyses based on random process models: (a) a trend analysis, based on fitting a monotonic trend to river heights; (b) an analysis of point process data, with ordinal-valued marks, for damage assessment following an earthquake, and (c) an analysis of spatial-temporal meteorological data to estimate the speed of motion of a 500 mbar surface. There is discussion of stochastic processes generally. © 1997 by John Wiley & Sons, Ltd.

Environmetrics, **8**, 269–281 (1997)
No. of Figures: 8. No. of Tables: 1. No of References: 22.

KEY WORDS earthquake damage; floods; geopotential height; monotone trend; ordinal data; point process; random process; risk; spatial-temporal process; time series; velocity estimation

1. INTRODUCTION

Random processes are fundamental to studying data in time and space and such data are basic to environmental problems. These processes provide a major interface of the field of statistics with the worlds of science and technology. They are employed in practice to address questions arising in environmental science, questions such as is there: trend? change? association? causation? high risk? predictability? periodicity? structure?

In this paper three examples of random process data analysis are presented. The first analysis addresses the question of whether the mean level of the Rio Negro is increasing, perhaps as a consequence of deforestation of the Amazon Basin. Figure 1 presents a graph of the mean monthly river stage or level. The second analysis goes on to compute a smooth display of damage following the Loma Prieta earthquake and develops an estimate of damage risk as a function of distance from the earthquake source. This earthquake occurred near San Francisco. The basic data are interesting for being ordinal-valued. Some of them are graphed in Figure 2. The final example is a study of world weather as described by the height of the 500 mbar geopotential surface. One image is displayed in Figure 3.

The character of the paper, as was the Lecture, is expository. It was an honour to be asked to present the Hunter Lecture, particularly since I was privileged to know Stu Hunter at Princeton. He has done so much for the field he named *environmetrics*.

* Presented at the 7th meeting of the International Environmetric Society, 22–26 July 1996, Sao Paulo, Brazil.
† Correspondence to: D. R. Brillinger, Statistics Department, University of California, Berkeley, CA 94720, U.S.A.

Contract grant sponsor: National Science Foundation
Contract grant numbers: MCS-9300002, MCS-9625574, INT-9600251

CCC 1180–4009/97/040269–13$17·50

Received 16 February 1997
Revised 28 February 1997

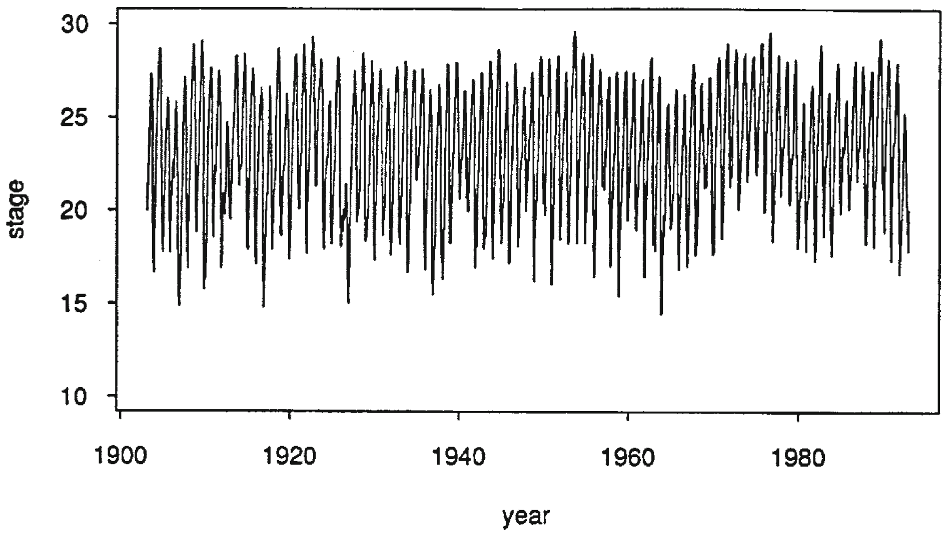


Figure 1. Monthly mean stages for the Rio Negro at Manaus, Brazil

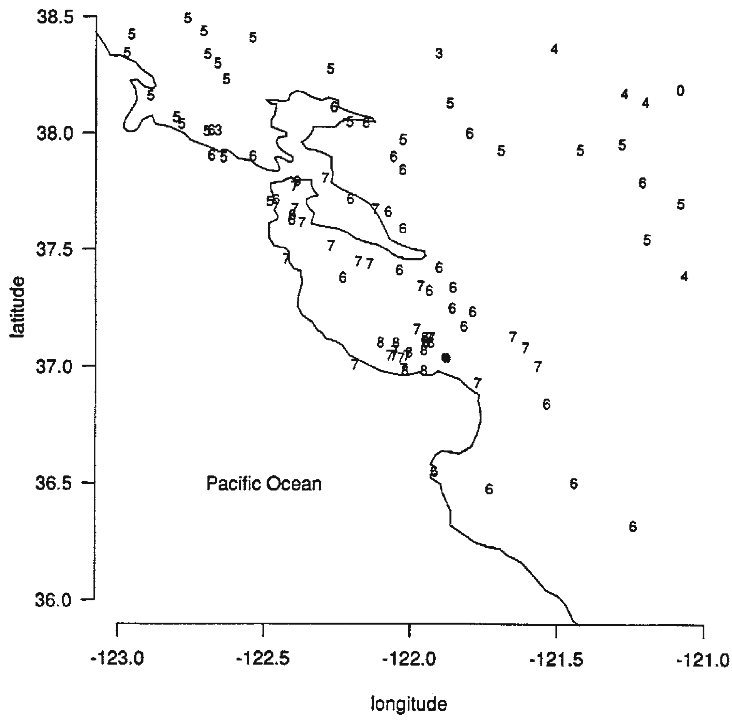


Figure 2. Modified Mercalli intensities observed for the Loma Prieta earthquake of 1989

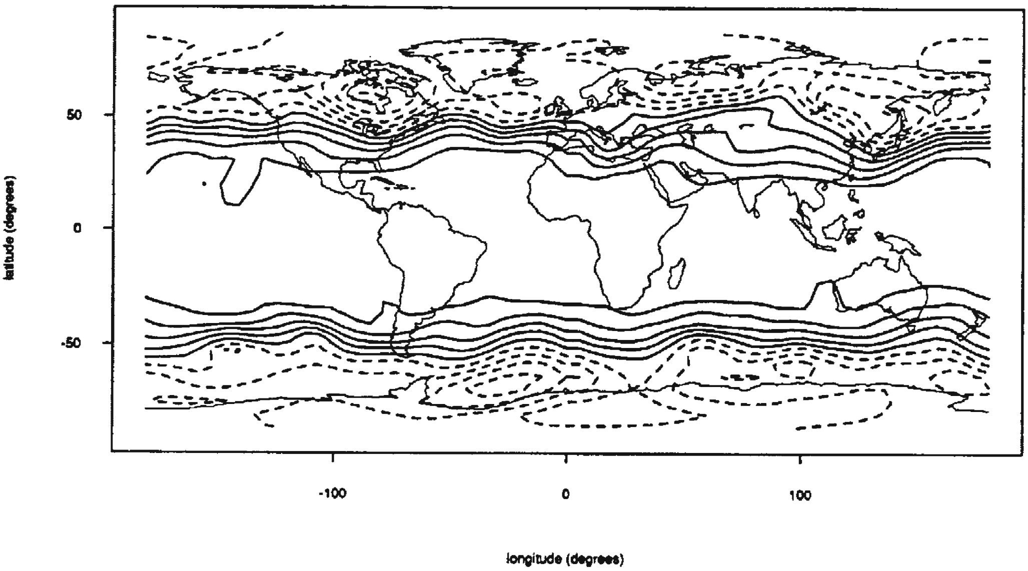


Figure 3. The height of the 500 mbar surface at 1200 GMT on 1 January 1986. The contour levels are spaced 100 m apart with 5300 m and below indicated by dashed lines

2. RANDOM PROCESSES

Random processes and accompanying statistical analyses benefit from formal development. Some of the basic concepts involved are next described.

By *data* are understood facts or measurements. *Statistics* may then be described as the science of using data wisely. A *process* is a collection of indexed values. For example, the index may be time or space or both. Particular cases include functions, curves, surfaces and changing surfaces. By *process data* are meant the measurements of (part) of a process. Some types of process data are: time series; point processes; images; spatial-temporal fields; marked point processes. A *system* is a collection of interacting processes. There may be a recognizable input and output. It is often useful to take a systems approach when working with process data.

In analysing data it can be convenient to introduce a probabilistic description. By *probability* is meant chance or long run frequency. A *stochastic* or *random process* is a probabilistic entity whose realizations are processes, in other words an indexed family of chance quantities or random variables. A random function is an example. It is usual to describe random processes by some of: joint densities; joint moments; stochastic difference and differential equations, or functional transforms. For example, the first two of these are given by:

$$\text{Prob}\{y_1 < Y(t_1) < y_1 + \mathrm{d}y_1, \ldots, y_k < Y(t_k) < y_k + \mathrm{d}y_k\} = f(y_1, \ldots, y_k)\, \mathrm{d}y_1 \ldots \mathrm{d}y_k$$

and

$$E\{Y(t_1) \ldots Y(t_k)\}$$

respectively, for a time series $\{Y(t)\}$. A *stochastic model* is a probabilistic description of a circumstance. In working with stochastic process data it is often convenient to develop a stochastic model. The examples of Sections 3.1, 3.2 and 3.3 involve 'signal plus noise' models.

Random process methodology allows a researcher to address problems involving spatial or temporal dependence. The process might be denoted as $Y(x, y, t)$ with (x, y) representing location (e.g. latitude and longitude) and t time. The analyses presented below involve a time series, $\{Y(t)\}$, a spatial marked point process, $\{(x_j, y_j), M_j\}$ and a time series of images, $\{Y(x, y, t)\}$, respectively. Using the Dirac delta function, the second process may be written in more familiar form as

$$Y(x, y) = \sum_j M_j \delta(x - x_j)\delta(y - y_j).$$

A broad variety of statistical methods are employed in the analysis of random process data: regression; smoothing; asymptotics; transforms; displays; likelihood.

3. SOME ANALYSES

3.1. Rio Negro water height

A gauge situated at the end of a pier in Manaus, Brazil, has been employed by the Manaus Harbour Limited to measure the daily height (or stage) of the Rio Negro river since 1903, see Sternberg (1987). The monthly mean levels for the period 1903–1992 are graphed in Figure 1. The values show a strong annual variation, of the order of magnitude of 10 metres. An environmental question, that these data might be useful in addressing, is whether the mean level of the river is rising as time is passing. This could be happening as a consequence of the deforestation.

The basic random process is a time series, $Y(t)$, with the index t referring to day. There are $T = 32{,}874$ daily values in all that were employed in the computations presented. A 'signal plus noise' model for the situation is

$$Y(t) = A(t) + S(t) + \varepsilon(t)$$

with $A(t)$ the annual component, $S(t)$ a monotonic non-decreasing trend component and $\varepsilon(t)$, $t = 0, \pm 1, \pm 2, \ldots$ a zero mean noise process. The mean level is

$$E\{Y(t)\} = A(t) + S(t).$$

The question of interest may now be written:

$$\text{Is } S(\cdot) \equiv 0?$$

The analysis proceeds by first estimating and then 'removing' the annual curve, $A(t)$. The estimate of $A(t)$ used is the median annual curve as described in Brillinger (1988). The series $Y(t) - \hat{A}(t)$ is graphed in Figure 4(a). No clear trend is apparent, but that issue will now be addressed. A monotonic estimate of $S(t)$ is computed, employing the 'pool the first violator and backaverage' algorithm of Friedman and Tibshirani (1984) based on the $Y(t) - \hat{A}(t)$ values. The resulting fitted monotonic trend, $\hat{S}(t)$, is graphed in Figure 4(b). Necessarily it shows a trend, but is it significant?

To address the question of whether $S(\cdot) \equiv 0$, the uncertainty level of the estimate, $\hat{S}(\cdot)$, is critical. Difficulties in determining such are that the estimate $\hat{S}(\cdot)$ is non-linear and that serial

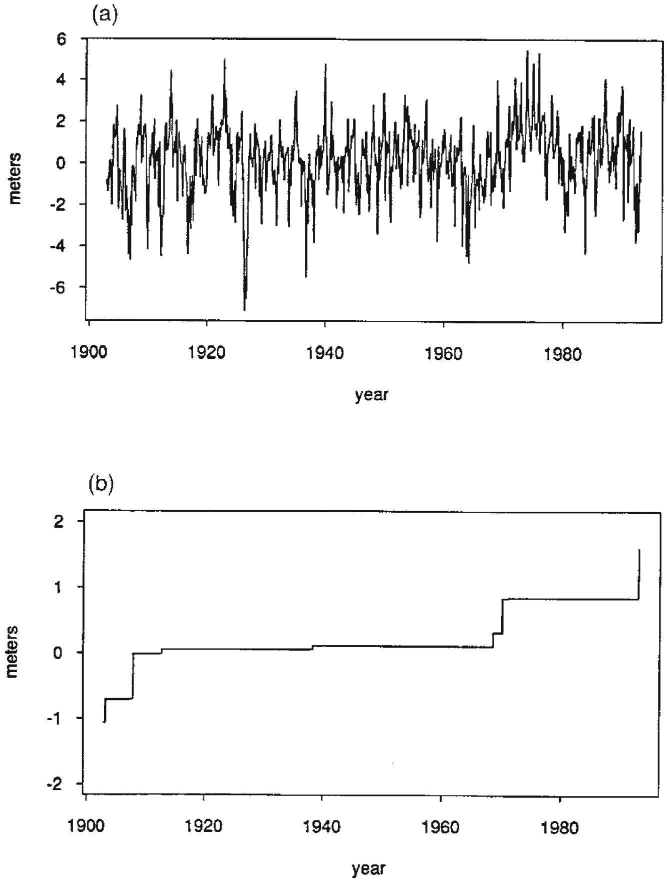


Figure 4. (*a*) The seasonally adjusted Rio Negro stages. (*b*) The fitted monotonic function $\hat{S}(t)$

correlation is likely to be present. To assess the uncertainty the following procedure was employed:

(i) Naive residuals, $\hat{\varepsilon}(t) = Y(t) - \hat{A}(t) - \hat{T}(t)$, were obtained. Here $\hat{T}(\cdot)$ is an 8 year running mean of the daily data, see Brillinger (1988).

(ii) These values were modelled as a (long) autoregressive process

$$\sum_{u=0}^{U} a(u)\varepsilon(t-u) = \eta(t), \quad a(0) = 1. \tag{1}$$

In the computations the value $U = 250$ was used. The estimated innovations $\hat{\eta}(t)$ of this fit were obtained.

(iii) The $\hat{\eta}(t)$ were permuted randomly and a reconstituted series, $\hat{\varepsilon}_P(t)$, formed recursively

$$\hat{\varepsilon}_P(t) = \hat{\eta}_P(t) - \sum_{u=1}^{U} \hat{a}(u)\hat{\varepsilon}_P(t-u).$$

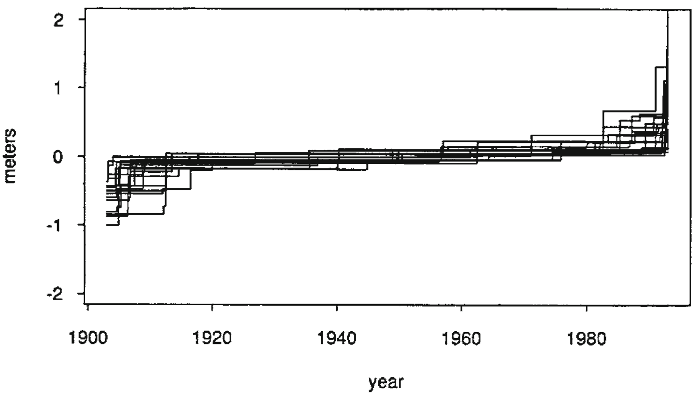


Figure 5. Nineteen fitted monotonic functions based on permuting the estimated innovations of the model (1)

(iv) Next an estimated monotonic mean level, $\hat{S}_P(t)$, based on the $\hat{\varepsilon}_P(t)$ was computed.
(v) This was done 19 times and the resultant $\hat{S}_P(t)$ graphed.

Figure 5 provides the 19 simulated estimates, $\hat{S}_P(t)$, on the same scale as Figure 4. One notices a higher level of variability in the early and the late time periods. Including the original estimate, $\hat{S}(t)$, there are 20 curves. One can ask whether the original curve is the most extreme of these and be working at the 1 in 20, i.e. 5 per cent level of significance. The conclusion, as in the analyses of Brillinger (1988, 1989), is that there is a soupçon of an increasing trend.

The logic of the resampling procedure is that the innovations, $\eta(t)$, of (1) are white noise and hence their distribution permutable. A naive justification of the procedure in the simplest case is provided in the Appendix. Research is in progress on the general case.

An alternative approach is suggested by the work of Bühlmann (1996). He develops a bootstrap procedure for estimates of smooth trend functions, via applying a bootstrapping to the residuals of fitted (long) autoregressive processes. His scheme involves resampling the estimated innovations, with replacement. In contrast the scheme in this paper involves randomly permuting the estimated innovations. Other estimates for the function $S(t)$ might be considered, e.g. monotone splines (Ramsay, 1988), or smoothed estimates (Mukerjee, 1988; Mammen, 1991).

Depending on the scientific question of interest, variants of the bootstrap might be employed to estimate the uncertainty of $\hat{S}(t)$ itself, rather than as here, where the variability is assessed in the case that no monotonic trend is present.

3.2. The Loma Prieta earthquake

After a sizeable earthquake observations are made of its effects on buildings and people. The observations are recorded on a descriptive scale, the scale of modified Mercalli (MM) intensities. The MM scale has 12 ordinal levels of increasing severity. For example the description of MM_{VIII} reads:

> *'Damage slight in specially designed structures; considerable in ordinary substantial buildings, with partial collapse; great in poorly built structures. Disturbs persons driving motor cars. Fall of chimneys, . . .'*

When such intensity data are examined, there is found to be a general fall-off in severity of effect with distance from the earthquake source. Figure 2 shows selected observations for the Loma Prieta event of 17 October 1989 and one can see that phenomenon. (Not all values were plotted there so as to reduce the effects of overstriking.) This event took place near Santa Cruz, California. The source of epicentre of the earthquake is marked in Figure 2 by a large dot. The event had magnitude 6·9, duration 10 seconds, and led to 63 deaths, 1300 buildings destroyed and 5·9 billion dollars damage. The largest MM intensity recorded was IX. There were 921 observations of MM intensity in all.

If (x_j, y_j) is the location of the jth measurement and M_j the MM intensity, then one can consider the data as a realization of a *spatial marked point process*, $\{(x_j, y_j), M_j\}$. A basic fact to be incorporated into the modelling of this circumstance is that the intensities are ordinal-valued. A convenient model leading to ordinal-valued data is the following. Let (x, y) denote location. Consider a latent variable

$$\zeta = g(x, y) + \varepsilon$$

where ε has an extreme value distribution. Suppose that the intensity, I, at that particular location is i if $a_{i-1} < \zeta \leqslant a_i$ for some cut values $\{a_i\}$. Then

$$\text{Prob}\{I = i\} = \text{Prob}\{a_{i-1} < \zeta \leqslant a_i\}$$

in this case leading to

$$\log(-\log(1 - \text{Prob}\{I \geqslant i \,|\, (x, y)\})) = \alpha_i + g(x, y) \tag{2}$$

for some constants α_i. Grouped continuous models were considered in McCullagh (1980) and McCullagh and Nelder (1989). The extreme value distribution leads to a generalized linear model with the complimentary log-log link. By conditioning one may act as if the successive cell values are independent, see Pregibon (1980), and fit via the usual GLM algorithms supposing the ζ_j to be independent of each other. The spatial dependence in this case is introduced via the smoothness of the function $g(\cdot)$. The motivation for the state variable ζ and the extreme value distribution is that ζ represents the strength of the earthquake effect at (x, y). The extreme value distribution is employed because the intensity recorded is the maximum observed at a site.

The model may be fit by a locally weighted analysis. To estimate $g(x, y)$ and the α_i one can form a log-likelihood with the terms inversely weighted to how far this location, (x_j, y_j) is from (x, y), see Cleveland *et al.* (1992) and Brillinger (1994b). Specifically in the calculations presented the procedure gam( ) with the loess( ) smoother of S-plus was employed, see Hastie (1992). Figure 6 provides the contours of the estimate $\hat{g}(x, y)$. As anticipated one sees a fall off of effect with distance from source.

Once the model has been fit, risk probabilities and expected losses may be estimated. Further analysis suggests that the fit may be reasonably approximated by the functional form

$$\log(-\log(1 - \text{Prob}\{I \geqslant i\})) = \alpha_j + \beta d + \gamma \log(d)$$

where d is the distance from a site to the earthquake source. Figure 7 gives the fit of this relationship for the cases $i = 0$, V, VIII. (0 corresponds to no earthquake effect noted.)

Using assumed loss ratio values for buildings of some type of interest, one may now estimate the expected loss for such a building situated a given distance from an earthquake source.

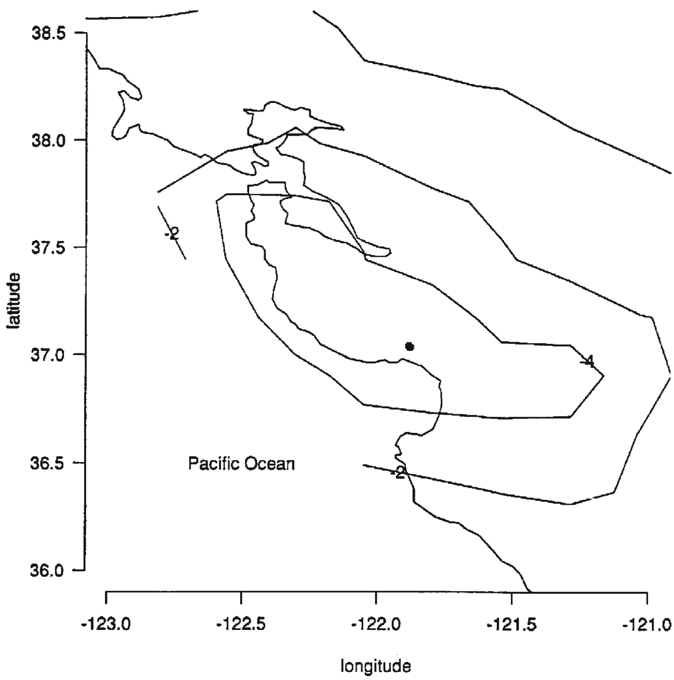


Figure 6. The Loma Prieta event estimated surface; $g(x, y)$ of (2)

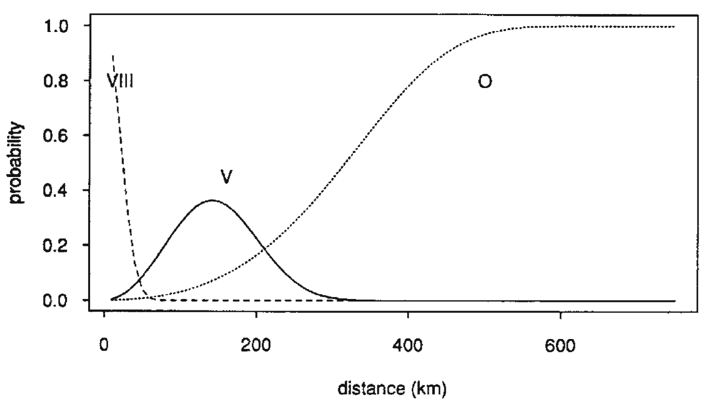


Figure 7. The estimated risk of MM intensity, I, as a function of distance from the source of the earthquake; Loma Prieta prob$\{I = \mathrm{i} \mid \text{source distance}\}$

Brillinger (1993, 1994b) estimated isoseismals for these data previously, but the ordinal character of the data was not taken specific account of.

3.3. The 500 mbar geopotential height

The weather moves from west to east and it can be of interest to estimate the velocity of that motion based on a sequence of images. The data now studied are a five day sequence of 0000 and

1200 Greenwich Mean Time (GMT) geopotential analyses. These are spatially interpolated estimates of the height of the 500 millibar (mbar) pressure field across the surface of the earth. This quantity provides the thickness of the atmosphere between the sea level and the 500 mbar level. It relates to temperature, being low for cold values and high for warm values. The period covered is 1200 GMT 1 January 1986 to 0000 GMT 6 January 1986. The time interval between images is 12 hours and there are 10 time slices. The measurements of 1200 GMT 1 January are graphed as contours in Figure 3. Values 5300 metres and below are indicated by dashed lines. The contours are 100 m apart. One sees, for example, a depression over Hudson Bay in Northern Canada. Further examination of the 10 such images shows the depression to move eastward and fill in over the eastern Atlantic on 5 January.

The problem of concern is how to estimate the velocity of a moving phenomenon, such as this 500 mbar field. Denote the data by $Y(x, y, t)$. One has a spatial-temporal process with index (x, y, t). Suppose one restricts attention to motion along a single latitude. Denote the values along a given latitude, y, by $Y(x, t)$ with t referring to time and x to longitude east. Being on the sphere, this process is periodic in x.

Consider a signal plus noise model with

$$E\{Y(x, t)\} = g(x) + h(x - vt) \tag{3}$$

where x is longitude, v is velocity and $g(\cdot), h(\cdot)$ are smooth and periodic. The function $g(\cdot)$ corresponds to stable features in the field, while $h(\cdot)$ refers to dynamic ones. The problem is to estimate the velocity, v. Because the principal parameter, v, is real-valued, what was done here was the following. First the $g(\cdot)$ term in (3) was 'eliminated' by taking first differences, i.e. the series

$$Y'(x, t) = Y(x + 1, t) - Y(x, t)$$

was studied. A corresponding model is now

$$Y'(x, t) = h'(x - vt) + \varepsilon(t) \tag{4}$$

with $\varepsilon(t)$ representing noise. For given v and smooth periodic $h'(\cdot)$ there are a variety of ways to estimate $h'(t)$, see Hastie and Tibshirani (1990). In the present case the function loess( ), see Cleveland *et al.* (1992), is employed to estimate $h'(\cdot)$. For each v let

$$R(v)^2 = 1 - \Sigma(Y'(x, t) - \hat{h}'(x - vt))^2 / \Sigma(Y'(x, t) - \bar{Y})^2. \tag{5}$$

Figure 8 graphs this function for six different latitudes. In each case one sees a hump. The location of its peak may be taken to be the estimate of v. The estimated velocities, in units of degrees per day, are given in Table I. The bracketed figures are the estimated standard errors.

Table I.

Latitude	Velocity
36·0	13·9 (5·0)
41·4	17·3 (7·4)
47·1	18·7 (14·0)
52·6	22·8 (16·6)
58·1	19·4 (10·4)
63·7	18·8 (10·6)

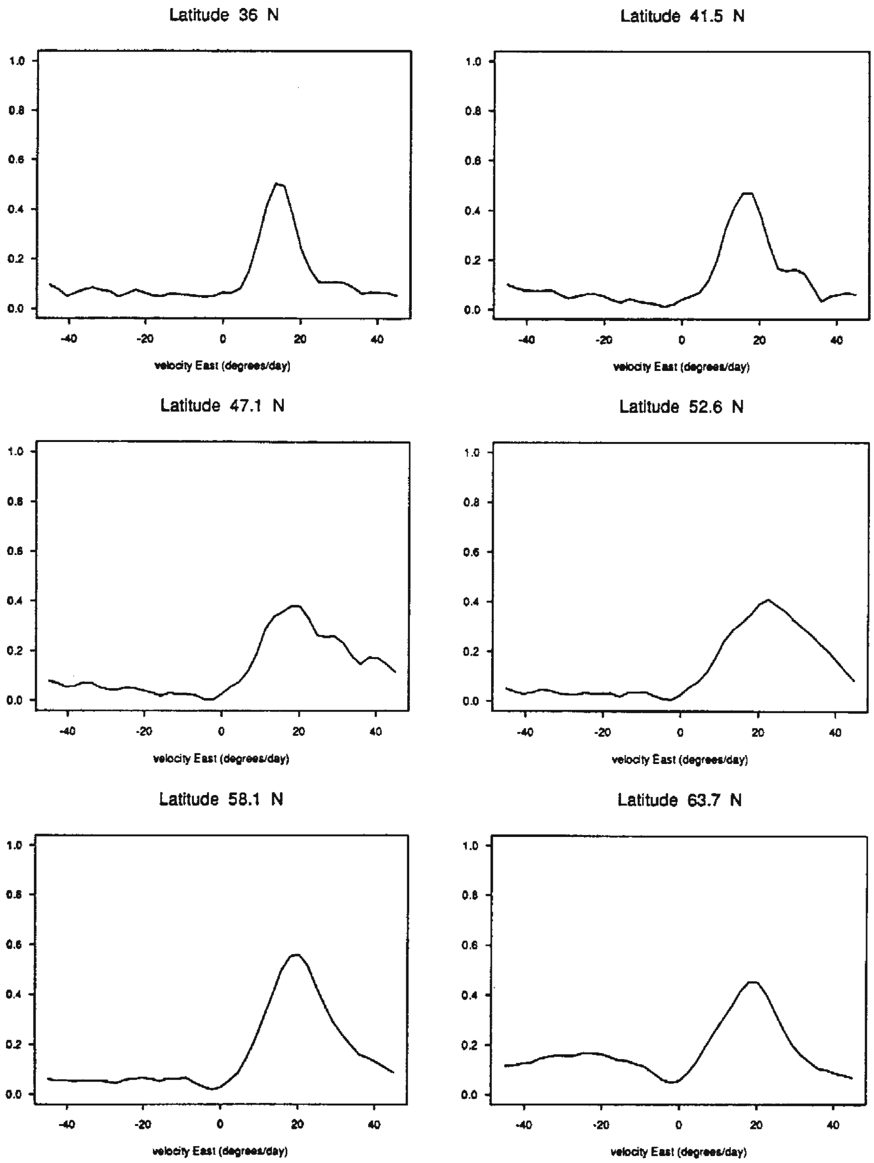


Figure 8. The function $R(v)^2$ of (5) as a function of velocity, v

These last are determined from the profile likelihood in the manner of Richards (1961). The idea of that paper is that for given v the other unknowns of the problems may be estimated and substituted for, leaving v as the sole unknown. To obtain the entries of Table I, it has been assumed that the $\varepsilon(t)$ are independent normals with mean 0 and common variance. Richards shows its variance may then be estimated ignoring the presence of the other unknowns. A general conclusion is that the surface, around the latitude of 50 degrees north, is moving eastward with a velocity of about 20 degrees/day.

More details on this example may be found in Brillinger (1997). The estimate and preliminary computations with it were presented in the 1989 Wilks Lectures at Princeton University.

4. VARIANTS AND PROBLEMS

In practice various difficulties arise. The data to be studied may be irregular, e.g. the values may have been aggregated, there may be bias in the collection procedures, there may be missing values, or censoring, or measurement noise. It is of interest to develop procedures for dealing with these possibilities.

There are variants of the techniques presented that may be considered including: non-linear relationships; random effects; replicated responses; data collected in an experimental design; wavelet versions. It is of interest to develop techniques for dealing with these.

Spatial-temporal data arise for processes of more general character. In particular one can consider graph-, tesselation- and colour-valued processes. The process of the earthquake data was ordinal-valued. A process might be proportion-, count-, categorical- or non-negative-valued. The time series proportion case is considered in Grunwald *et al.* (1993).

5. DISCUSSION AND SUMMARY

Questions of trend existence, risk assessment and motion estimation have been addressed. A defining aspect for the problems was the dependence of the data on time and space. Random process concepts and techniques were found to be a powerful way to conceptualize and address these situations.

The approach to the problems studied has been that of building, fitting and manipulating stochastic models for curves, surfaces and point masses. Dependence has been introduced into the models by including signals (here $S(t)$, $g(x, y)$, $h(x - vt)$).

Looking towards the future one can say that it has never been easier to work with large complex data sets, one has such an array of computing devices, display devices, storage devices and analytic tools to employ.

A basic general remark is that in problems such as these the importance of collaboration and learning the pertinent subject matter cannot be overemphasized.

APPENDIX

The resampling procedure for assessing uncertainty described in Section 3.1 is here motivated for one simple case.

Consider the naive trend model

$$Y(t) = \alpha + \beta t + \varepsilon(t), \quad t = 0, \ldots, T-1$$

with $\varepsilon(\cdot)$ noise. The least squares estimate of β is

$$\begin{aligned} \hat{\beta} &= \sum_t Y(t)(t - \bar{t}) / \sum_t (t - \bar{t})^2 \\ &= \beta + \sum_t \varepsilon(t)(t - \bar{t}) / \sum_t (t - \bar{t})^2. \end{aligned}$$

One is interested in the permutation distribution of $\hat{\beta}$.

Let $\{\varepsilon_P(0), \ldots, \varepsilon_P(T-1)\}$ denote a random permutation of $\{\varepsilon(0), \ldots, \varepsilon(T-1)\}$. One has

$$E_P\left\{\sum_t \varepsilon_P(t)(t-\bar{t})\right\} = \left(\frac{1}{T}\sum_t \varepsilon(t)\right)\sum_t (t-\bar{t}) = 0$$

and after some computations

$$\underset{P}{\operatorname{var}}\left\{\sum_t \varepsilon_P(t)(t-\bar{t})\right\} = \Sigma(\varepsilon(t)-\bar{\varepsilon})^2\Sigma(t-\bar{t})^2/(T-1)$$

and so

$$\underset{P}{\operatorname{var}}\ \hat{\beta} = s_\varepsilon^2/\Sigma(t-\bar{t})^2.$$

One reference for the computations is Kendall and Stuart (1961), Section 31.19. In the case that the $\varepsilon(t)$ are assumed i.i.d. with variance σ^2

$$\operatorname{var}\ \hat{\beta} = \sigma^2/\Sigma(t-\bar{t})^2.$$

It is seen that this last may be estimated by an estimate of

$$\underset{P}{\operatorname{var}}\ \hat{\beta}$$

which may in turn be estimated by evaluating the estimate for various randomly selected permutations. The asymptotic normality of the estimate under the permutation distribution follows from the results of Ho and Chen (1978).

ACKNOWLEDGEMENTS

I wish to thank Professor Hilgard O'Reilly Sternberg for providing the basic data collected at Manaus, Francis Zwiers of Environment Canada who provided the geopotential data and Lindy Brewer of the US Geological Survey for providing the Loma Prieta data. The work was carried out with the support of the National Science Foundation Grants MCS-9300002, MCS-9625574 and INT-9600251. It was also supported by the University of California Campus-Laboratory Collaboration project 'Advanced Earthquake Hazard Research'.

REFERENCES

Brillinger, D. R. (1988). 'An elementary trend analysis of Rio Negro levels at Manaus, 1903–1985', *Revista Brasileira Prob. Est.*, **2**, 63–79.

Brillinger, D. R. (1989). 'Consistent detection of a monotonic trend superposed on a stationary time series', *Biometrika*, **76**, 23–30.

Brillinger, D. R. (1993). 'Earthquake risk and insurance', *Environmetrics*, **4**, 1–21.

Brillinger, D. R. (1994a). 'Trend analysis: time series and point process problems', *Environmetrics*, **5**, 11–19.

Brillinger, D. R. (1994b). 'Examples of scientific problems and data analyses in demography, neurophysiology and seismology', *Journal of Computational and Graphical Statistics*, **3**, 1–22.

Brillinger, D. R. (1997). 'An Application of Statistics to Meteorology: Estimation of Motion', *in* Pollard, D., Torgersen, E. and Yang, G. L. (eds), *Festschrift for Lucien LeCam*, Springer, New York, pp. 93–105.

Bühlmann, P. (1996). 'Confidence regions for trends in time series: a simultaneous approach with a sieve bootstrap'. Preprint.

Cleveland, W. S., Grosse, E. and Shyu, W. M. (1992). 'Local regression models', *in* Chambers, J. M. and Hastie, T. J. (eds), *Statistical Models in S*, Wadsworth, Pacific Grove, pp. 309–376.

Friedman, J. and Tibshirani, R. (1984). 'The monotone smoothing of scatterplots', *Technometrics*, **26**, 243–250.

Grunwald, G. K., Raftery, A. E. and Guttorp, P. (1993). 'Time series of continuous proportions', *Journal of the Royal Statistical Society, Series B*, **55**, 103–116.

Hastie, T. J. (1992). 'Generalized additive models', *in* Chambers, J. M. and Hastie, T. J. (eds), *Statistical Models in S*, Wadsworth, Pacific Grove, pp. 249–308.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, New York.

Ho, S.-T. and Chen, L. H. Y. (1978). 'An Lp bound for the remainder in a combinatorial central limit theorem', *Annals of Probability*, **6**, 231–249.

Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics, Vol. 2*, Griffin, London.

Mammen, E. (1991). 'Nonparametric regression under qualitative smoothness assumptions', *Annals of Statistics*, **19**, 741–759.

McCullagh, P. (1980). 'Regression models for ordinal data', *Journal of the Royal Statistical Society, Series B*, **42**, 109–127.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn, Chapman and Hall, New York.

Mukerjee, H. (1988). 'Monotone nonparametric regression', *Annals of Statistics*, **16**, 741–750.

Pregibon, D. (1980). 'Discussion of paper by P. McCullagh', *Journal of the Royal Statistical Society, Series B*, **42**, 139.

Ramsay, J. O. (1988). 'Monotone regression splines in action', *Statistical Science*, **3**, 425–441.

Richards, F. S. G. (1961). 'A method of maximum-likelihood estimation', *Journal of the Royal Statistical Society, Series B*, **23**, 469–475.

Sternberg, H. O'R. (1987). 'Aggravation of floods in the Amazon River as a consequence of deforestation?', *Geografiska Annaler*, **69A**, 201–219.

Statistical Science
2008, Vol. 23, No. 1, 48–64
DOI: 10.1214/07-STS246

The 2005 Neyman Lecture: Dynamic Indeterminism in Science[1]

David R. Brillinger

Abstract. Jerzy Neyman's life history and some of his contributions to applied statistics are reviewed. In a 1960 article he wrote: "*Currently in the period of dynamic indeterminism in science, there is hardly a serious piece of research which, if treated realistically, does not involve operations on stochastic processes. The time has arrived for the theory of stochastic processes to become an item of usual equipment of every applied statistician.*" The emphasis in this article is on stochastic processes and on stochastic process data analysis. A number of data sets and corresponding substantive questions are addressed. The data sets concern sardine depletion, blowfly dynamics, weather modification, elk movement and seal journeying. Three of the examples are from Neyman's work and four from the author's joint work with collaborators.

Key words and phrases: Animal motion, ATV motion, elk, Jerzy Neyman, lifetable, monk seal, population dynamics, sardines, stochastic differential equations, sheep blowflies, simulation, synthetic data, time series, weather modification.

1. INTRODUCTION

This paper is meant to be a tribute to Jerzy Neyman's substantive work with data sets. There is an emphasis on scientific questions, statistical modeling and inference for stochastic processes.

The title of this work comes from Neyman (1960) where one finds,

> "The essence of dynamic indeterminism in science consists in an effort to invent a hypothetical chance mechanism, called a 'stochastic model,' operating on various clearly defined hypothetical entities, such that the resulting frequencies of various possible outcomes correspond approximately to those actually observed."

Here and elsewhere Neyman appeared to use the adjective "indeterministic" where others would use "stochastic," "statistical" or "nondeterministic"; see, for example, Neyman and Scott (1959). Perhaps Neyman had some deeper or historical context in mind, but that is not clear. In this paper the emphasis is on the word "dynamic."

Jerzy Neyman (JN) led a full life. Reid (1998) contains many details and anecdotes, a lot of them in Neyman's own words. Other sources include the papers: Neyman (1970), Le Cam and Lehmann (1974), Kendall, Bartlett and Page (1982), Scott (1985), Lehmann (1994) and Le Cam (1995).

The article has six sections: 1. Introduction, 2. Jerzy Neyman, 3. Some formal methods, 4. Three examples of JN's applied statistics work, 5. Four examples of random process data analysis, 6. Conclusion. The focus is on applied work in the environmental sciences and phenomena. This last is a word that Neyman often employed.

In particular the examples show how random process modeling can prove both helpful and not all that difficult to implement. The thought driving this paper is that by examining a number of examples, unifying methods and principles may become apparent. One connecting thread is "synthetic" data, in the language of Neyman, Scott and Shane (1953) and Neyman and Scott (1956). Synthetic data, based on simulations, are

David Brillinger is Professor, Department of Statistics, University of California, Berkeley, California 94720, USA (e-mail: brill@stat.berkeley.edu).

[1]Discussed in 10.1214/07-STS246B and 10.1214/07-STS246A; rejoinder at 10.1214/07-STS246REJ.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_25, © Springer Science+Business Media, LLC 2012

an exploratory tool for model validation that has the advantage of suggesting how to create another model if the resemblance of the simulation to the actual data is not good.

There are quotes throughout to create a flavor of JN's statistical approaches.

2. JERZY NEYMAN

> "His devotion to Poland and its culture and traditions was very marked, and when his influence on statistics and statisticians had become worldwide it was fashionable ... to say that 'we have all learned to speak statistics with a Polish accent' ..." (Kendall, Bartlett and Page, 1982).

The life of Neyman is well documented by JN and others; see, for example, Reid (1998), LeCam and Lehmann (1974) and Scott (1985). Other sources are cited later. Neyman was of Polish ancestry and as the above quote makes clear he was very Polish! Table 1 records some of the basic events of his life. One sees a flow from Poland to London to Berkeley with many sidetrips intermingled throughout his life. These details are from Scott (1985) and Reid (1998).

Neyman's education involved a lot of formal mathematics (integration, analysis, ...) and probability. He often mentioned the book, *The Grammar of Science* (Pearson, 1900) as having been very important for his scientific and statistical work. He described Lebesgue's *Leçons sur l'intégration* as "the most beautiful monograph that I ever read."

TABLE 1
A timeline of Jerzy Neyman's life

Date	Event
1894	Born, Bendery, Monrovia
1916	Candidate in Mathematics, University of Kharkov
1917–1921	Lecturer, Institute of Technology, Kharkov
1921–1923	Statistician, Agricultural Research Institute, Bydgoszcz, Poland
1923	Ph.D. in Mathematics, University of Warsaw
1923–1934	Lecturer, University of Warsaw Head, Biometric Laboratory, Nencki Institute
1934–1938	Lecturer, then Reader, University College, London
1938–1961	Professor, University of California, Berkeley
1955	Berkeley Statistics Department formed
1961–1981	Professor Emeritus, University of California, Berkeley
1981	Died, Oakland, California

The Author's Note to the *Early Statistical Papers* (Neyman, 1967) comments on the famous and influential teachers he had at Kharkov. They included S. Bernstein ("my teacher in probability"), C. K. Russyan, and A. Przeborski. Others he mentions as influential include E. Borel, R. von Mises, A. N. Kolmogorov, E. S. Pearson and R. A. Fisher.

Neyman came to Berkeley in 1938. That appointment had been preceded by a triumphant U.S. tour in 1937. The book Neyman (1938b) resulted from the tour. After Neyman's arrival, internationally renowned probabilists and statisticians began to visit Berkeley regularly and contributed much to its research atmosphere and work ethic.

In Neyman's time the lunch room used to play an important role in the Berkeley Department. JN, Betty Scott (ELS) and Lucien Le Cam enthralled students, colleagues, visitors and the like with their conversation. They involved everyone in the stories and discussions.

Neyman had a seminar Wednesday afternoons. It began with coffee and cakes. Then there was a talk, often by a substantive scientist, but theoretical talks did occur from time to time. The talk's discussion was followed by drinks at the Faculty Club including the famous Neyman toasts. "To the speaker. To the international intellectual community. To the ladies present and some ladies absent." Up until perhaps the mid-1970s there was a dinner to end the event.

Neyman's work ethic was very strong. It typically included Saturdays in the Department, and for those who came to work also there were cakes at 3 pm.

3. SOME FORMAL METHODS

> "Every attempt at a mathematical treatment of phenomena must begin by building a simplified mathematical model of the phenomena." (Neyman, 1947).

This section provides a few of the technical ideas and methods that are basic to the examples presented. The examples involve dynamics, time, spatial movement, Markov processes, state-space models, stochastic differential equations (SDEs) and phenomena.

3.1 Random Process Methods

> "..., modern science and technology provide statistical problems with observable random variables taking their values in functional spaces." (Neyman, 1966).

By a random process is meant a random function. Their importance was already referred to in Section 1.

In particular Neyman was concerned with "phenomena developing in time and space" (Neyman, 1960). The random processes describing these are the backbone of much of modern science.

3.2 Markov Processes

Neyman was taken with Markov processes. Reid (1998) quotes him as saying,

> "So what Markov did—he considered changes from one position to another position. A simple example. You consider a particle. It's maybe human. And it can be in any number of states. And this set of states may be finite, may be infinite. Now when it's Markov—Markov is when the probability of going—let's say—between today and tomorrow, whatever, depends only on where you are today. That's Markovian. If it depends on something that happened yesterday, or before yesterday, that is a generalization of Markovian."

Time and Markovs play key roles in Fix and Neyman (1951). An advantage of working with a Markov process is that when there is a parameter one can set down a likelihood function directly.

3.3 Stochastic Differential Equations (SDEs)

> "It seems to me that the proper way of approaching economic problems mathematically is by equations of the above type, in finite or infinitesimal differences, with coefficients that are not constants, but random variables; or what is called random or stochastic equations. ... The theory of random differential and other equations, and the theory of random curves, are just starting." (Neyman, 1938a).

To give an example, let $\mathbf{r}(t)$ refer to the location of a particle at time t in R^p space. The path that it maps out as t increases is called the trajectory. (Trajectory is an old word used for a stochastic process.) Its vector-valued velocity will be denoted

$$\boldsymbol{\mu}(t) = d\mathbf{r}(t)/dt.$$

Rewriting this equation in terms of increments and adding a random disturbance leads to a so-called stochastic differential equation

$$(1) \qquad d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t), t)\,dt + \boldsymbol{\sigma}(\mathbf{r}(t), t)\,d\mathbf{B}(t)$$

or in integrated form,

$$(2) \quad \mathbf{r}(t) = \mathbf{r}(0) + \int_0^t \boldsymbol{\mu}(\mathbf{r}(s), s)\,ds + \int_0^t \boldsymbol{\sigma}(\mathbf{r}, s)\,d\mathbf{B}(s).$$

If, for example, the process $\mathbf{B}$ is Brownian, that is, the increments $\mathbf{B}(t_{i+1}) - \mathbf{B}(t_i)$ are $IN(\mathbf{O}, (t_{i+1} - t_i)\mathbf{I})$, then, under conditions on $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, a solution of the equation exists and is a Markov process. The function $\boldsymbol{\mu}$ is called the drift rate and $\boldsymbol{\sigma}$ the diffusion coefficient.

A particular case of an SDE is the Ornstein–Uhlenbeck process given by

$$d\mathbf{r}(t) = \alpha\big(\mathbf{a} - \mathbf{r(t)}\big)\,dt + \sigma\,d\mathbf{B}(t)$$

with $\alpha > 0$ and σ a scalar. This models a particle being attracted to the point $\mathbf{a}$ with the motion disturbed randomly.

An approximate solution to (1) is given, recursively, by

$$(3) \quad \begin{aligned} \mathbf{r}(t_{i+1}) - \mathbf{r}(t_i) &\approx \boldsymbol{\mu}(\mathbf{r}(t_i), t_i)(t_{i+1} - t_i) \\ &\quad + \boldsymbol{\sigma}(\mathbf{r}(t_i), t_i)\mathbf{Z}_i\sqrt{t_{i+1} - t_i} \end{aligned}$$

with the t_i an increasing sequence of time points filling in the time domain of the problem; see Kloeden and Platen (1995). The $\mathbf{Z}_i$ are independent p-variate standard normals. This solution procedure to (1) is known as the Euler method. In fact Itô (1951) used an expression like (3) to demonstrate that, under conditions, (1) had a unique solution.

There has been a substantial amount of work on statistical inference for SDEs; references include Heyde (1994) and Sørensen (1997). There are parametric and nonparametric fitting methods. Inferential work may be motivated by setting down the above approximation and taking the t_i to be the times of observation of the process.

Assuming that $\boldsymbol{\mu}(\mathbf{r}, t) = \boldsymbol{\mu}(\mathbf{r})$, that $\boldsymbol{\sigma}(\mathbf{r}(t), t) = \sigma\mathbf{I}$, σ scalar, and that $\mathbf{r}$ is p vector-valued, one can consider as an estimate of σ^2

$$(4) \quad \begin{aligned} \hat{\sigma}^2 = \frac{1}{pI}\sum_i \big\| \mathbf{r}(t_{i+1}) - \mathbf{r}(t_i) - \hat{\boldsymbol{\mu}}(\mathbf{r}(t_i)) \\ \cdot (t_{i+1} - t_i)\big\|^2 / (t_{i+1} - t_i), \end{aligned}$$

$i = 1, \ldots, I$, having determined an estimate of $\boldsymbol{\mu}$.

If the region of motion, say D, is bounded with boundary ∂D, one can proceed via the SDE

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t), t)\,dt + \boldsymbol{\sigma}(\mathbf{r}(t), t)\,d\mathbf{B}(t) + d\mathbf{A}(t)$$

with the support of $\mathbf{A}$ on the boundary ∂D. This construction pushes the particle into D.

3.4 A Potential Function Approach

The choice of the function $\boldsymbol{\mu}$ in (1) may be motivated by Newtonian dynamics. Suppose there is a scalar-valued potential function, $H(\mathbf{r}(t), t)$; see Taylor (2005). Such a function H can control a particle's direction and velocity.

In a particular physical situation the Newtonian equations of motion may take the form

$$d\mathbf{r}(t) = \mathbf{v}(t)\,dt,$$

$$d\mathbf{v}(t) = -\beta\mathbf{v}(t)\,dt - \beta\nabla H(\mathbf{r}(t), t)\,dt, \quad (5)$$

with $\mathbf{r}(t)$ the particle's location at time t, $\mathbf{v}(t)$ the particle's velocity and $-\beta\nabla H$ the external force field acting on the particle. The parameter β represents the coefficient of friction. Here $\nabla = (\partial/\partial x, \partial/\partial y)^\tau$ is the gradient operator. For example, Nelson (1967) makes use of the form (5).

In the case that the relaxation time, β^{-1}, is small (or in other words, the friction is high), (5) is approximately

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r}(t), t)\,dt = \boldsymbol{\mu}(\mathbf{r}, t)\,dt.$$

Writing the velocity $\mathbf{v}(t) = \boldsymbol{\mu}(\mathbf{r}, t)$ one is led to a stochastic gradient system

$$d\mathbf{r}(t) = -\nabla H(\mathbf{r}(t), t)\,dt + \sigma\,d\mathbf{B}(t).$$

The function H might be a linear combination of elementary known functions, a combination of thin plate splines placed around a regular grid or based on a kernel function. Example 7 below will indicate the method. The method is further elaborated in Brillinger (2007a, 2007b).

4. THREE EXAMPLES OF JN'S APPLIED STATISTICS WORK

> "...the delight I experience in trying to fathom the chance mechanisms of phenomena in the empirical world." (Neyman, 1970).

Neyman was both an exceptional mathematical statistician and an exceptional applied statistician. The applied work commenced right at the beginning of his career and continued until the very end. This section presents examples from astronomy, fisheries and weather modification. These examples were chosen as they are interesting and they blend into the later examples in the paper.

Neyman's work was special in applied statistics in that he set down specific "postulates" or assumptions. Tools of his applied work included sampling, best asympotically normal (BAN) estimators, C(α) tests, chi-squared, randomization and synthetic data. His work was further characterized by the very careful preparation of the data by his Statistical Laboratory workers.

JN's applied papers typically include substantial introductions to the scientific field of concern. Topics include farfield effects of cloud seeding, estimation of the dispersion of the redshift of galaxies, higher-order clustering of galaxies, and sardine depletion.

Given Neyman's concern with the scientific method, one can wonder how he validated or appraised his models. On reading his papers, hypothesis testing seems to include assessment. There were lots of data, and fit components (observed–expected) and chi-squared (residuals). There was smooth chi-squared to get alternative hypotheses. There was often the remark, "appears reasonable."

4.1 Example 1. ASTRONOMY

> "By far the strongest and most sustained effort expended for us in studying natural phenomena through appropriately selected aspects of the process of clustering referred to astronomy, specifically to galaxies. ..., the stimulus came from the substantive scientists, that is from astronomers." (Neyman and Scott, 1972).

The work of Neyman, and his collaborators in this case, is a model for applied statistics. The question is made clear. Substantive science is involved. Statistical theory is employed and developed as necessary. Empirical analyses are carried out.

In a series of papers Neyman, Scott, Shane and Swanson addressed the issue of galaxy clustering. They applied mathematical models to the Lick galaxy counts of Shane and Wirtanen. They were the first to compare the observed galaxy distribution to synthetic images of the Universe. They assumed that clusters occur around centers distributed as a spatial Poisson process. Each center was assigned a random number of galaxies and the latter placed independently at random distances from the center. This model, the so-called Neyman–Scott model, seemed to fit reasonably. However, when Neyman and Scott produced a simulated realization, or synthetic plate, of the sky from their model they were surprised. The actual pictures of the sky were a lot more lumpy than those their simulation had produced.

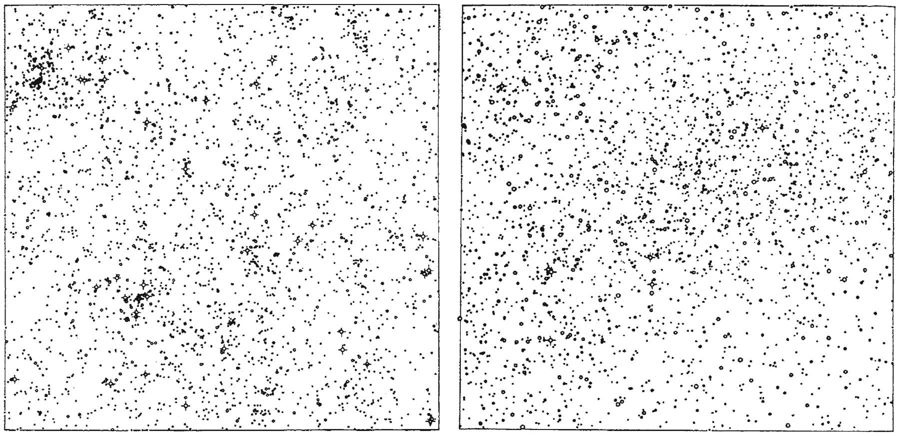


FIG. 1. *Left-hand panel is an image of an actual photographic plate. The right-hand panel is a synthetic plate. See Scott, Shane and Wirtanen* (1954).

> "When the calculated scheme of distribution was compared with the actual distribution of galaxies ..., it became apparent that the simple mechanism postulated could not produce a distribution resembling the one we see" (Neyman and Scott, 1956).

More clustering was needed in the model. Neyman and Scott proceeded to introduce it. With a two-stage clustering process the simulated appearance of the sky looked much more realistic. Figure 1, taken from Scott, Shane and Wirtanen (1954), presents an example.

In summary,

> "... it was shown that the visual appearance of a 'synthetic' photographic plate, obtained by means of a large-scale sampling experiment, conforming exactly with the assumptions of the theory, is very similar to that of an actual plate" (Neyman, Scott and Shane 1954).

4.2 Example 2. SARDINE DEPLETION

> "Biometry is an interdisciplinary domain aimed at the understanding of biological phenomena in terms of chance mechanisms." (Neyman, 1976).

In 1947–1948 Neyman was called upon by the California Council of the Congress of Industrial Organizations to study the decrease in sardine catches. The decrease was of great concern and strongly affected the canneries and commerce of the workers along the west coast of the United States.

In particular JN was consulted regarding the natural and fishing mortality of the sardines. A specific purpose of his work was "... to study the methods of estimating the death rates of the sardines." JN wrote three reports on sardine fishery. They are collected in Neyman (1948) and titled, 1. *Evaluations and Observations of Material and Data Available on the Sardine Fishery*, 2. *Natural and Fishing Mortality of the Sardines, and* 3. *Contribution to the Problem of Estimating Populations of Fish with Particular Reference to Fish Caught in Schools, Such as Sardines.* A revision of the third report appeared as Neyman (1949).

At the outset of Neyman (1949), he provides Table 2. From it he infers a "rapid decline ... observed in spite of a reported increase in fishing effort..." A second table, Table 3, gives the amount (in arbitrary units) of sardines landed on the West Coast in the seasons 1941–1946, classified by age and season. Figure 2 graphs the

TABLE 2
Seasonal catch of California sardines 1943–1948 *in* 1000 *tons*

Year	Seasonal catch
1943–1944	579
1944–1945	614
1945–1946	440
1946–1947	248
1947–1948	110

TABLE 3
Numbers, $m_{t,a}$, of sardines caught by age and year

Season	41–2	42–3	43–4	44–5	45–6
age = 1	926.0	718.0	1030.0	951.0	493.0
2	6206.0	2512.0	1308.0	2481.0	1634.0
3	3207.0	4496.0	2245.0	1457.0	1529.0
4	868.0	1792.0	2688.0	1298.0	799.0
5	361.0	478.0	929.0	1368.0	407.0
6	95.1	169.4	327.0	498.5	299.2
7	47.2	36.0	98.4	148.0	111.2

amounts with lines joining the values for the same sardine age. One sees the high numbers in the early 1940s followed by decline. The interpretation is tricky because the numbers reflect both the fish available and the effort put into catching them. Neyman (1948) discussed the effect of migration and concluded that it was unimportant for his current purposes.

Turning to analysis Neyman remarks,

> "Certain publications dealing with the survival rates of the sardines begin with the assumption that both the natural death rate and the fishing mortality are independent of the age of the sardines, at least beginning with a certain initial age." (Neyman, 1948).

and goes on to say,

> "In the present note a method is suggested whereby it is possible to a (sic) test the hypothesis that the natural death rate is independent of the age of the sardines" (Neyman, 1949).

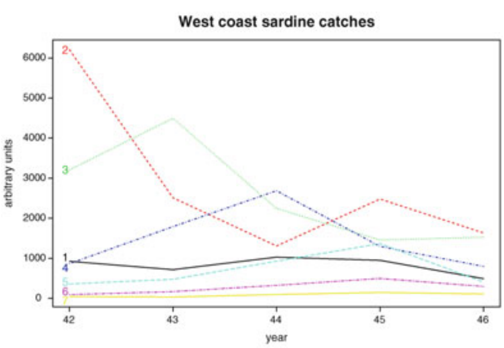


FIG. 2. *The data of Table 3 plotted versus year. The curve labels 1–7 index the age groups.*

To address the independence issue, and possibly motivated by Table 3, Neyman sets up a formal structure as follows. Let $N_{t,a}$ be "the number of fish available aged a at the beginning of season t and exposed to the risk of being caught." Here these numbers are collected into a vector, $\mathbf{N}(t) = [N_{t,a}]$. Next $n_{t,a}$ is set to be the expected number of sardines aged a caught during season t, and $P_t = 1 - Q_t$ set to be the "fishing survival rate in the tth year." Continuing, $p_a = 1 - q_a$ denotes the "natural survival rate at age a" and q_a the "rate of disappearance." The rate of mass emigration during season t is denoted by M_t.

The following null hypothesis may be set down concerning the mortality rates,

$$H_0 : q_{a_0} = q_{a_0+1} = \cdots = q_a, \quad a > a_0.$$

Specific assumptions Neyman considered were:

(i) $Q_t = n_{t,a}/N_{t,a}$, season t fishing mortality,
(ii) $N_{t+1,a+1} = N_{t,a}(1 - Q_t)(1 - q_a)$,
(iii) $N_{t+1,a+1} = N_{t,a}(1 - Q_t)(1 - M_t)(1 - q_a)$.

Assumptions (ii) and (iii) involve separation of the age and season variables. For identifiability of the model Neyman writes

$$n_{t+1,a+1} = n_{t,a} R_t p_a = n_{t,a} r_t p_a^*$$

with

$$R_t = \frac{P_t(1 - M_t)}{Q_t} Q_{t+1}, \quad r_t = R_t/R_1, \quad p_a^* = R_1 p_a.$$

One notes from these expressions that $n_{t+1,a+1}/n_{t,a}$ separates into a function of t and a function of a. This last led Neyman to work with logs of ratios in his analyses. (There will be more on this choice later.) He estimates $p_a^* = R_1 p_a$, which is proportional to p_a under his definitions, from the data.

The p_a^* estimates are provided in Table 4 and graphed in Figure 3. One sees a steady decrease with age. Table 5 provides $\hat{n}_{t,a}$ based on assumptions (i) and (ii) [or (i) and (iii)].

Neyman's conclusions included,

TABLE 4
Parameter estimates (these are the values obtained in calculations for this article)

Season	41–2	42–3	43–4	44–5
p_a^*	0.5944	0.4854	0.4629	0.4056
r_t	1.0	1.2252	1.0695	0.6259

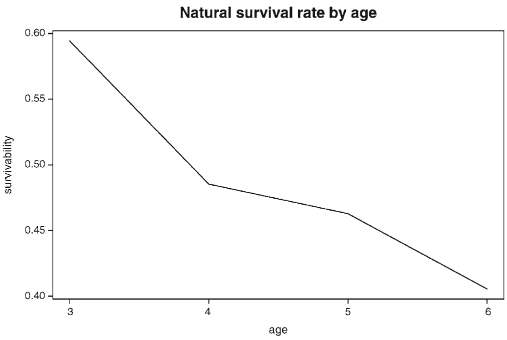


FIG. 3. *Estimates of the natural survival rate, p^* as a function of age.*

> "While in certain instances the differences between Tables IV (here Table 3) and VII (here Table 5) are considerable, it will be recognized that the general character of variation in the figures of both tables is essentially similar" (Neyman, 1948, pages 14–15).

No formal test of H_0 was set down, but Neyman concludes that,

> "Since the *estimates* of the p_a^* decrease rather regularly, it seems that the true natural survival rates must decrease with the increase in age..." (Neyman, 1948).

Basic elements of this example include working with empirical data, noting the age and season structure explicitly, and working with a Markov-like setup. Interestingly Neyman talks of an expected value, but no full probability model is set down.

In part this example is meant to get the reader in the mood for an age-structured population analysis to appear later in the paper.

The final example taken from Neyman's work follows.

TABLE 5
Estimates of the $n_{t,a}$, the expected numbers of sardines

Season	1	2	3	4	5
age, 3	2810.0	3556.3	2117.9	1761.6	—
4	1059.3	1684.3	2611.7	1355.7	661.0
5	383.7	514.2	1001.7	1355.7	412.5
6	91.9	77.6	291.6	495.9	391.7
7	—	37.3	88.2	126.5	125.9

4.3 Example 3. WEATHER MODIFICATION

> "The meteorological aspects of planning an experiment with cloud seeding depend upon the past experience, upon what the experimenter is prepared to adopt as a working hypothesis and upon the questions that one wishes to have answered by the experiment" (Neyman and Scott, 1965–1966).

Cloud seeding became an interest of Jerzy Neyman starting in the early 1950s. He and his collaborators studied data from the Santa Barbara and Arizona rainfall experiments. Neyman and Scott moved on to study data from a Swiss weather modification experiment that had been designed to see if cloud seeding could reduce hailfall. The experiment was carried out in the Canton of Ticino during the period 1957–1963 and was called Grossversuch III.

The experimental design involved each day deciding whether conditions were suitable to define an "experimental day." If a day was suitable seeding was or was not carried out the following day, randomly. Seeding, if any, lasted from 0730 to 2130 hours local time. Rainfall measurements that had been made in Zurich, about 120 km away from Ticino, were studied.

In the course of their work Neyman and Scott discovered so-called "far-away effects," that is, an apparent increase in amount of rainfall at a distance. See Neyman, Scott and Wells (1969).

Figure 4 provides a reconstruction of a graph that Neyman and Scott (1974) employed to highlight the result. It presents average hourly rainfall totals smoothed by a running mean of 3, for the experimental days when a "warm" stability layer and southerly winds were present.

To obtain the data of Figure 4 the values were read off a graph in Neyman and Scott (1974). The solid curve refers to experimental days with seeding, the dashed to those without. There were 53 experimental days with seeding and 38 without.

What Neyman and Scott focused on in the figure was an apparent effect of seeding in Zurich starting about 1400 hours in the afternoon.

They wrote as follows,

> "...the curves...represent averages of a number of independent realizations of certain stochastic processes. The 'seeded' curves are a sample from a population of one kind of processes and the 'not seeded' curve a sample from another. For an initial period of a number of hours...the two kinds

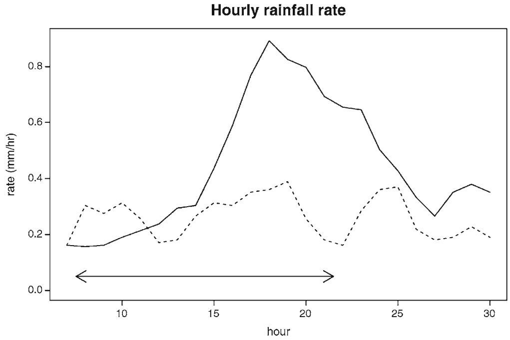


FIG. 4. *Comparison of seeded and not seeded hourly precipitation amounts on days with southerly upper winds. The solid line is rainfall for seeded days and the dashed line for unseeded. The horizontal line with arrowheads represents the seeding period at Ticino. A three-hour moving average had been employed to smooth hourly totals.*

> of processes coincide. Thereafter, at some unknown time T, the two processes may become different. Presumably, all the experimental days differ from each other, possibly depending on the direction and velocity of prevailing winds. Therefore, the time T must be considered as a random variable with some unknown distribution. *The theoretical problem is to deduce the confidence interval for the expectation of T, ...*" (Neyman and Scott, 1974).

This problem will be returned to later in the paper.

4.4 Neyman and Exploratory Data Analysis (EDA)

Given my statistical background it would be remiss not to provide some discussion of EDA in Neyman's work. Quotes are one way to bring out pertinent aspects of Neyman's attitude to EDA. One can conclude that exploratory data analysis was one of his talents.

> "... while hunting for a big problem I certainly established the habit, ..., to neglect rigour" (Neyman, 1967).

> "PAGE asked whether the elimination of outliers–supposed projected foreground or background objects recognized by discordant velocities–would not in itself introduce unwanted selection effects. NEYMAN advised that the investigator try calculations with and without outliers, then make up his mind 'which he likes best', while retaining both."

> "Compared with the old style experiments, characterized by the attitude 'to prove,' the proposed experiment would be substantially richer. ... This, then, will implement the attitude 'to explore' contrasted with that 'to prove' " (Neyman and Scott, 1965–1966).

> "We emphasize that such an investigation is only exploratory; whatever may be found are only clues which must be studied further and hopefully verified in other experiments" (Dawkins, Neyman and Scott, 1977).

JN did not seem to use residuals much. However, in Neyman (1980) one does find,

> "... one can observe a substantial number of consecutive differences that are all negative while all the others are positive. ... the 'goodness of fit' is subject to a rather strong doubt, irrespective of the actual computed value of χ^2, even if it happens to be small" (Neyman, 1980).

Neyman et al. (1953) proposed an innovative EDA method to examine variability: specifically, given values X and Y with the same units, plot $X - Y$ and $|X - Y|$ versus $(X + Y)/2$. Figure 5 compares Tables 3 and 5 of the sardine analysis this way. In the two panels one sees wedging, that is, an increase of variability with size. This suggests that a transformation of the data might simplify the matter. Neyman did employ the log transform in his analysis of the sardine data consistent with the multiplicative character of the model.

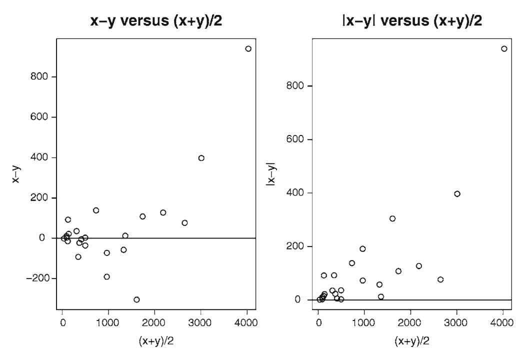


FIG. 5. *Comparisons of Table* 3, *x-values and Table* 5, *y-values. The left panel plots* $(x - y)$ *versus* $(x + y)/2$ *and the right* $|x - y|$ *versus* $(x + y)/2$.

5. FOUR EXAMPLES OF RANDOM PROCESS DATA ANALYSIS

The following examples report some of my work, typically with collaborators. They were suggested in part by my exposure to JN and to the preceding examples.

5.1 Example 4. SHEEP BLOWFLIES

In Example 2 above Neyman studied data on sardines that included the actual age information. However, it can be the case that, even though a population is age-structured, only aggregate data are available, and actual age information is unavailable. This is the case in the example that follows. To deal with it a state-space model is set down. The (unobserved) state vector is taken to be the counts of individuals in the various age groups. The story and details follow.

The tale begins with the mathematician John Guckenheimer and the then entomologist George Oster coming to meet with DRB. They had in hand data on a population of *lucilia cuprina* (Australian sheep blowflies). The data concerned an experiment maintained from 1954 to 1956 under constant, but limited conditions by A. J. Nicholson, then Chief Division of Entomology, CSIRO, Australia.

At the beginning of the experiment 1000 eggs were placed in a cage. Every other day counts were made of the number of eggs, of nonemerging flies' eggs, of the number of adult flies emerging, and of the number of adult fly deaths. The life stages, and corresponding time periods, of these insects are given in Table 6. Further details of the experiment may be found in Nicholson (1957). To get digital values Oster and a student took a photo of one of the figures in that paper. The photo was then projected on a wall and numerical values read off. Unfortunately some of the populations' sizes went off the top of the figure. The values for these cases were obtained when DRB later visited CSIRO.

Guckenheimer and Oster's question was whether these data displayed the presence of a strange attractor, a concept from nonlinear dynamic systems analysis; see Brillinger et al. (1980) and Guckenheimer and Holmes (1983). The behavior evidenced in the second half of the series graphed in Figure 6 is what attracted Guckenheimer and Oster's attention. The initial oscillations come from the usual lifespan of the adults.

TABLE 6
Life stages and their lengths for sheep blowflies

Life stage	Length
egg	12–24 hours
larva	5–10 days
pupa	6–8 days
adult	1–35 days

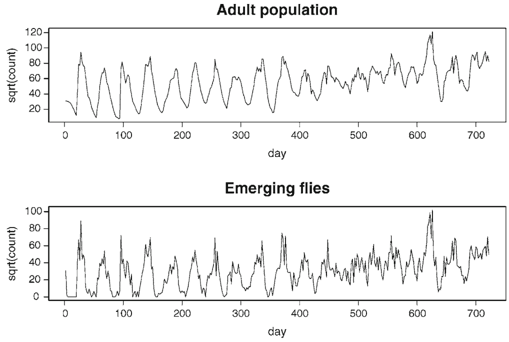


FIG. 6. *Square roots of counts for the Nicholson blowfly data. The top panel provides the number of adults and the bottom the number of emerging pupae.*

In the particular experiment studied here the amount of food put in the fly cage was deliberately restricted. This meant that the fecundity of the females was reduced. When much food was available many eggs were laid. With insufficient food the number of eggs was reduced. This led to boom periods and bust periods in the population size.

Figure 6 graphs the square roots of total adult population count, as well as of the number of flies emerging. The time points are every other day over a period of approximately two years. In the graphs one sees an initial periodic behavior in both series followed by rather irregular behavior. The square roots were plotted to make the variability of the display more nearly constant.

Brillinger et al. (1981) proceeded by setting down a formal state-space model for the situation as follows:

$t = 0, 1, 2, \ldots,$ represents time, observations being made every other day,

E_t, the number of emerging flies in time period $(t, t + 1]$,

$\mathbf{E}_t$, the entrant column vector; it has E_t in row 1 and 0 elsewhere,

N_t, the adult population at time t.

Constructs include:

$\mathbf{N}_t = [N_{it}]$, the state vector; in it row i gives the number of population members aged $i - 1$ at time t,

$\mathbf{P}_t = \mathbf{P}(\mathbf{H}_t) = [p_{i,t}]$, the survival matrix. The entry in row $i + 1$, column i gives the proportion surviving age i to age $i+1$. $\mathbf{P}_t$ is taken as depending on the history $\mathbf{H}_t$, that is, the collection of the data values up to and including time t.

The available data are E_t and N_t.

The measurement equation, corresponding to the observed population size is, $N_t = \mathbf{1}'\mathbf{N}_t$. The dynamic equation is

$$\mathbf{N}_{t+1} = \mathbf{P_t}\mathbf{N}_t + \mathbf{E}_{t+1} + \textit{fluctuations}.$$

This expression updates the counts of adult flies in each age group, starting from $\mathbf{N}_0 = 0$. The fluctuations represent variabilities in those numbers.

In one analysis (Brillinger et al., 1980), the following nonlinear age and density model was employed:

$$\begin{aligned} p_{i,t} &= 1 - \text{Prob}\{\textit{individual aged } i, \\ &\qquad \textit{dies aged } i \textit{ at time } t|\mathbf{H}_t\} \\ &= (1-\alpha_i)(1-\beta N_t)(1-\gamma N_{t-1}). \end{aligned} \tag{6}$$

This model allows survival dependence on age, i, on the current population size, N_t and on the preceding population size, N_{t-1}. The final term allows the possibility that it takes some time for the limited or excess food situation to take effect.

Weighted least squares was employed in the fitting of model (6). On the basis of residual plots weights were taken to be N_t^2. Hence writing $D_t = N_{t-1} - N_t + E_t$ one seeks

$$\min_\theta \sum_t \left(D_{t+1} - \sum_i q_i m_{i,t}\right)^2 \Big/ N_t^2,$$

where $\theta = \{\alpha_i, \beta, \gamma\}$ and $m_{i,t}$ is the conditional expected value, $E\{N_{i,t}|\mathbf{H}_t\}$. Graphs of the estimates of the individual entries of $\mathbf{N}_t$ are provided in Brillinger et al. (1980).

Synthetic series were computed to assess the reasonableness of the model (6). In the simulations counts of deaths in the time period $(t-1, t]$, are computed. The deaths, D_t, are plotted in the top panel of Figure 7. The value D_t is thought of as fluctuating about the value

$$\sum_i q_{i,t} N_{i,t}$$

where $N_{i,t}$ is the population aged i at time t.

The results of two simulations are provided in Figure 7. In the first, the middle series, the variability is taken as binomial. In the second, the bottom series, the variability is taken as independent normal, mean 0, standard deviation $\hat{\sigma} N_t$ with $\hat{\sigma}$ estimated from the weighted least squares results. That the appearances of

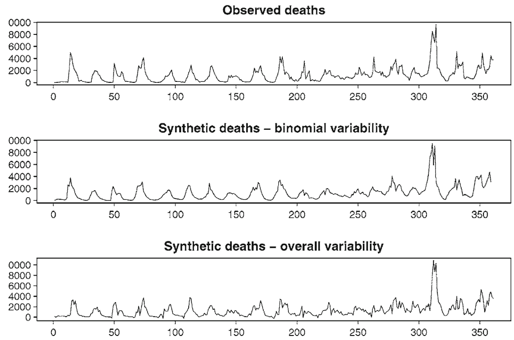


FIG. 7. *Death series and synthetic death series using the model (6).*

the synthetic series are so close to the actual series relates to the use of the common stimulus series, E_t.

A byproduct of this analysis is that because the measurement equation, $N_t = \mathbf{1}'\mathbf{N}_t$, is of simple addition form by this analysis one has developed a decomposition of the population total series into individual age series. These are graphed in Brillinger et al. (1980).

The fitted death rates were nonlinear in the population size, so mathematically a strange attractor might be present (Brillinger, 1981).

In this situation one is actually dealing with a nonlinear closed loop feedback system with time lags. Guttorp (1980), in his doctoral thesis, completed the analysis of the feedback loop modeling the births.

5.2 Example 5. WEATHER MODIFICATION REVISITED

Neyman and Scott's problem referred to in Example 3 was addressed in Brillinger (1995). At issue was making inferences concerning the travel time of seeding effects from Ticino to Zurich. The approach of the paper was to envisage a succession of travel time effects that started at times throughout the seeding period. This way one had replicates to allow employment of statistical characteristics. A conceptual model involving a gamma density for the travel velocity of the seeding effect was employed. The data themselves were graphed in Figure 4 above.

The model employed is the following. Suppose that "rain particles" created at Ticino move off toward Zurich with a possibility of leading to a cluster of rain drops there. Suppose that the particles are born at Ticino at the times σ_j of a point process M, at rate $p_M(t)$. Suppose that the travel times from the particles' times of creation, U_j, to Zurich are independent of each other with density $f_U(\cdot)$. Let N denote the point process of times, τ_j, at which the particles arrive at Zurich and $p_N(t)$ denote the rate of that process.

If the jth particle moves with velocity v_j and the distance to be traveled is Δ, then its travel time is $u_j = \Delta/v_j$ and since

$$\sum_j \delta(t-\tau_j) = \sum_j \delta(t-\sigma_j - u_j)$$

with $\delta(\cdot)$ the Dirac delta, one has

$$p_N(t) = \int p_M(t-u) f_U(u)\,du.$$

Let the amounts, R_j, of rain falling at Zurich associated with the individual particles, be statistically independent of the particles. Let μ_R denote $E\{R_j\}$. Then the rate of rainfall at Zurich at time t is

$$p_X(t) = \mu_R \int p_M(t-u) f_U(u)\,du.$$

Next let $X(t)$ denote the cumulative amount of rain falling at Zurich from time 0 to time t. Its expected value is

$$E\{X(t)\} = \int_0^t p_X(v)\,dv.$$

Turning to Figure 4, Neyman and Scott employed a running mean of order 3 of the hourly totals to get the values graphed. These are the data available for analysis. (The hourly values appear to be lost.) The running mean may be written

$$Y(t) = \tfrac{1}{3}(X(t+1) - X(t-2))$$

for $t = 2, 3, \ldots$. Its expected value is

$$\begin{aligned} &\tfrac{1}{3}\int_{t-2}^{t+1} p_X(v)\,dv \\ &\quad = \tfrac{1}{3}\mu_R \int_{t-2}^{t+1} \int p_M(v-u) f_U(u)\,du\,dv. \end{aligned} \tag{7}$$

One can now view the Neyman–Scott problem as related to estimating $f_U(\cdot)$ of (7), that is, estimating the travel time density given the available data.

To proceed, the seeding rate $p_M(t)$ will be taken to be constant on the time interval from 0730 to 2130 hours and to be 0 otherwise. It will be further assumed that the travel time of U has the form θ/W with θ a parameter, and with W Weibull, having scale 1, and shape s. Brillinger (1995) took the gamma as the density, but a review of the literature of wind speeds suggests that the Weibull would be more appropriate.

Writing $p_M(t) = C$ for $A < t < B$ (here A = 7.5 and B = 21.5 hr) one has the regression function

$$\begin{aligned} E\{Y(t)\} = \alpha + \frac{C}{3}\mu_R\Big[&\int_{t-2-A}^{t+1-A} F_U(u)\,du \\ &- \int_{t-2-B}^{t+1-B} F_U(u)\,du\Big], \end{aligned} \tag{8}$$

where $F_U(\cdot)$ denotes the distribution function of U, in the case of seeding and α is the natural level of rainfall. With the assumed Weibull velocity distribution, (8) may be evaluated in terms of G the distribution function of the Weibull. Specifically,

$$\begin{aligned} \int_0^x F_U(u)\,du = x\Big[1 - G\Big(\frac{1}{x}, s\Big)\Big] \\ - \frac{s}{s-1}\Big[1 - G\Big(\frac{1}{x}, s-1\Big)\Big]. \end{aligned}$$

(To derive this one replaces $\mathrm{Prob}\{1/W \le u\}$ by $\mathrm{Prob}\{W \ge 1/u\}$ and integrates by parts.)

The estimates of the unknowns $\mu = \theta\Gamma((s-1)/s)$ (the average travel time), s, α, $\beta = C\mu_R/3$ were determined by ordinary least squares, weighting the seeded terms by 53 and the unseeded by 38 to handle the unequal numbers of seeded and unseeded cases.

Figure 8, left-hand panel, presents the data (solid curve) and the fitted (dotted) curve. The parameter estimates obtained are:

$$\hat{\mu} = 4.78(0.47)\ \text{hr},$$
$$\hat{s} = 6.68(5.12),$$
$$\hat{\alpha} = 0.24(0.02),$$
$$\hat{\beta} = 1.69(0.19).$$

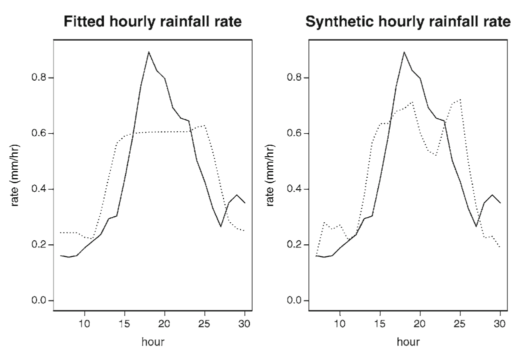


FIG. 8. *Left panel—actual and fitted (dotted line) rainfall when seeding. Right panel—actual and synthetic in the case of seeding (dotted line).*

[The standard errors, assumed the errors to be i.i.d.]

One sees in the left-hand panel that the actual data have a peak near 1800 during 0730 and 2130 hr, whereas the fitted has a flat top. Perhaps the birthrate, $p_M(t)$, of particles is not approximately constant as assumed above. Perhaps the distribution, $f_U(u)$, depends on time. Perhaps the result is due to natural variability.

A synthetic plot is generated to examine the fit. Specifically the fluctuations of the unseeded days have been added to the fitted curve and graphed in the right-hand panel of Figure 8. Still the fitted curve is quite flat on the top, in contrast to the Neyman–Scott data curve which is noticeably peaked. The added fluctuations do not bring the curve up to the data level.

Returning to the Neyman–Scott problem of Section 3, the second quotation there refers to T, a random time at which seeding first shows up in Zurich. The U's represent the lengths of time it takes for an effect just initiated to arrive. One can take the expected value, EU, to be ET. Using the parameter estimates above, an approximate 95% confidence interval for the expectation of T is

$$4.78 \pm 2 * 0.47 \text{ hours.}$$

More work needs to be done with this example. A indication of how to proceed is provided by Figure 8. The data graph is pointed, whereas the fitted is flat-topped.

5.3 Example 6. ELK MOTION

The data now studied were collected at the Starkey Experimental Forest and Range (Starkey), in Northeastern Oregon. Quoting from the website, fs.fed.us/pnw/starkey/publications/by_keyword/Modelling_Pubs.shtml.

Starkey was set up by the US Forest Service for

> "Long-term studies of elk, deer, and cattle—examining the effects of ungulates on ecosystems."

A specific management question of concern is whether recreational uses by humans would affect the animals there substantially. Further details about Starkey and the recreation experiment may be found in Brillinger et al. (2001a, 2001b, 2004), Preisler et al. (2004) and Wisdom (2005).

In the first analysis presented the elk were not deliberately disturbed and their paths were sampled at discrete times. This gave control data for an experiment. An all-terrain vehicle (ATV) was introduced and driven around on the roads in the NE Meadow of Starkey. The analysis to be presented quantifies the effect of the disturbance. The locations of both the ATV and the elk were monitored by GPS methods.

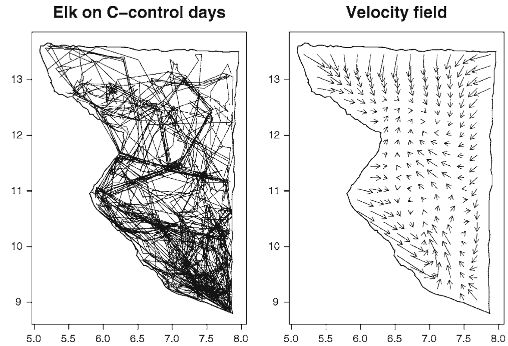


FIG. 9. *Northeast pasture of the Starkey Reserve and the elk motion on control days. The left panel shows the paths of 8 elk, superposed. The right panel displays the estimated velocity field* $\hat{\boldsymbol{\mu}}(\mathbf{r})$ *as a vector field.*

There were 8 elk in the study. The ATV was introduced into the meadow over 5-day periods. This was followed by 9-day "control" periods with no ATV. In the control periods the animals were located every 2 hours. In the ATV case elk locations were estimated about every 5 min. The ATV's locations were determined every second.

Figure 9, left-hand panel, shows observed elk trajectories superposed. One sees the animals constrained by the fence, but moving about most of the Reserve. They often visit the SE corner. The straight line segments result from the locations being obtained only every 2 hours in this control case.

The animal motion will be modeled by the SDE

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t))\,dt + \boldsymbol{\sigma}\,d\mathbf{B}(t) \tag{9}$$

with $\mathbf{r}(t)$ the location at time t, $\mathbf{B}$ a bivariate standard Brownian motion and σ a scalar. The function $\boldsymbol{\mu}$ is assumed to be smooth. The discrete approximation (3) becomes a generalized additive model with Gaussian errors; see Hastie and Tibshirani (1990).

The resulting estimate is displayed as a velocity vector field $(\hat{\mu}_1(\mathbf{r}), \hat{\mu}_2(\mathbf{r}))$ in the right-hand panel of Figure 9 employing arrows. One sees the animals moving along the boundary and toward the center of the pasture. The fence can be ignored in this data analysis.

The fence is important in preparing a synthetic trajectory. What was done in that connection was to employ the relation (3) with the proviso that if it generated a point outside the boundary, then another point was

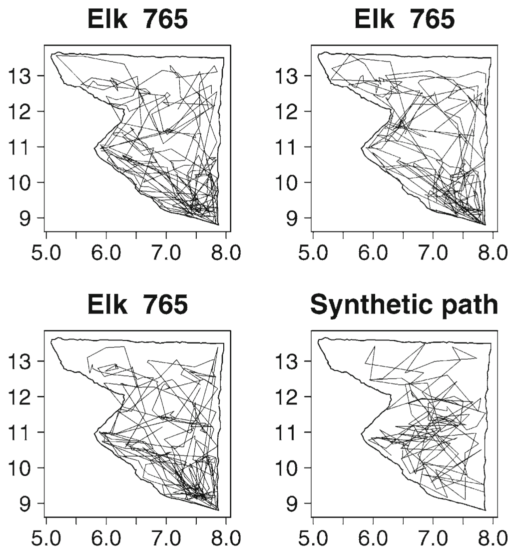


FIG. 10. *The first three panels display the tracks of the indicated animals. The final panel, lower right, is a synthetic path.*

generated until one stayed within the boundary. This is a naive but effective method if the t_i of (3) are close enough together. Better ways for dealing with boundaries are reviewed in Brillinger (2003).

Figure 10 shows the trajectories of three of the animals. The lower right panel presents a synthetic path generated including 188 location points. The synthetic trajectory does not appear unreasonable.

Consideration now turns an analog of regression analysis for trajectories, that is, there is an explanatory variable. The explanatory variable is the changing location, $\mathbf{x}(t)$, of the ATV. The left-hand panel of Figure 11 shows the routes of the ATV cruising around the roads

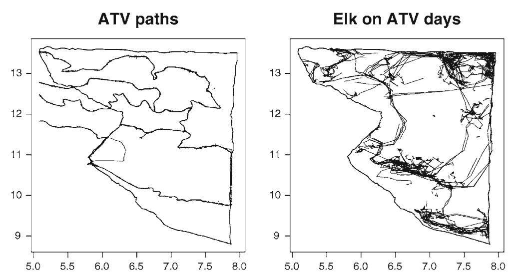


FIG. 11. *The left panel shows the ATV's route, while the right shows the elk paths in the presence of the ATV. The ATV passes in and out some gates on the lefthand side.*

of the Meadow. The right-hand panel provides the superposed trajectories of the 8 elk. One sees, for example, the elk heading to the NE corner, possibly seeking refuge. The noise of the ATV is surely a repellor when it is close to an elk, but one wonders at what distance does the repulsion begin?

The following model was employed to study that question. Let $\mathbf{r}(t)$ denote the location of an elk, and $\mathbf{x}(t)$ the location of the ATV, both at time t. Let τ be a time lag to be studied. Consider

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t))\,dt + \boldsymbol{\nu}\big(|\mathbf{r}(t) - \mathbf{x}(t-\tau)|\big)\,dt + \boldsymbol{\sigma}\,d\mathbf{B}(t). \tag{10}$$

The times of observation differ for the elk and the ATV. They are every 5 minutes for the elk when the ATV is present and every 1 sec for the ATV itself. In the approach adopted location values, $\mathbf{x}(t)$, of the ATV are estimated for the elk observation times via interpolation. The ATV observed times are close in time, namely 1 second, so the interpolation should be reasonably accurate.

Expression (10) allows the change in speed of an elk to be affected by the location of the ATV τ time units earlier. Assuming that $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in (10) are smooth functions, then the model may be fit as a generalized additive model. Figure 12 graphs $|\hat{\boldsymbol{\nu}}(d)|$, d being the distance of the elk from the ATV. (The norm $|\boldsymbol{\nu}| = \sqrt{\nu_1^2 + \nu_2^2}$ here.) One sees an apparent increase in the speed of the elk, particularly when an elk and the ATV are close to each another. The increased speed is apparent at distances out to about 1.5 km. An upper 95% null level is indicated in Figure 12 by a dashed line. One sees less precise measurement at increasing large values of τ.

The estimation of $|\boldsymbol{\nu}(d)|$ was also carried out in the absence of the $\boldsymbol{\mu}$ term in the model. The results were very similar. This gives some validity to interpreting the estimate $\hat{\boldsymbol{\nu}}(d)$ on its own despite the presence of $\boldsymbol{\mu}$ in the model.

In conclusion, the ATV is having an apparent effect and it has been quantified to an extent by the graphs of Figure 12.

These results were presented in Brillinger et al. (2004). Also Wisdom (2005) and Preisler et al. (2004) modeled the probability of elk response to ATVs in a different way. They used data for the year 2002, and measured the presence of an effect in another manner.

5.4 Example 7. MONK SEALS: A POTENTIAL FUNCTION APPROACH

Hawaiian monk seals are endemic to the Hawaiian Islands. The species is endangered and has been declin-

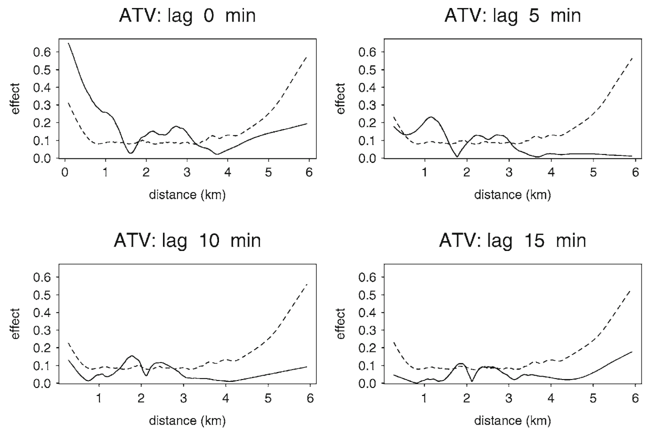


FIG. 12. *The function* $|\hat{\mathbf{v}}|$ *of* (10) *for the time lags* 0, 5, 10, 15 *minutes.*

ing for several decades. It now numbers about 1300. One hypothesis accounting for the decline in numbers is the poor growth and survival of young seals owing to poor foraging success. In consequence of the decline data have been collected recently on the foraging habitats, movements, and behaviors of these seals throughout the Hawaiian Islands Archipelago. Specific questions that have been posed regarding the species include:

What are the geographic and vertical marine habitats that Hawaiian monk seals use?

How long is a foraging trip?

For more biological detail see Stewart et al. (2006) and Brillinger, Stewart and Litnnan (2006, 2008).

The data set studied is for the west side of the main Hawaiian Island of Molokai. The work proceeds by fitting an SDE that mimics some aspects of the behavior of seals. It employs GPS location data collected for one seal. An SDE is found by developing a potential function.

The data are from a three-month journey of a juvenile male while he foraged and occasionally hauled out onshore. The track started 13 April 2004 and ended 27 July 2004. The animal was tagged and released at the southwest corner of Molokai; see Figure 13, top left panel. The track is indicated for six contiguous 15-day periods. The seal had a satellite-linked radio transmitter glued to his dorsal pelage. It was used to document geographic and vertical movements as proxies of foraging behavior.

There were 754 location estimates provided by the Argos satellite service, but many were suspicious. Associated with a location estimate is a prediction of the location's error (LC or location class). The LC index takes on the values 3, 2, 1, 0, A, B, Z. When LC = 3, 2 or 1 the error in the location is predicted to be 1 km or less, and these are the cases employed in the analysis here.

The estimated times of locations are irregularly spaced and not as close together as one might like. This can lead to difficulties of analysis and interpretation.

The motivating SDE of the analysis is

$$(11)\quad d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t))\,dt + \sigma\,d\mathbf{B}(t), \quad \mathbf{r}(t) \in F,$$

with $\boldsymbol{\mu} = -\nabla H$, H a potential function, σ scalar, $\mathbf{B}$ bivariate Brownian and F the region inside the 200-fathom line up to Molokai. There was discussion of the

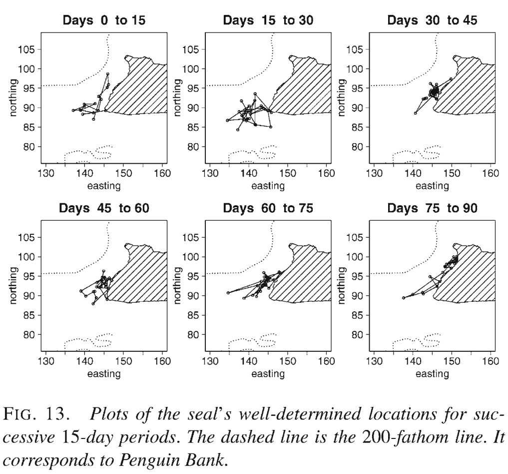


FIG. 13. *Plots of the seal's well-determined locations for successive* 15*-day periods. The dashed line is the* 200*-fathom line. It corresponds to Penguin Bank.*

potential approach in Section 3. The potential function employed here is

$$H(x, y) = \beta_{10}x + \beta_{01}y + \beta_{20}x^2 + \beta_{11}xy + \beta_{02}y^2 + C/d_M(x, y) \tag{12}$$

where d_M is the shortest distance to Molokai from the location (x, y). The final term in (12) is meant to keep the animal off Molokai.

The model was fit by ordinary least squares taking $C = 7.5$. In the analysis the number of data points was 142 and the parameter estimates obtained were $\hat{\boldsymbol{\beta}} = (93.53, 8.00, -0.47, 0.47, -0.41)$, and $\hat{\sigma} = 4.64$ km. Figure 14 shows the estimated potential function, $\hat{H}$. This seal is pulled into the middle of the concentric contours, with the Brownian term pushing it about.

Synthetic plots were generated to assess the reasonableness of the model and to suggest departures. Figure 15 shows the results of a simulation of the process (only one path was generated) having taken the parameter values to be those estimated and having broken the overall trajectory down into six segments as in Figure 13, to which it may be compared. The sampling interval, dt, employed in the numerical integration of the fitted SDE is 1 hour. The paths were constrained to not go outside the 200-fathom line and not to go on the island. (See Brillinger, 2003, for methods of doing this.) The locations of the time points of the synthetic track are the times of the observed locations. This allows direct comparison with the data plot of Figure 13. The variability of Figure 15 is not unlike that of Figure 13.

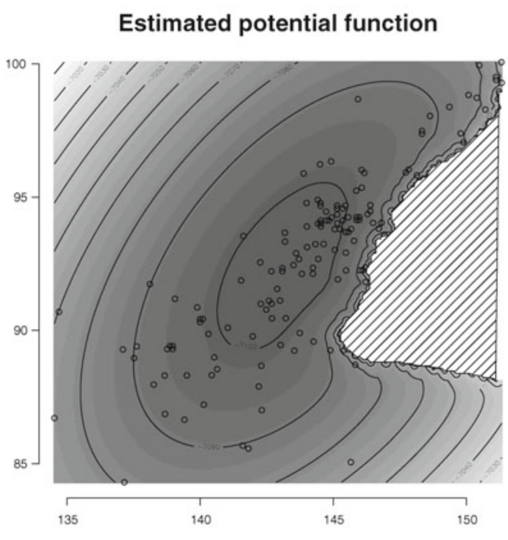


FIG. 14. *The fitted potential function obtained using the potential function* (12). *The darker the values are, the deeper the potential function is. The slanted line region is Molokai.*

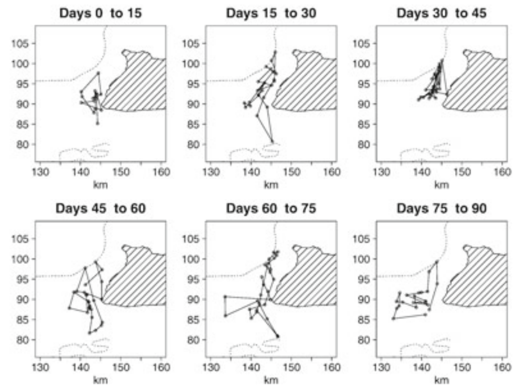


FIG. 15. *Synthetic plots of the model* (11) *having fit the potential function* (12). *The times are those of the data of Figure* 13.

In this work the scattered, sometimes unreasonable, satellite locations have been cleaned up and summarized by a potential function. The general motion of the animal on a foraging trip has been inferred and simulated. It has been learned that the animal stays mostly within Penguin Bank and tends to remain in an area off the west coast of Molokai.

There are other examples of potential function estimation in Brillinger, Stewart and Littnnan (2006, 2008) and Brillinger (2007a, 2007b).

6. CONCLUSION

> "Say what you are going to say, say it, then say what you said" (Neyman, Personal communication).

It was a great honor to be invited to present the Neyman Lecture. I attended many Neyman Seminars and made quite a few presentations as well. A side effect of the work was the very pleasant experience of reading through many of Neyman's papers in the course of preparing the lecture and the article. So many personal memories returned.

The emphasis has been placed on dynamic and spatial situations. There are three examples of JN and ELS; two concern temporal functions and one spatial. Four examples are provided of the work of DRB with collaborators. Two are temporal and two are spatial-temporal. The data are from astronomy, fisheries, meteorology, insect biology, animal biology and marine biology. The models and analyses were not all that difficult. The statistical package R was employed.

The field of sampling was another one to which Neyman made major contributions; see Neyman (1934, 1938a). It can be argued that work in sampling had a more profound impact on the United States than any of his other applied work. I looked hard but did not find reference to repeated sample surveys in JN's work. Had I, there would have been some discussion of dynamic sample survey.

The reader cannot have missed the many references to Elizabeth Scott. In fact in many places in my lecture the title could have been the Neyman–Scott Lecture. From the year 1948 on, 55 out of 140 of JN's papers were with her. Some 118 of Betty's publications are listed in Billard and Ferber (1991). One in the spirit of this lecture, Scott (1957), concerns the Scott effect, a biasing effect that occurs in galaxy observations because at greatest distances only the brightest would be observed. She developed a correction method (Scott, 1957).

I end with a wonderful and enlightening story concerning Jerzy Neyman. It was told by Alan Izenman at the lecture in Minneapolis. In the early 1970s the Berkeley Statistics Department voted to do away with language requirements. (There had been exams in two non-English languages.) In response in the graduate class that JN was teaching he announced that he was going to ask various people to give their presentation in their native, non-English, language. This continued for a number of weeks and languages.

ACKNOWLEDGMENTS

I have to thank my collaborators and friends regarding the matters described in the paper. These include A. Ager, J. Guckenheimer, P. Guttorp, A. Izenman, J. Kie, C. Littnan, H. Preisler, G. Oster and B. Stewart. I also thank the Editor and referees for their comments. They improved the article. Brillinger (1983) was an earlier attempt at the topic.

The research was supported by various NSF grants, the most recent being DMS-05-04162, "Random processes: data analysis and theory."

REFERENCES

BILLARD, L. and FERBER, M. A. (1991). Elizabeth Scott: Scholar, teacher, administrator. *Statist. Sci.* **6** 206–216. MR1127863

BRILLINGER, D. R. (1981). Some aspects of modern population mathematics. *Canad. J. Statist.* **9** 173–194. MR0640015

BRILLINGER, D. R. (1983). Statistical inference for random processes. In *Proceedings of the International Congress of Mathematicians* **1, 2** 1049–1061. PWN, Warsaw. MR0804757

BRILLINGER, D. R. (1995). On a weather modification problem of Professor Neyman. *Probab. Math. Statist.* **15** 115–125.

BRILLINGER, D. R. (2003). Simulating constrained animal motion using stochastic differential equations. *Probability, Statistics and Their Applications. IMS Lecture Notes Monograph Series* **41** 35–48. IMS, Beachwood, OH. MR1999413

BRILLINGER, D. R. (2007a). A potential function approach to the flow of play in soccer. *J. Quant. Anal. Sports* **3** 1–21. MR2304568

BRILLINGER, D. R. (2007b). Learning a potential function from a trajectory. *Signal Processing Letters* **14** 867–870.

BRILLINGER, D. R., GUCKENHEIMER, J., GUTTORP, P. and OSTER, G. (1980). Empirical modeling of population time series data: the case of age and density dependent vital rates. *Lectures Mathematics Life Sciences* **13** 65–90.

BRILLINGER, D. R., PREISLER, H. K., AGER, A. A. and KIE, J. (2001a). The use of potential functions in modelling animal movement. In *Data Analysis from Statistical Foundations* 369–386. Nova Science, Hutington. MR2034526

BRILLINGER, D. R., PREISLER, H. K., AGER, A. A., KIE, J. and STEWART, B. S. (2001b). Modelling movements of free-ranging animals. Technical Report 610, UCB Statistics.

BRILLINGER, D. R., PREISLER, H. K., AGER, A. A. and WISDOM, M. J. (2004). Stochastic differential equations in the analysis of wildlife motion. In *2004 Proceedings of the American Statistical Association, Statistics and the Environment Section*.

BRILLINGER, D. R., STEWART, B. S. and LITTNAN, C. L. (2006). A meandering *hylje*. In *Festschrift for Tarmo Pukkila* 79–92. Univ. Tampere, Finland.

BRILLINGER, D. R., STEWART, B. S. and LITTNAN, C. L. (2008). Three months journeying of a Hawaiian monk seal. *IMS Collections. Probability and Statistics: Essays in Honor of David A. Freedman* **2** 246–264. IMS, Beachwood, OH.

DAWKINS, S., NEYMAN, J. and SCOTT, E. L. (1977). The Grossversuch Project. *Transactions of Workshop on Total-Area Effects of Weather Modification.* Fort Collins, Colorado.

FIX, E. and NEYMAN, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* **23** 205–241.

GUCKENHEIMER, J. and HOLMES, P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* Springer, New York. MR0709768

GUTTORP, P. M. (1980). Statistical modelling of population processes. Ph.D. thesis, Univ. California, Berkeley.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London. MR1082147

HEYDE, C. (1994). A quasi-likelihood approach to estimating parameters in diffusion type processes. *J. Applied Probab.* **31A** 283–290. MR1274731

ITO, K. (1951). On stochastic differential equations. *Mem. Amer. Math. Soc.* **1951** 1–51. MR0040618

KENDALL, D. G., BARTLETT, M. S. and PAGE, T. L. (1982). Jerzy Neyman, 1894–1981. *Biog. Memoirs Fellows Royal Society* **28** 379–412.

KLOEDEN, P. E. and PLATEN, E. (1995). *Numerical Solution of Differential Equations*. Springer, Berlin.

LE CAM, L. (1995). Neyman and stochastic models. *Probab. Math. Statist.* **15** 37–45. MR1369790

LE CAM, L. and LEHMANN, E. (1974). J. Neyman—on the occasion of his 80th birthday. *Ann. Statist.* **2** vii-xiii. MR0342345

LEHMANN, E. (1994). Jerzy Neyman. *Biographical Memoirs* **63** 395–420. National Acad. Sci.

NELSON, E. (1967). *Dynamical Theories of Brownian Motion*, 2nd ed. Princeton Univ. Press. MR0214150

NEYMAN, J. (1934). On two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.* **97** 558–606.

NEYMAN, J. (1938a). Contributions to the theory of sampling human populations. *J. Amer. Statist. Assoc.* **33** 101–116.

NEYMAN, J. (1938b). *Lectures and Conferences on Mathematical Statistics and Probability Theory*. USDA, Washington.

NEYMAN, J. (1947). Outline of the statistical treatment of the problem of diagnosis. *Public Health Reports* **62** 1449–1456.

NEYMAN, J. (1948). *Reports on the Sardine Fishery.* Research Department, California Council of the Congress of Industrial Organizations.

NEYMAN, J. (1949). On the problem of estimating the number of schools of fish. *Univ. California Publ. Statist.* **1** 21–36. MR0038046

NEYMAN, J. (1960). Indeterminism in science and new demands on statisticians. *J. Amer. Statist. Assoc.* **55** 625–639. MR0116393

NEYMAN, J. (1966). Behavioristic points of view on mathematical statistics. In *On Political Economy and Econometrics* 445–462. Polish Scientific Publishers, Warsaw.

NEYMAN, J. (1967). *A Selection of Early Statistical Papers of J. Neyman*. Univ. California Press. MR0222983

NEYMAN, J. (1970). A glance at some of my personal experiences in the process of research. In *Scientists at Work* (T. Dalenius et al., eds.) 148–164. Almqvist and Wiksells, Uppsala. MR0396141

NEYMAN, J. (1976). Descriptive statistics vs. chance mechanisms and societal problems. In *Proc. 9th Int. Biometrics Conf.* **II** 59–68. Boston.

NEYMAN, J. (1980). Some memorable incidents in probabilistic/statistical studies. In *Asymptotic Theory of Statistical Tests and Estimation* (I. M. Chakravarti, ed.) 1–32. Academic, New York. MR0571333

NEYMAN, J. and SCOTT, E. L. (1956). The distribution of galaxies. *Scientific American* **September** 187–200.

NEYMAN, J. and SCOTT, E. L. (1959). Stochastic models of population dynamics. *Science* **130** 303–308. MR0107571

NEYMAN, J. and SCOTT, E. L. (1965–1966). Planning an experiment with cloud seeding. *Proc. Fifth Berkeley Symposium on Math. Statistics and Probab.* 327–350. Univ. California Press, Berkeley.

NEYMAN, J. and SCOTT, E. L. (1972). Processes of clustering and applications. In *Stochastic Point Processes* (P. A. W. Lewis, ed.) 646–681. Wiley, New York.

NEYMAN, J. and SCOTT, E. L. (1974). Rain stimulation experiments: design and evaluation. In *Proc. WMO/IAMAP Sci. Conf. Weather Modification* 449–457. WMO, Geneva.

NEYMAN, J., SCOTT, E. L. and SHANE, C. D. (1953). On the spatial distribution of galaxies a specific model. *Astrophys. J.* **117** 92–133.

NEYMAN, J., SCOTT, E. L. and SHANE, C. D. (1954). The index of clumpiness of the distribution of images of galaxies. *Astrophys. J. Supp.* **8** 269–294.

NEYMAN, J., SCOTT, E. L. and WELLS, M. A. (1969). Statistics in meteorology. *Rev. Inter. Stat. Inst.* **37** 119–148.

NICHOLSON, A. J. (1957). The self-adjustment of populations to change. *Cold Spring Harbor Symp. Quant. Biol.* **22** 153–173.

PEARSON, K. (1900). *The Grammar of Science*. Dent, London.

PREISLER, H. K., AGER, A. A., JOHNSON, B. K. and KIE, J. G. (2004). Modelling animal movements using stochastic differential equations. *Environmetrics* **15** 643–657.

REID, C. (1998). *Neyman*. Springer, New York. MR1480666

SCOTT, E. L. (1957). The brightest galaxy in a cluster as a distance indicator. *Astron. J.* **62** 248–265.

SCOTT, E. L. (1985). Neyman, Jerzy. *Encycl. Stat. Sci.* **6** 214–223.

SCOTT, E. L., SHANE, C. D. and WIRTANEN, C. A. (1954). Comparison of the synthetic and actual distribution of galaxies on a photographic plate. *Astrophys. J.* **119** 91–112.

SØRENSEN, M. (1997). Estimating functions for discretely observed diffusions: A review. *IMS Lecture Notes Monograph Series* **32** 305–325. IMS, Hayward, CA. MR1837812

STEWART, B. S., ANTONELIS, G. A., YOCHEM, P. K. and BAKER, J. D. (2006). Foraging biogeography of Hawaiian monk seals in the northwestern Hawaiian Islands. *Atoll Research Bulletin* **543** 131–145.

TAYLOR, J. R. (2005). *Classical Mechanics*. Univ. Science, Sausalito.

WISDOM, M. J. ed. (2005). *The Starkey Project.* Alliance Communications Group, Lawrence Kansas.

Part IV
Point Processes

Commentary by Peter Guttorp

This part will start with a couple of earth sciences papers. We then proceed via an influential methodology paper to describing some work in neurophysiology. In addition, a paper on latent variables and two on robustness of regression to misspecification of the regression function are discussed.

An empirical investigation of the Chandler wobble and two proposed excitation processes [1973]

David was not so keen on including this paper, since it does not solve the problem of what excites the Chandler wobble: the periodic deviation of Earth's pole from its mean direction. I like it, because it illustrates very clearly the power of frequency analysis tools. The wobble has two dominant frequencies, one corresponding to the annual cycle, and the other with the Chandler period of about 14.1 months.

Describing the pole of rotation of Earth at time t by $Z(t)$, a complex number with the real part corresponding to deviation towards Greenwich , the imaginary part the perpendicular direction, this has been modeled by the stochastic differential equation

$$dZ(t) = \alpha Z(t)dt + d\Phi(t)$$

where $\alpha = \beta + i\gamma,\ \beta > 0$. Assuming that the excitation function $\Phi(t)$ has stationary increments (see Brillinger 1972a for the analysis of such processes), it is easy to see that Z will inherit the peaks of Φ, plus a new peak at frequency γ with spread β. To model the spectrum, David looks at first differences of a seasonally adjusted series, and fits the corresponding model using exponential likelihood, approximately valid for Fourier frequencies of the periodogram.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_26,

A complex demodulation at the estimated Chandler frequency shows high amplitudes 1910-14 and very high 1948-55, with the phase lower for lower amplitude. To explain the Chandler wobble, two proposed excitation processes are earthquakes or atmospheric motion. Data on the times of large earthquakes, although forming a point process, can be demodulated similarly to the observed deviations. There is little similarity between these two demodulations, indicating that this is not the actual excitation, at least not by itself. To see if atmospheric motion (or more precisely its moment of inertia) could be the cause, demodulation shows again no similarity to the wobble, although the motion of course has a strong annual frequency.

Estimation of uncertainties in eigenspectra estimates from decaying geophysical time series [1979]

This is one of the papers David wrote with long-term collaborator and dear friend Bruce Bolt. They consider the fundamental seismological problem of estimating the eigenfrequencies of the earth. The response of the earth to an earthquake is essentially a linear combination of cosines at these frequencies with decaying parameters. In order to estimate simultaneously both frequencies and decay rates, they propose to do a complex demodulation (at preliminary estimates of the frequencies), with the phase diagram helping to assess both decay rates and amplitudes. The results yield estimates and standard errors of these parameters for the first time in the geophysical literature.

Statistical inference for stationary point processes [1975]

This is probably the paper of David's I have gone back to the most. Here he sets out some parameters for point processes on the line with points of k different types. For example, the kth order product density is

$$P(\text{type } a_j \text{ point in } (t_j, t_j + dt_j),\ j = 1 \ldots k) = p_{a_1 \cdots a_k}(t_1, ..., t_k) dt_1 \cdots dt_k$$

After working out the details for a couple of examples, he turns to the stationary case, where one can start looking at spectral properties, setting down cross spectra as the Fourier transform of the corresponding product density. One can use the spectral representation to compare point process to corresponding time series parameters (Brillinger 1978e).

David shows how to write down interesting models based on linear systems. Let

$$P(N_1(t,t+dt] = 1\,|N_2(u)) = (\mu + \int_{-\infty}^{\infty} a(t-u)dN_2(u))dt$$

This model lets the intensity of type 1 points near t be dependent on the process of type 2 points. If there is a causal connection we must have $a(u) = 0,\ u < 0$. The sign of a determines whether type 2 points tend to inhibit or excite points of type 1. When $a \equiv 0$, type 1 points are not dynamically influenced by type 2 points, so the series are casually unaffected.

The paper then turns to inference. Under a mixing condition (which sometimes can be hard to check) the natural histogram-type estimator for a kth order parameter will be asymptotically Poisson or normal, depending on which bandwidth conditions are used. Estimators at lags sufficiently separated are asymptotically independent. Spectral tools enable estimation of the function a in the linear system, and Whittle type approximate likelihood tools (using the periodogram) are presented.

Once you master the techniques in this paper, no analysis of point processes on the line will ever seem too daunting!

The identification of point process systems [1975]

The linear model which was used as an example in the previous paper comes to extensive use in this one. It was motivated by the study of systems of connected nerve cells. The tools in the previous paper: spectral analysis to identify the excitation function a, and Whittle likelihood to estimate parametric models, are applied (a rare enough phenomenon in Annals of Probability) to two neighboring nerve cells of a sea slug, with neurophysiological and causal interpretations.

Measuring the association of point processes, a case history [1976]

The lovely pedagogic structure of this paper is as follows: here is a problem (testing the association of two neural spike trains). Here is a model (stationary point processes on the line), and a relevant parameter $(p_{NM}(u)/p_N p_M)$. Next an estimate, which asymptotically is Poisson, so a square root transformation should give approximate normality and constant variance in u. Finally, a plot of $\sqrt{\hat{p}_{NM}(u)/\hat{p}_N\hat{p}_M}$ with associated confidence band shows quite clearly the inhibitory effect of one spike train on the other.

Empirical examination of the threshold model of neuron firing [1979]

José Segundo is the other main scientific collaborator (as well as another dear friend) of David's. This paper models the idea that a nerve cell fires when a membrane potential $U(t)$ upcrosses a threshold $\theta(t)$. $U(t)$ is modeled as a linear functional of the input current, and the goal is to estimate the kernel of this functional, the average threshold, and the probability of firing as a function of U from data on input currents and observed firings. Tools include likelihood, spectral and correlational analysis, all of which require different simplifying assumptions but give similar results. There is an indication that an input of uniform white noise is better than Gaussian white noise at estimating the firing probability.

Nerve cell spike train data analysis: a progression of techniques [1992]

The Fisher lecture, given at the Joint Statistical Meetings in Atlanta in 1991, is a typical Brillinger production: a tour de force through neurophysiology, point process theory, an extended nonparametric analysis from that in the previous paper, causal analysis and coherence in a system of three neurons, and finally some tools to study a network of eight nerve cells. As always in David's papers, the right picture is used to illustrate the points.

A generalized linear model with "Gaussian" regressor variables [1982]
The identification of a particular nonlinear time series systems [1977]

These two papers were really the beginning of what today is called model robustness. The basic tool is a simple identity for bivariate normal variables V and U and a real function G:

$$Cov(G(U), V) = Cov(G(U), U)Cov(U, V)/Var(V)$$

This now can be used in regression, to show that regression coefficients *even when the regression is a nonlinear function* of a linear combination of regressors are proportional to the coefficients of the linear combination, at least as long as the regressors are normal or "behave like" normals. The second paper extends the result to the time series case, where a nonlinear function of an autoregressive function again allows you to estimate the relative sizes of the autoregressive coefficients. David tells me that when he told Jerzy Neyman and Erich Lehmann about this result, they did not believe him!

An Empirical Investigation of the Chandler Wobble and Two Proposed Excitation Processes

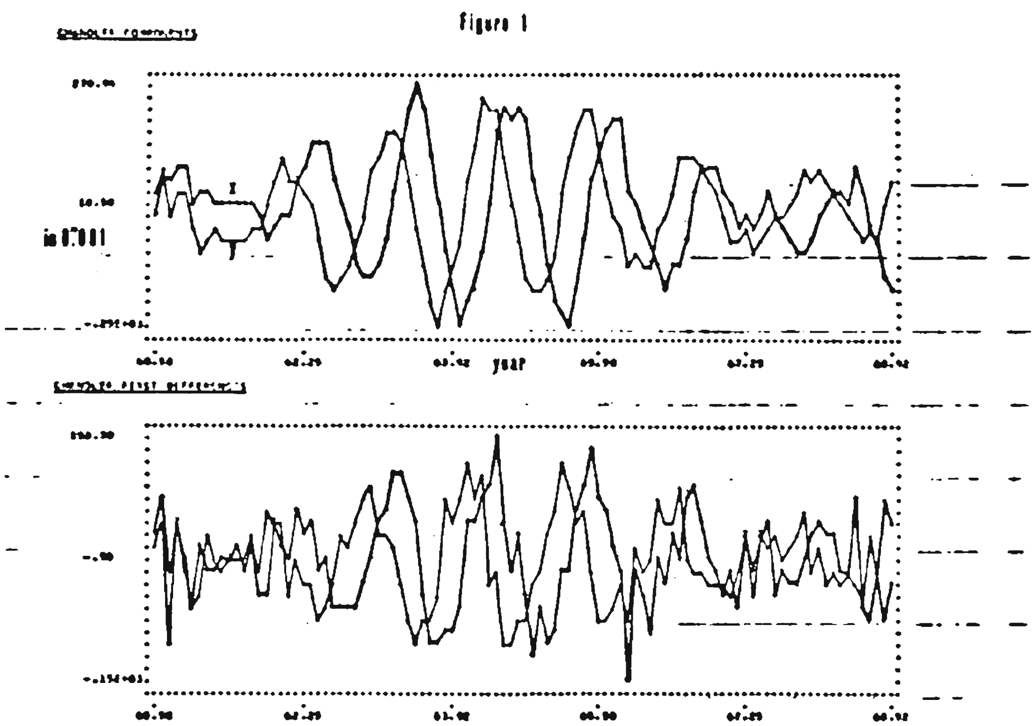
By

David R. Brillinger, Berkeley, U.S.A.

1. Introduction

The axis of instantaneous rotation of the Earth does not remain fixed relative to the body of the Earth, rather, its points of interception with the surface wander about within a region approximately the size of a tennis-court. This wandering was predicted by Euler in 1765 and confirmed by observation in 1891. The top graph of Figure 1 provides the x and y coordinates of the deviation of the North pole from its mean position for the period 1960–1969. (In units of 0″.001 = .101 ft.) The motion of the pole produces a variation in the latitude which may be used to deduce the time path of the pole. We mention briefly how this is done.

The zenith is the direction opposite to local gravity. The altitude of a star is the complement of its zenith distance. The fundamental method of determining the latitude of an observatory is to take the average of the altitudes of a circumpolar star when it crosses the meridian above and below the pole. Since 1899 the International Latitude Service has measured the variation of



P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_27, © Springer Science+Business Media, LLC 2012

latitude at five stations apread along 39°08′ north latitude. A conventional pole of rotation (the C.I.O.) has been adopted. Suppose $X(t)$ denotes the displacement of the instantaneous north pole at time t from the C.I.O. towards Greenwich and $Y(t)$ the displacement towards 90° west of Greenwich. Let $\Delta\varphi_j(t)$ denote the increment in latitude at observatory j, from its mean latitude. Then estimates $x(t)$, $y(t)$ of $X(t)$, $Y(t)$ are determined by the least squares fit of the regression equation

$$\Delta\varphi_j(t) = Z(t) + X(t)\cos\lambda_j + Y(t)\sin\lambda_j + \varepsilon_j(t) \tag{1.1}$$

$j = 1, \ldots, 5$ where λ_j denotes the longitude of the j-th observatory. For t at monthly intervals, these values are given in Vicente and Yumi (1969, 1970), which is the source of the data used in the computations of this paper. The values of $x(t)$ and $y(t)$ fall in the intervals $-0''.37$, $0''.47$ and $-0''.28$, $0''.50$ respectively. The probable errors given in Table 12 of Yumi, Ishii and Sato (1968) may be used to deduce the standard errors of $x(t)$, $y(t)$ from the above linear fit. These are $0''.057$ and $0''.048$ respectively.

Chandler (1891) suggested that the polar motion was made up of two principal components with periods one year and 428 days $\doteq$ 14 months respectively. Figure 3 below gives the logarithm of the periodogram of the data. Two peaks, at frequencies near these periods are apparent. In the next section we shall set down a differential equation that describes the motion of the pole when the Earth is subjected to arbitrary excitations. Scientific workers seem to be agreed that the component of the motion with period one year results from the excitation function possessing a strong seasonal component. 428 days corresponds to the Euler frequency of vibration of the Earth; however the source of the energy that stimulates the natural vibration is not agreed upon. We shall consider earthquakes and shifts of the mass of the atmosphere as possible sources of the energy. This 428 day component is called the Chandler component. The associated motion of the Earth is called the Chandler wobble.

In the next section we present a variety of harmonic analyses of the polar variation including; power spectrum estimation, maximum likelihood fit of a model of the spectrum, bispectrum estimation and complex demodulation. In Section 3 we carry out cross-spectrum analysis of the polar motion series with two earthquake series as well as complex demodulation of the latter. In Section 4 we repeat this analysis with an atmospheric series.

Munk and MacDonald (1960) is an excellent source of basic material concerning the rotation of the Earth. The proceedings of two symposia on the topic have appeared. These are Mansinha, Smylie and Beck (1970) and Melchior and Yumi (1972). These works show that the problem of understanding the rotation of the Earth is exceedingly rich in geophysical terms. It is also rich in statistical aspects. We mention the papers; Walker and Young (1955, 1957), Arato, Kolmogorov and Sinai (1962), Mandlebroit and McCamy (1970).

2. Analyses of the Polar Motion

The position of the pole of rotation at time t is conveniently described by the complex number

$$Z(t) = X(t) + i\,Y(t) \tag{2.1}$$

where $X(t)$, $Y(t)$ are the displacements from the C.I.O. towards Greenwich and towards 90° west of Greenwich respectively. Munk and MacDonald have investigated the dynamics of the spinning Earth. Let $\Phi(t)$ denote an excitation function whose increments, $d\Phi(t)$, describe the change in the Earth's inertia tensor in the time interval $(t, t + dt) \cdot d\Phi(t)$ is complex-valued with Re $d\Phi(t)$ giving the change towards Greenwich and Im $d\Phi(t)$ the change towards 90° west of Greenwich. [In the next section we shall make use of a formula for $d\Phi(t)$ when the change results from a shift of mass in an earthquake.] From classical mechanics, Munk and MacDonald deduce the equation of motion

$$dZ(t) = a\,Z(t)\,dt + d\Phi(t) \tag{2.2}$$

with $a = -\beta + i\gamma$ complex-valued and $\beta > 0$. If $\Phi(t) = 0$, then a solution of (2.2) is provided by

$$Z(t) = e^{at} = e^{-\beta t}(\cos\gamma t + i\sin\gamma t) \tag{2.3}$$

This motion is one of a damped oscillation of requency γ. The greater β, the greater will be the damping.

Suppose now that $\Phi(t)$, $-\infty < t < \infty$, is a random process with stationary increments and power spectrum $f_{\Phi\Phi}(\lambda)$. [See Brillinger (1970) for a discussion of the spectral analysis of processes with stationary increments. The definitions given there must be modified trivially to apply to complex-valued processes.] Then (2.2) will have a solution with stationary increments and power spectrum

$$f_{ZZ}(\lambda) = |\,i\lambda - a\,|^{-2} f_{\Phi\Phi}(\lambda) = [\beta^2 + (\lambda - \gamma)^2]^{-1} f_{\Phi\Phi}(\lambda) \tag{2.4}$$

This expression shows that $f_{ZZ}(\lambda)$ may be expected to inherit the peaks of $f_{\Phi\Phi}(\lambda)$ and to possess a new peak, of spread β, at $\lambda = \gamma$.

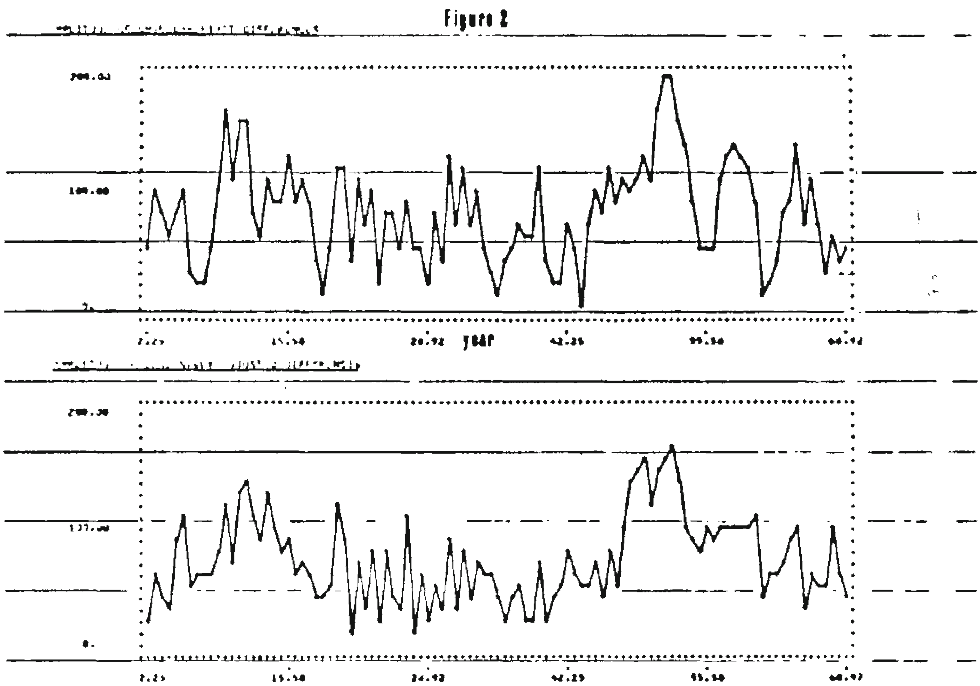
Were $Z(t)$ available for an interval $0 < t \leqslant T$, we would be led to base a spectral analysis of it on the Fourier-Stieltjes transform

$$\int_0^T \exp\{-i\lambda t\}\, dZ(t) \tag{2.5}$$

The polar motion values we use are given at monthly intervals, and so we are led to take as basic statistic the finite Fourier transform of the first differences of $Z(t)$, namely

$$d_Z^{T}(\lambda) = \sum_{t=0}^{T-1} \exp\{-i\lambda t\}\,[(Zt + 1) - Z(t)] \tag{2.6}$$

$-\infty < \lambda < \infty$. The second graph of Figure 1 is a plot of the series $Z(t + 1) - Z(t)$ for the time period 1960–1969. The first graph of Figure 2 is a plot of $|\,Z(t + 1) - Z(t)\,|$ for the period 1902–1969.



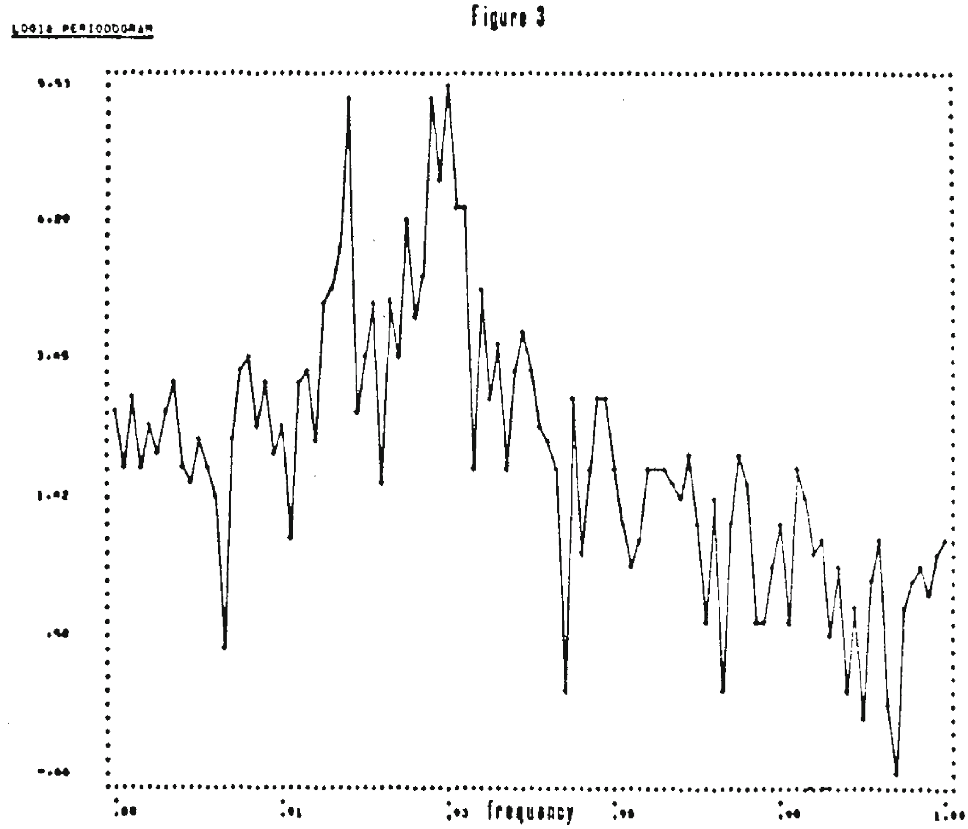
In Figure 3 we have plotted $\log_{10}$ of the periodogram $I_{ZZ}^{(T)}(\lambda) = (2\pi T)^{-1} | d_Z^{(T)}(\lambda) |^2$ for $.88 \leq \lambda/2\pi \leq 1.00$. $I_{ZZ}^{(T)}(\lambda)$ may be considered to be a highly unstable estimate of $f_{ZZ}(\lambda)$. In the case that the process $\Phi(t)$, $-\infty < t < \infty$, is mixing, the periodogram will be asymptotically exponential with mean $f_{ZZ}(\lambda)$. The standard deviation of the curve in Figure 3 will be approximately .43. Peaks are present in this graph at frequencies $\lambda/2\pi = .917$, .929 corresponding to rotations in a negative direction with periods $\doteq$ 12 months, 14.1 months respectively. It has long been understood that the process $\Phi(t)$, $-\infty < t < \infty$, would contain a strong component of period 12 months because of the seasonal variation of the loading of the Earth through, shifts of the atmosphere, melting of snow, tides and the like. [See Jefferys (1959).] This would account for the peak in Figure 3 corresponding to a period of 12 months. Before smoothing the periodogram in order to obtain a more stable estimate of the power spectrum, we therefore removed the seasonal variation from the series of first differences by subtracting monthly means. The values subtracted are given in Table 1. They correspond to a figure of ellipsoidal shape. Figure 4 is $\log_{10}$ of the spectral estimate obtained by

Table 1

(units of 0″.001)

	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
x	−41	−17	−2	22	33	43	49	31	−2	−34	−40	−42
y	11	28	43	34	22	7	−12	−35	−40	−45	−19	5

Figure 3



smoothing 8 adjacent periodogram ordinates based on the seasonally corrected values. The bandwidth of this estimate is .01 cycles/month. Its asymptotic standard error is .15.

The smooth curve in Figure 4 corresponds to a fitted model whose construction we now describe. Suppose that we denote the seasonally corrected version of $\boldsymbol{Z}(t)$, $\boldsymbol{\Phi}(t)$ by $\boldsymbol{Z}'(t)$, $\boldsymbol{\Phi}'(t)$ respectively. The second graph of Figure 2 is a plot of $|\boldsymbol{Z}'(t+1) - \boldsymbol{Z}'(t)|$, the amplitude of the seasonally adjusted first differences. It is seen to peak around the years 1910 and 1950. The model (2.2) retains the form

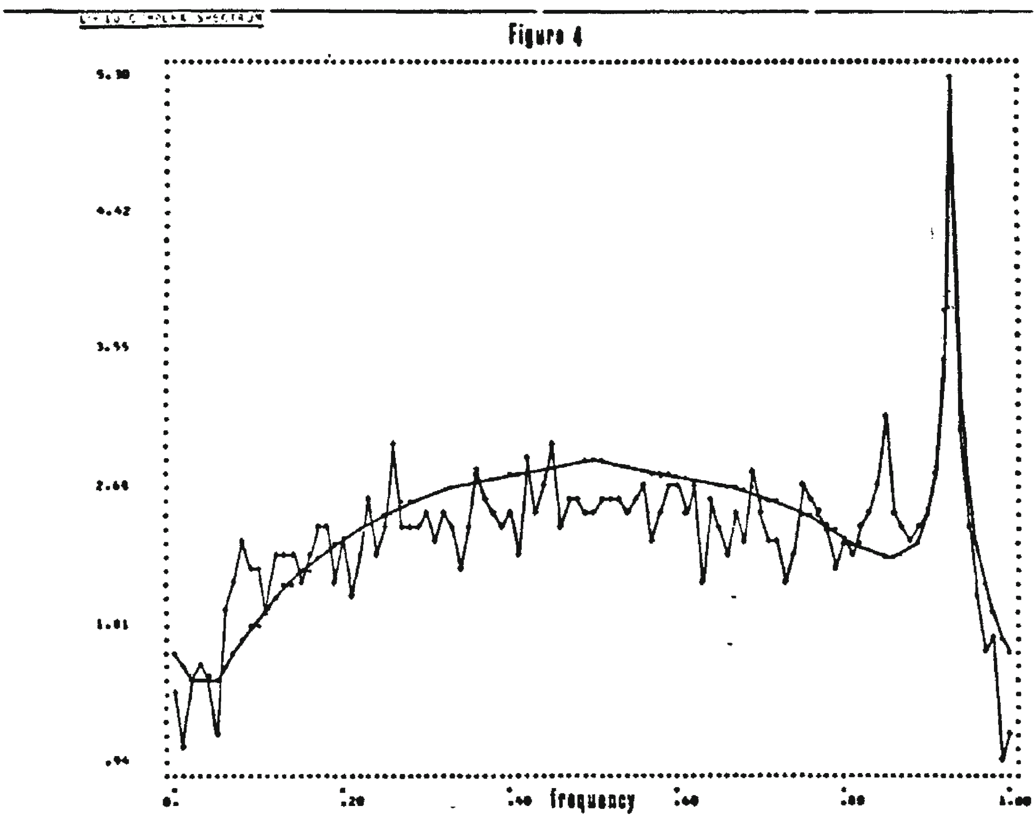
$$d\boldsymbol{Z}'(t) = a\,\boldsymbol{Z}'(t) + d\boldsymbol{\Phi}'(t) \tag{2.7}$$

We may solve the equation (2.7) and obtain

$$\boldsymbol{Z}'(t) = \int_{-\infty}^{t} e^{a(t-u)}\, d\boldsymbol{\Phi}'(u) \tag{2.8}$$

Noting the assumed removal of seasonal components from $\boldsymbol{\Phi}(t)$, we now assume that $\boldsymbol{\Phi}'(t)$ is a noise process with stationary orthogonal increments and var $\{d\Phi'(t)\} = \sigma^2$. [Were we to assume it Gaussian as well, then (2.7) would be the model of Arato et al. (1962).] Consider the series of increments

$$\Delta\boldsymbol{Z}'(t) = \boldsymbol{Z}'(t+1) - \boldsymbol{Z}'(t) \tag{2.9}$$



$t = 0, \pm 1, \ldots$. We see from the representation (2.8) that this series has autocovariance function

$$c_{\Delta Z', \Delta Z'}(u) = \sigma^2 \exp\{-\beta \mid u \mid\} \exp\{i \gamma u\}/2\beta \qquad (2.10)$$

$u = 0, \pm 1, \ldots$ and hence power spectrum

$$f_{\Delta Z', \Delta Z'}(\lambda) =$$

$$= \frac{\sigma^2}{2\pi} \frac{1 - \exp\{-2\beta\}}{2\beta} \frac{1}{1 - 2\exp\{-\beta\}\cos(\lambda - \gamma) + \exp\{-2\beta\}}$$

for $-\infty < \lambda < \infty$.

We have mentioned previously that the series $Z(t)$ is not observed directly, but is measured subject to error. Let

$$z'(t) = Z'(t) + \varepsilon(t) \qquad (2.11)$$

denote the observed series, corrected for seasonal effects, where we assume that $\varepsilon(t)$, $t = 0, \pm 1, \ldots$ is a stationary noise series with var $\varepsilon(t) = \psi^2$. It follows that the power spectrum of the first differences of $z'(t)$ will be given by

$$f_{\Delta z', \Delta z'}(\lambda) = f_{\Delta Z', \Delta Z'}(\lambda) + \frac{\psi^2}{2\pi} \mid 1 - e^{-i\lambda} \mid^2 \qquad (2.12)$$

This last constitutes our proposed model for the spectrum of Figure 4. It is seen to involve four unknown parameters; γ the frequency of the Chandler wobble, β the damping constant, σ the standard deviation of the seasonally corrected excitation function and ψ the standard deviation of the measurement errors.

We fit the model by the method of maximum likelihood. Set

$$\hat{f}_s = I^{(T)}_{\Delta\tau', \Delta\tau'}\left(\frac{2\pi s}{T}\right), f_s = f_{\Delta\tau', \Delta\tau'}\left(\frac{2\pi s}{T}\right)$$

for $s = 0, \ldots, T-1$. Under a variety of conditions, the variates $\hat{f}_s/f_s$, $s = 0, \ldots, T-1$ are approximately independent standard exponentials. The likelihood function of the data therefore has the approximate form

$$L = \prod_s f_s^{-1} \exp\{-\hat{f}_s/f_s\}$$

Let θ, θ' denote any two of the parameters, then

$$\frac{\partial \log L}{\delta \theta} = -\sum_s \frac{(f_s - \hat{f}_s)}{f_s^2} \frac{\partial f_s}{\partial \theta} \tag{2.13}$$

$$E\left[\frac{\partial \log L}{\partial \theta} \frac{\partial \log L}{\partial \theta'}\right] = \sum f_s^{-2} \frac{\partial f_s}{\partial \theta} \frac{\partial f_s}{\partial \theta'}$$

The maximum likelihood equations are obtained by setting (2.13) equal to 0 for the various parameters. We solve these equations by the method of scoring [see Rao (1965), p. 302]. This procedure has the advantage of producing estimates of the asymptotic standard errors incidentally. We began the recursion with estimates determined by the method of moments. The procedure stabilised after two rounds. We obtained the following results.

Table 2

parameter	$\gamma/2\pi$	β	σ	ψ
estimate	.9294	.0050	7.1	31.9
s. e.	.0026	.0023	.33	.62

The indicated results for γ lead to a 95 per cent confidence interval for the Chandler period to be from 13.2 months to 15.3 months. A 95 per cent confidence interval for β is from .0005 month^{-1} to .0095 month^{-1}. It has not been determined accurately at all. The estimate of σ is important in searching for the source of the excitation of the whobble. It suggests that the non-seasonal fluctuations of the excitation have standard deviation of order 0''.007 for monthly values. It is interesting to compare the magnitude of the standard deviation of the observational errors as estimated here with the values 0″.057, 0″.048 mentioned in the introduction. In the present notation they correspond to $\psi = 0''.075$ a value larger than the 0″.032 found here. In either case the observational errors are large compared to the magnitude of the phenomenon under study.

In Figure 4 we have plotted expression (2.12) using the parameter values of Table 2. The fit seems consistent with the standard error .15 of the estimate except for the peak just to the left of the Chandler peak. Surprisingly, this peak is centered at a frequency, (.846) (2π) that is near the sum of the seasonal and Chandler frequencies. This occurrence led us to suspect the presence of a non-linear phenomenon. We therefore estimated the modulus of the bicoherency

$$\frac{|f_{\Delta z\Delta z,\Delta z}(\lambda_1, \lambda_2)|}{\sqrt{f_{\Delta z,\Delta z}(\lambda_1) f_{\Delta z,\Delta z}(\lambda_2) f_{\Delta z,\Delta z}(\lambda_1 + \lambda_2)}} \tag{2.14}$$

(See the Appendix for the definition of the third-order spectrum appearing here.) Table 3 below presents an estimate for frequencies in the immediate neighborhood of the seasonal, the Chandler and the seasonal plus the Chandler. The bandwidth of the estimate is .01. In the null case the square of the estimate is distributed asymptotically as an exponential with mean $T/2\pi N$, if N denotes the number of third-order periodograms averaged in forming the estimate. [The sampling properties of such estimates are discussed in Brillinger and Rosenblatt (1967) and Huber et al. (1970).] The 99 per cent point for the values of Table 3 is 2.52, corresponding to $N = 95$. There is a clear suggestion that the values of Table 3 are larger than would be expected in the null case. No dramatic peaks are present in the table however. Our conclusion is that the excitation process or the measurement error process, is not quite normal. Table 4 presents an estimate of (2.14) covering the whole frequency domain. Here the bandwidth adopted was .05 and $N = 1950$. The 99 per cent point of the null distribution is now .56. Again there are no dramatic peaks in the function, rather the whole collection of values is larger than would be expected in the null case. It appears that the data are somewhat non-normal.

Table 3

frequency/2π

	.87	.88	.89	.90	.91	.92	.93	.94	.95	.96
.87	2.5	.7	.2	.7	.9	1.2	1.6	1.6	.9	1.5
.88		1.1	1.1	.3	2.1	.4	.9	1.3	.9	.2
.89			2.3	.7	.6	1.3	1.8	.7	.9	.5
.90				1.1	.8	.4	2.6	1.1	2.2	1.2
.91					.5	1.4	1.3	.9	2.2	.5
.92						2.9	2.5	2.7	.1	1.9
.93							.8	.2	1.5	.3
.94								1.3	.6	.5
.95									2.5	2.3
.96										.7

Table 4

frequency/2π

	.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
.00	1.9	.7	.8	.3	.4	.5	.3	.6	.5	.5	.9	.8	.7	.7	.6	.5	.4	.3	.4	.5
.05		.2	.2	.3	.3	.3	.3	.5	.1	.2	.2	.4	.4	.2	.2	.4	.3	.5	.1	.1
.10			.5	.2	.3	.0	.4	.3	.2	.2	.1	.2	.3	.2	.2	.6	.2	.5	.2	.3
.15				.3	.4	.1	.3	.3	.1	.2	.6	.3	.3	.6	.3	.4	.1	.2	.7	.3
.20					.6	.3	.3	.1	.2	.1	.2	.1	.4	.3	.1	.2	.4	.3	.1	.4
.25						.3	.3	.3	.2	.2	.2	.5	.1	.6	.3	.1	.1	.3	.1	.3
.30							.3	.2	.4	.3	.2	.2	.2	.3	.1	.3	.3	.2	.2	.2
.35								.6	.2	.3	.1	.3	.2	.6	.4	.1	.2	.4	.4	.2
.40									.8	.6	.5	.0	.2	.5	.4	.1	.1	.4	.3	.1
.45										.1	.1	.3	.4	.3	.4	.6	.3	.2	.2	.4
.50											.8	.2	.2	.3	.3	.0	.4	.3	.7	.3
.55												.3	.4	.5	.3	.2	.5	.2	.5	.3
.60													.7	.5	.1	.2	.1	.2	.2	.3
.65														.4	.2	.2	.1	.1	.1	.2
.70															.1	.5	.4	.5	.3	.2
.75																.4	.1	.3	.3	.5
.80																	.2	.0	.4	.2
.85																		.3	.4	.4
.90																			1.1	.8
.95																				.0

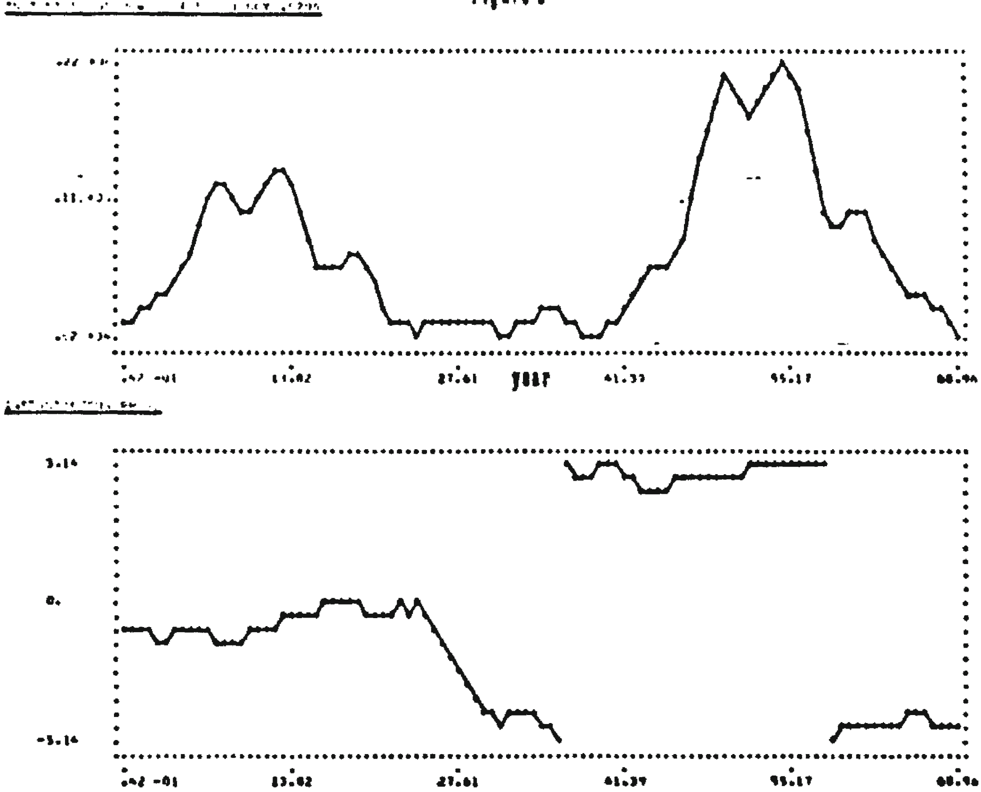
In order to be able to better understand the behavior of the polar motion and in order to get an idea of the character of the excitation process, $\Phi(t)$, we carried out a complex demodulation of the series $\Delta z(t)$ at several frequencies. [This procedure is described for real-valued series in Tukey (1961).] Specifically we formed the series

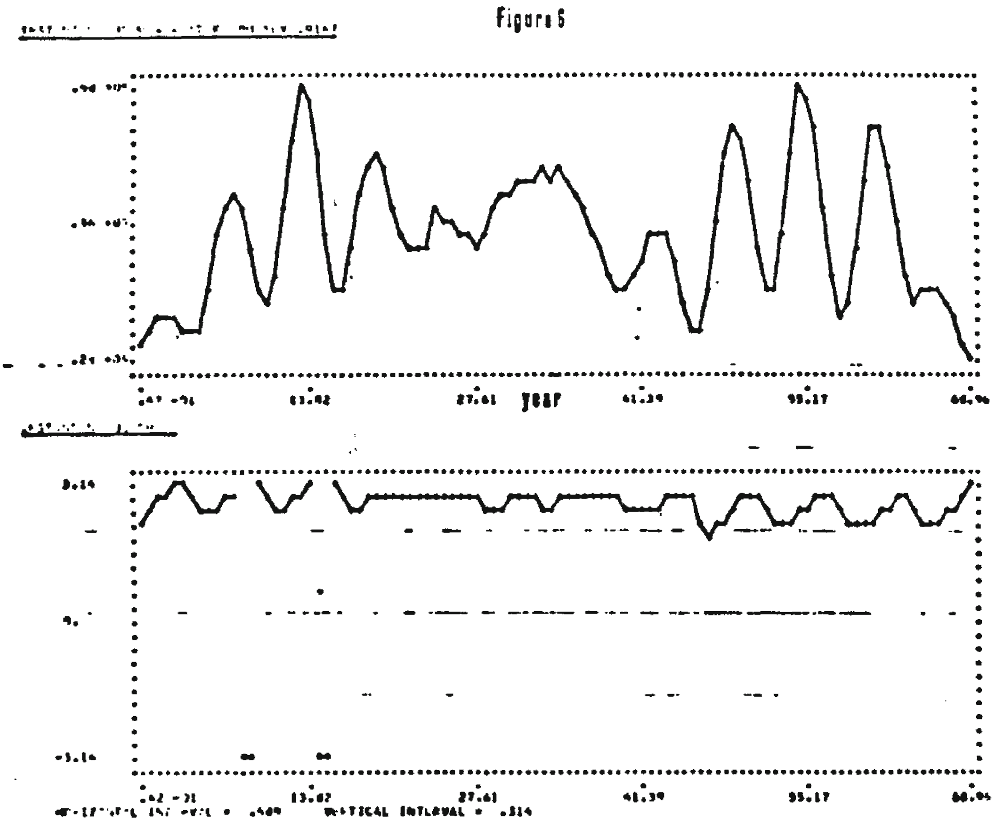
$$\zeta_{\Delta z}(t, \lambda) = \frac{1}{\sqrt{2\pi(2L+1)}} \sum_{|t-u| \leq L} \Delta z(u) \exp\{-iu\lambda\} \tag{2.15}$$

$t = 0, 1, \ldots, T-1$ for a variety of λ and $L = 48$. We note that $|\zeta_{\Delta z}(t, \lambda)|^2$ $t = 0, \ldots, T-1$, is a running periodogram for the data at frequency λ. Its average across the whole time domain would provide an alternate estimate of $f_{\Delta z, \Delta z}(\lambda)$. Variations in it are indicative of temporal variations in the power at frequency λ. The time path of $\arg \zeta_{\Delta z}(t, \lambda)$ gives information concerning the value of the dominant frequency component in the neighborhood of λ. If its path is a straight line with slope ν for some time period, then the component with frequency $\lambda + \nu$ is dominant in that time period.

Figure 5 is a plot of $|\zeta_{\Delta z'}(t, \lambda)|^2$, $\arg \zeta_{\Delta z'}(t, \lambda)$, $t = 0, \ldots, T-1$ for $\lambda/2\pi = .9294$, the Chandler frequency. The Chandler component is seen to be strong in the period 1910—1914 and very strong in the period 1948—1955.

Figure 5



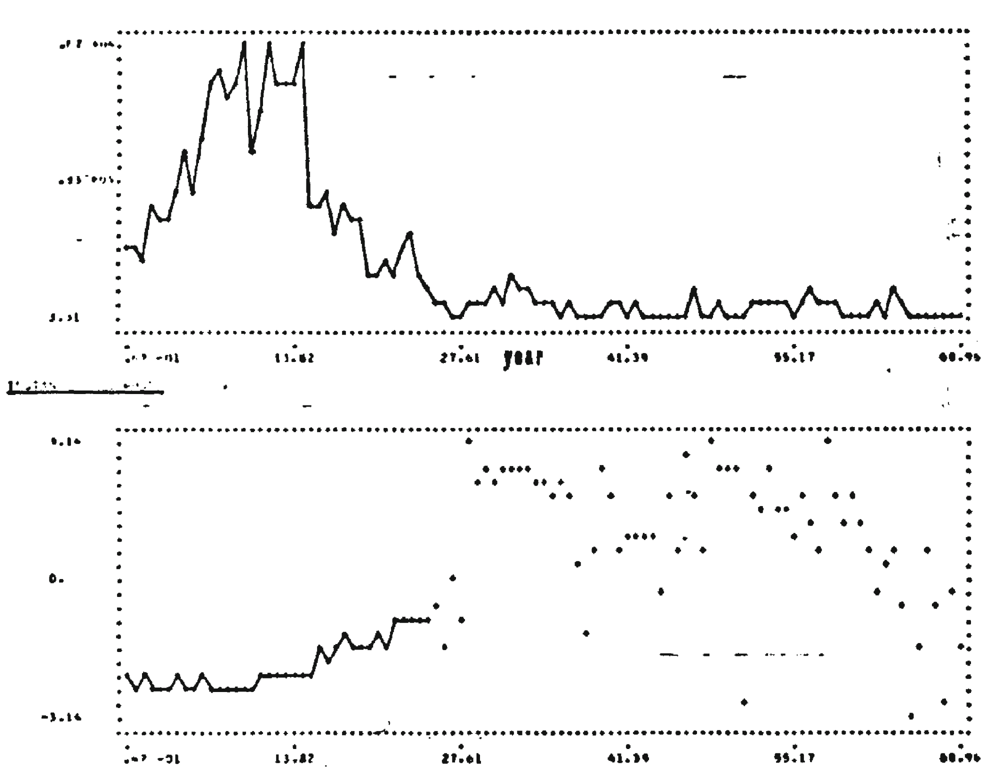


The second graph suggests that the Chandler frequency decreased in the period 1925–1940 when the power was low. Figure 6 is a plot of $|\zeta_{\Delta x}(t,\lambda)|^2$ and $\arg \zeta_{\Delta x}(t,\lambda)$ for $\lambda/2\pi = .9167$, the annual component. The power of the seasonal component is seen to be reasonably stationary across the whole period. Figure 7 is especially interesting. It is a plot of $|\zeta_{\Delta x'}(t,\lambda)|^2$, $\arg \zeta_{\Delta x'}(t,\lambda)$ for $\lambda/2\pi = .8460$, corresponding to the Chandler frequency plus the seasonal. The component at this frequency is seen to be present effectively only for the period 1905–1914. I have no explanation for this behavior. Perhaps it is due to a fault in the data processing for the early years. A perturbation analysis of the equations of motion leading to (2.2) suggests that non-linearities would lead to the appearance of harmonics of the seasonal, but not the appearance of a seasonal plus Chandler frequency component.

We now turn to the problem of discerning, if possible, the character of the excitation process $\Phi(t)$. An examination of expression (2.12), (with the parameter values of Table 2), as plotted in Figure 4 suggests that the true process $Z'(t)$ dominates only for frequencies in the immediate neighborhood of the Chandler frequency. It follows that we will have

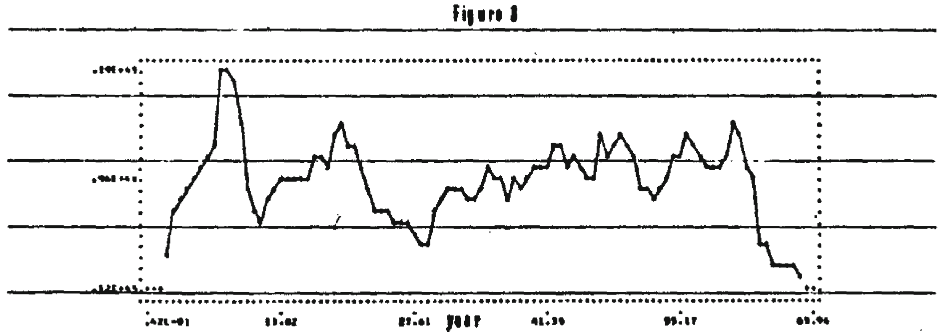
$$\begin{aligned} \zeta_{\Delta x'}(t,\lambda) &\doteq \zeta_{\Delta z'}(t,\lambda) \\ &\doteq (i\lambda - a)^{-1}\,\zeta_{\Delta\Phi'}(t,\lambda) \end{aligned} \tag{2.16}$$

Figure 7



for, and only for, $\lambda/2\pi = .9294$. Figure 5 is therefore especially important in the search for the process, $\Phi(t)$, exciting the Chandler wobble. The instantaneous power of this process, at frequency .9294, must have behaved in the manner of the top graph of Figure 5, namely been high for the period 1910–1914 and very high for the period 1948–1955. I do not know of a process that has behaved in this manner. I would appreciate suggestions that anyone has. In the next sections I examine two processes that have been proposed.

Figure 8

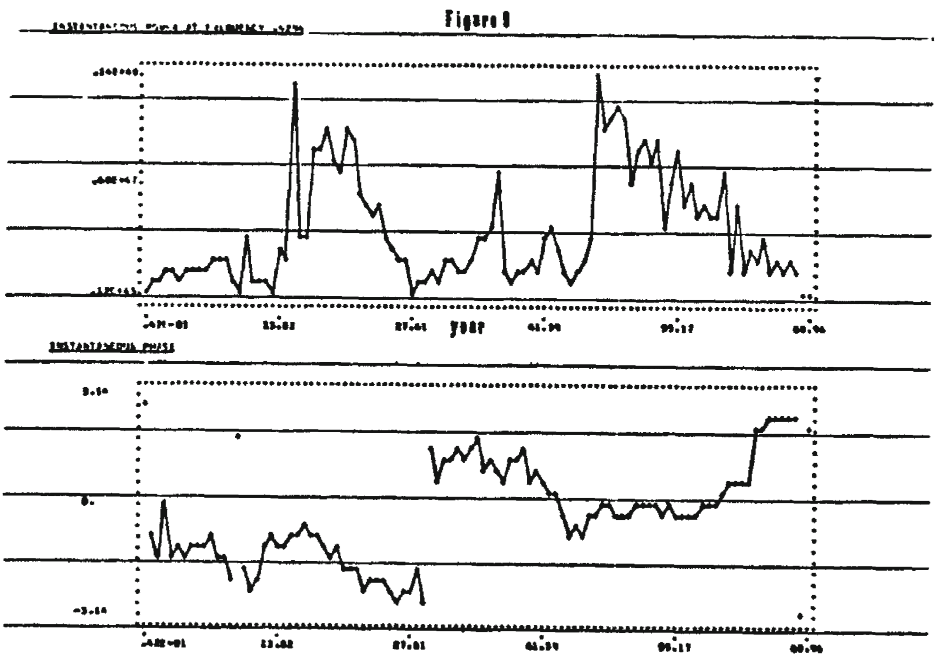


3. Excitation by Earthquakes

Earthquakes were proposed at an early time as a cause of the Chandler wobble [see Cecchini (1928)]. A number of serious investigations of this possibility have been carried out recently [see Mansinha et al. (1970), Dahlen (1971, 1972) for example]. We begin this section with an examination of a series of monthly earthquake energy. We computed such a series from values given in Dubourdieu (1972), which were based on the data in Duda (1965). The series is one of earthquake energy (in ergs) released per month by earthquakes of magnitude $\geqslant 7.0$ throughout the world in the period 1904–1965. Figure 8 is a plot of the square of an 8 year running average of this series. We notice that the energy released was greatest in the period 1904–1910.

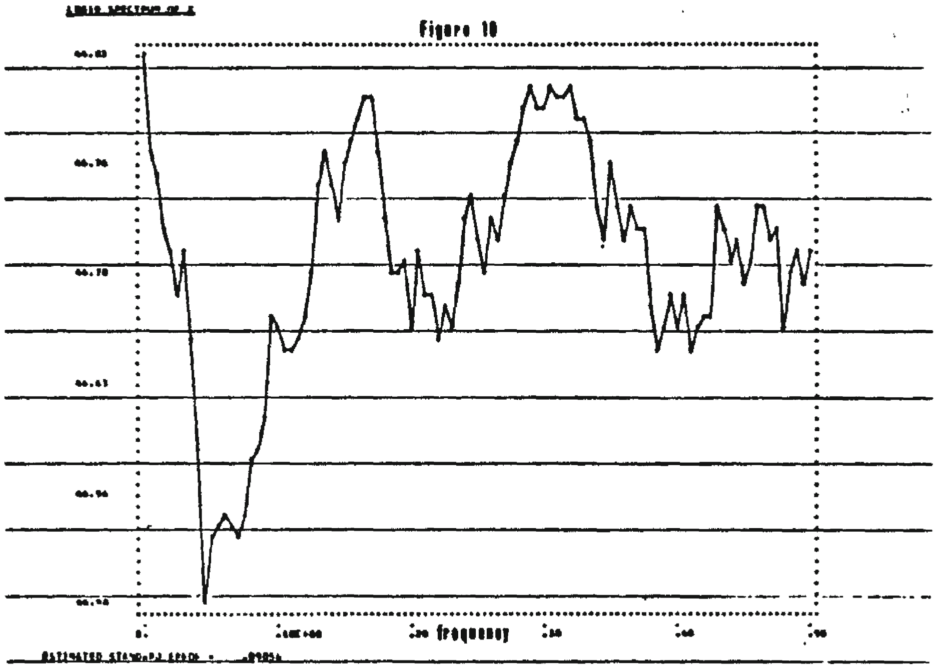
Figure 9 gives $|\zeta(t, \lambda)|^2$, $\arg \zeta(t, \lambda)$, $t = 0, \ldots, T-1$ for this series with λ corresponding to the Chandler frequency. The instantaneous power at this frequency is seen to be greatest for the periods 1914–1925, 1946–1954. A comparison of this plot with the top graph of Figure 5 suggests that the two do not match too well with respect to either shape or timing. Figure 10 is a graph of $\log_{10}$ of the estimated power spectrum of this series. The approximate standard error of this estimate is .09, suggesting that the population spectrum is not far from constant. The bandwidth is .05.

Clearly the effect any earthquake has on the motion of the pole will depend on the location of the earthquake within the Earth and on the direction of its movement. The above analysis takes no note of this dependence. Dahlen (1971, 1972) has developed expressions for the change in the Earth's inertia



tensor as a function of an earthquake's latitude, longitude, depth, strike, dip, slip and magnitude. Let θ_j denote all of these parameters for the j-th earthquake. Let τ_j denote the time of occurrence of the j-th earthquake. Dahlen (1972) has derived an expression for $\boldsymbol{C}(\theta)$, the change in the Earth's inertia tensor for an earthquake with parameter θ. The excitation function $\boldsymbol{\Phi}'(t)$ may now be written

$$\boldsymbol{\Phi}'(t) - \boldsymbol{\Phi}'(s) = \sum_{s < \tau_j \leq t} \boldsymbol{C}(\theta_j) \tag{3.1}$$



The model (2.11) therefore takes the form

$$z'(t) = \boldsymbol{Z}'(t) + \varepsilon(t) \tag{3.2}$$

with

$$d\boldsymbol{Z}'(t) = a\,\boldsymbol{Z}'(t) + d\boldsymbol{\Phi}'(t) \tag{3.3}$$

and $d\boldsymbol{\Phi}'(t)$ given. This is a model of a linear causal relationship involving two observed processes, $z'(t)$, $\boldsymbol{\Phi}'(t)$. If we assume that these processes have stationary increments, then we can carry out a frequency domain analysis of the processes in the manner of Section 4 of Brillinger (1970).

A difficulty presents itself at the beginning of the analysis. Not all of the components of θ are available for most of the earthquakes. After reading Dahlen (1971) and discussions with Professor B. A. Bolt, Professor T. V. McEvilly and W. Peppin, I have approached this difficulty as follows. Earthquakes tend to occur in belts along the edges of great plates on the Earth's surface [see Calder (1972)]. I constructed 11 strips at plate to plate boundaries within which the majority of the events occurred (166 out of 187) using the map of Chase (1972). (I took as basic data the earthquakes of magnitude $\geqslant 7.9$ occurring in the period 1900–1971.) For the unknown components of θ, I then took the parameters suggested by the direction and overall motion of the plates. I read average strike angles from the map of Chase (1972). I used the dip angles of Davies and Brune (1971). I assigned slip angles by assuming that the oceanic plates were plunging at a 45° angle under the continental plates at the oceanic transform faults. The remaining 19 events had unknown parameters assigned at random.

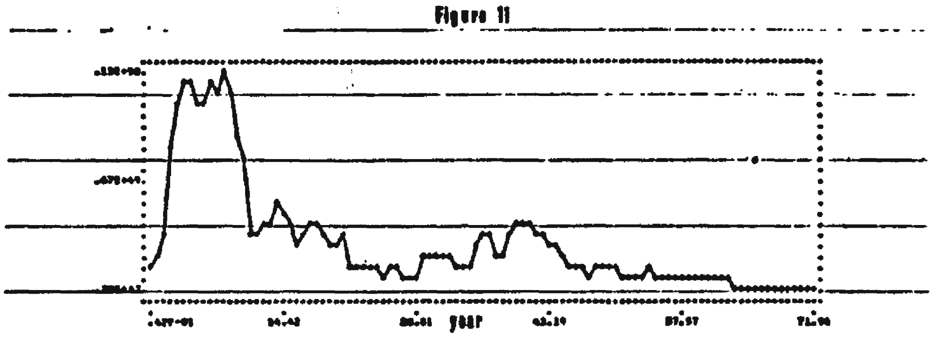


Figure 11 is a plot of the square of an 8 year running average of the energy released by earthquakes of magnitude $\geqslant 7.9$ for the period 1900–1971. Notice the large value of this function for the period 1904–1911. Figure 12 is an estimate of the $\log_{10}$ spectrum of the series of times of these events. The horizontal line is an estimate of the asymptote of the curve. The approximate standard error of the estimate is .072. The graph is suggestive that the corresponding population curve is near constant. (It would be constant for a stationary Poisson process.)

Figure 13 is a plot of the complex demodulate at the Chandler frequency of the process $\Phi'(t)$. Because of the point process with ancillary variate character of $\Phi'(t)$ we compute the demodulate differently from (2.15). We compute instead

$$\zeta_{\Phi'}(t, \lambda) = \frac{1}{\sqrt{2\pi(2L+1)}} \sum_{|t-\tau_j|<L} C(\theta_j) \exp\{-i\lambda\tau_j\} \qquad (3.4)$$

Figure 13 has no semblance with Figure 5, as it should were the earthquakes exciting the pole. Figure 14 is an estimate of the spectrum of the process $\Phi'(t)$. The bandwidth is .05 and approximate standard error .086. The Figure suggests that the population spectrum is near constant. The average level suggests a monthly excitation standard error of .74, far below the value 7.1 found from the polar coordinate data in Section 2. Figure 15 gives an estimate of the coherence between $z'(t)$ and $\Phi'(t)$ computed in the manner of Brillinger (1970). Notice the low level of the curve. The 95 per cent point of the approx-

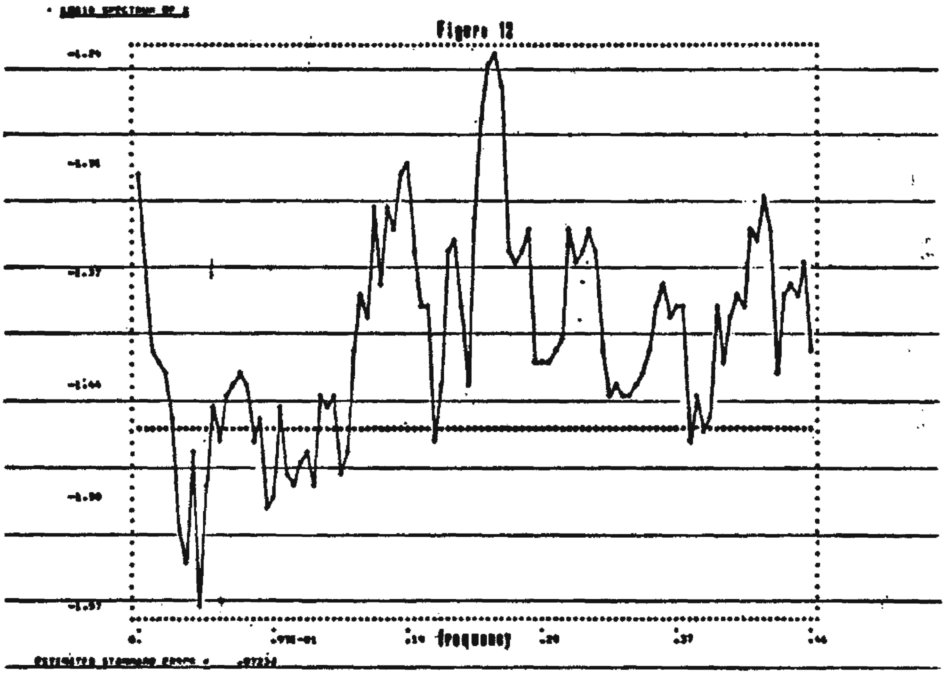


Figure 12
frequency

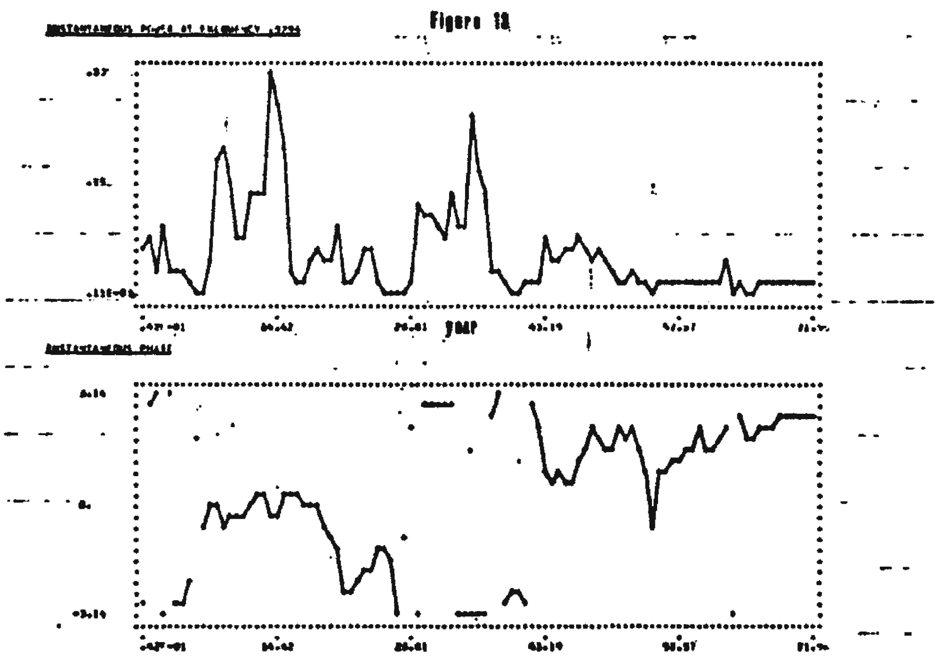


Figure 13

Figure 14

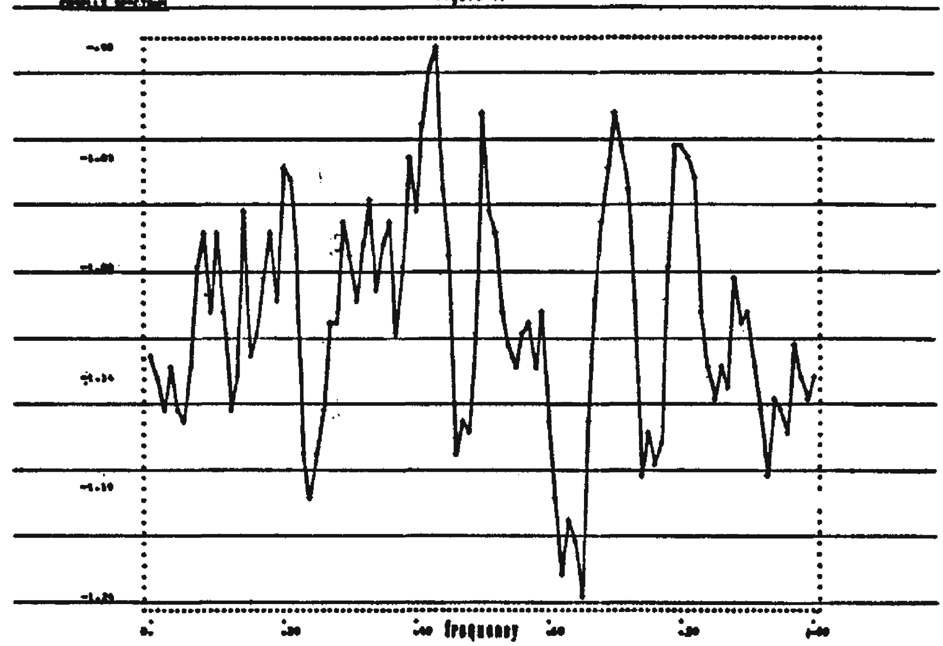
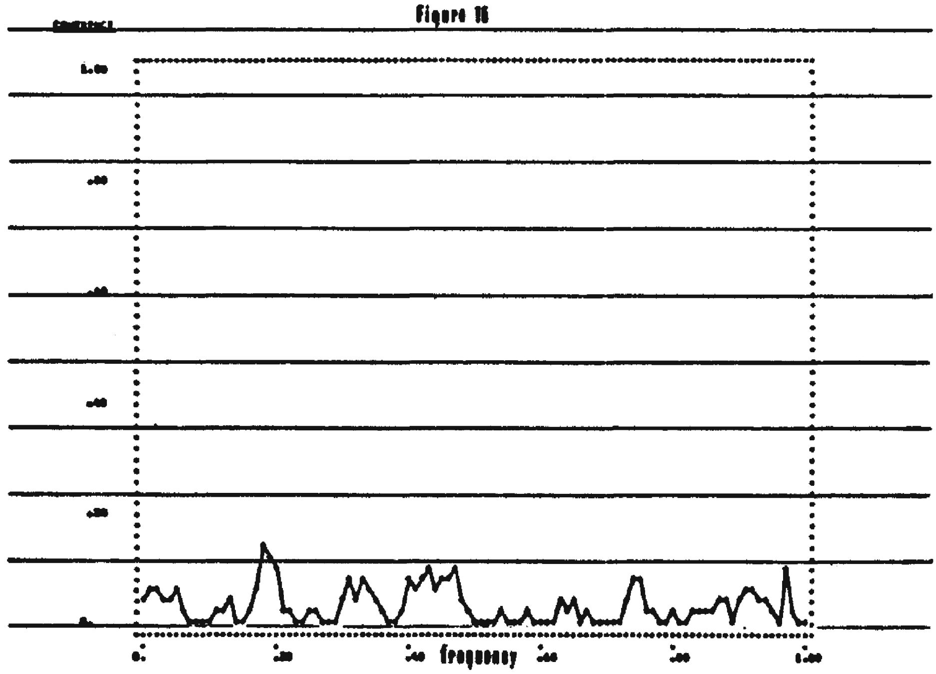


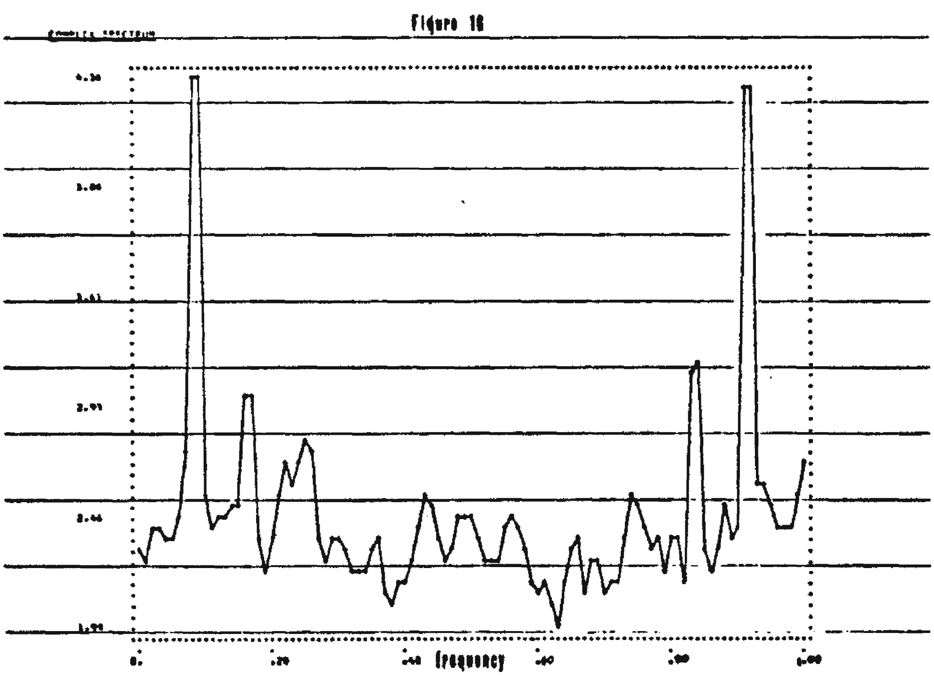
Figure 16



imate null distribution is .069. We have no evidence for a linear time invariant connection between the process $z'(t)$ and the computed series of earthquake effect. Perhaps the most telling thing against earthquakes being a principal cause of the Chandler wobble is an elementary comparison of Figures 11 and 2. Earthquake energy was high at the beginning of this century and has been trailing off since. The wobble amplitude has not been trailing off, in fact it reached its highest level around 1950.

4. Excitation by the Atmosphere

In 1901 Spitaler suggested that the seasonal component of polar motion was due to changes in the inertia tensor of the atmosphere. Hassan (1960) estimated the atmospheric product of inertia, $\psi(t)$, on a monthly basis for the period 1900–1950. Munk and Hassan (1961) carried out a cross-spectral analysis of this data with the polar motion. We carry out a further analysis here. Figure 16 gives $\log_{10}$ of the estimated power spectrum of this atmospheric data. The bandwidth of the estimate is .02. The approximate standard error of the curve is .13. The peaks appearing occur at the seasonal frequency and its harmonics. Figure 17 gives $|\zeta_\psi(t, \lambda)|^2$, arg $\zeta_\psi(t, \lambda)$ for λ at the seasonal frequency and $L = 48$. The instantaneous power is quite level after 1913. The phase is near constant also. The coherence between the series $\Delta z(t)$ and the series $\psi(t)$ is .88 at the seasonal frequency corresponding to polar motion in a negative direction. Figure 18 gives $|\zeta_{\psi'}(t, \lambda)|^2$, arg $\zeta_{\psi'}(t, \lambda)$ for λ the Chandler frequency. There is clearly not much power at this frequency, nor does its variation appear the same as that of the Chandler wobble. Figure 19



gives an estimate of the coherence between the process $\Delta z'(t)$ and the seasonally corrected process $\psi'(t)$. The bandwidth here is .05. The 95 per cent point of the null distribution is .095. We have no evidence to suggest that there exists a linear time invariant relation between the polar series and the atmospheric series at any but the seasonal frequency.

Figure 17

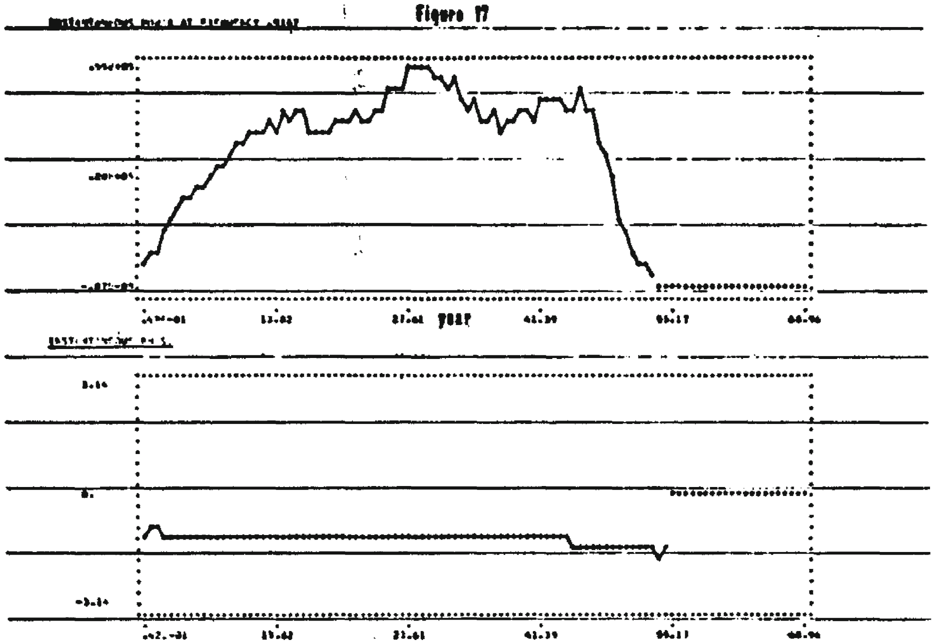


Fig. 18

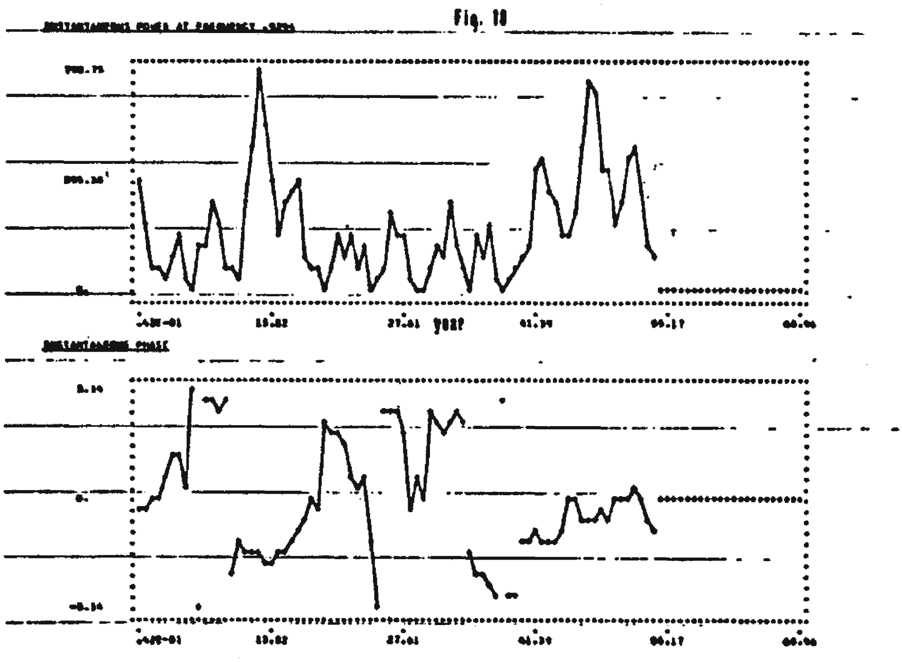
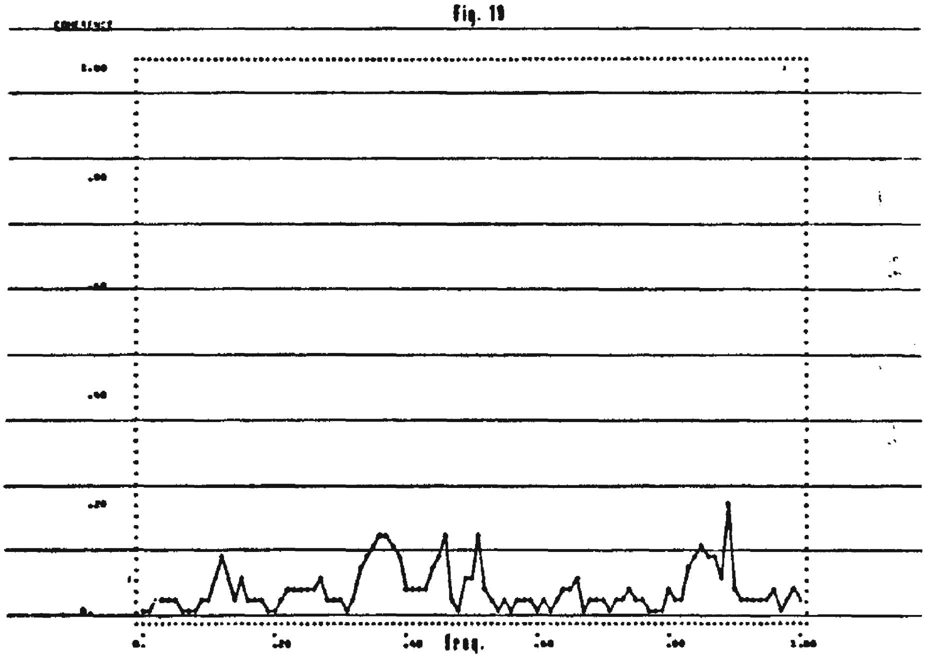


Fig. 11



Acknowledgement

I would like to acknowledge helpful conversations concerning the material of Section 3 with Professors B. A. Bolt. T. V. McEvilly, D. Vere-Jones and Mr. W. Peppin. The research was carried out through the support of NSF Grant GP-31411.

References

Arato, M., Kolmogorov, A. N. and Sinai, Ya. G. (1962): Estimation of the parameters of a complex stationary Gaussian Markov process. *Soviet Math.* 1368–1371.

Brillinger, D. R. (1970): The spectral analysis of stationary interval functions. 483–513 in *Proc. Sixth Berkeley Symp. Math. Stat. Prob*, Vol. 1. Calif. Press, Berkeley.

Brillinger, D. R. and Rosenblatt, M. (1967): Computation and interpretation of k-th order spectra. 189–232 in *Spectral Analysis of Time Series* (ed. B. Harris). J. Wiley, New York.

Calder, N. (1972): *The Restless Earth.* Viking, New York.

Cecchini, G. (1928): Il problema della variazione delle latitudini. *Publ. Reale Obs. Astr. Brera in Milano 61.*

Chandler, S. C. (1891): On the variation of latitude. *Astron. J. 11*, 65, 83.

Chase, C. G. (1972): The N-plate problem of plate tectonics. *Geophys. J. R. astr. Soc. 29*, 117–122.

Dahlen, F. A. (1971): The excitation of the Chandler wobble by earthquakes. *Geophys. J. R. astr. Soc. 25*, 157–206.

Dahlen, F. A. (1972): A correction to the excitation of the Chandler wobble by earthquakes. *Geophys. J. R. astr. Soc. 31.*

Davies, G. F. and Brune, J. N. (1971): Regional and global fault slip rates from sesmicity. *Nature Physical Science 229*, 101–107.

Dubourdieu, G. (1972): *Sur la Periodicite et les Perturbations Astronomiques du Mecanisme Terrestre.* Imp. du Neubourg, France.

Duda, S. J. (1965): Secular seismic energy release in the Circum-Pacific belt. *Tectonophysics 2*, 409–452.
Hassan, E. S. M. (1961): Fluctuations in the atmospheric inertia 1873–1950. *Meteorological Monographs 4*, 1–40.
Huber, P. J., Kleiner, B., Gasser, Th., Dumermuth, G. (1970): Statistical methods for investigating phase relations in stationary stochastic processes. *International Seminar on Digital Processing of Analog Signals*.
Jeffreys, H. (1959): *The Earth*. Cambridge Press, England.
Mandelbrot, B. B. and McCamy, K. (1970): On the secular pole motion and the chandler wobble. *Geophys. J. R. astr. Soc. 21*, 217.
Mansinha, L., Smylie, D. E. and Beck, A. E., editors (1970): *Earthquake Displacement Fields and the Rotation of the Earth*. Reidel, Holland.
Melchior, P. and Yumi, S., editors (1972): *Rotation of the Earth*. Reidel, Holland.
Munk, W. H. and MacDonald, G. J. F. (1960): *The Rotation of the Earth*. Cambridge Press, England.
Rao, C. R. (1965): *Linear Statistical Inference and its Applications*. J. Wiley, New York.
Sinai, Ya. G. (1963): On higher order spectral measures of ergodic stationary processes. *Theory Prob. Appl. 8*, 429–435.
Tukey, J. W. (1961): Discussion emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics 3*, 191–228.
Walker, A. M. and Young, A. (1955): The analysis of the observations of the variation of latitude. *Mon. Not. astr. Soc. 115*, 443.
Walker, A. M. and Young, A. (1957): Further results on the analysis of the variation of latitude. *Mon. Not. Roy. astr. Soc. 117*, 119–141.
Vicente, O. and Yumi, S. (1969): Coordinates of the pole (1899–1968) referred to the Conventional International Origin. *Publ. Inter. Lat. Obs. Mizusawa 7*, 41–50.
Vicente, O. and Yumi, S. (1970): Revised values (1941–1961) of the coordinates of the pole referred to the CIO. *Publ. Inter. Lat. Obs. Mizusawa 7*, 109–112.
Yumi, S., Ishii, H. and Sato, K. (1968): Assumed deformation of the Earth deduced from the observations of the polar motion. *Publ. Inter. Lat. Obs. Mizusawa 6*, 161–184.

Appendix on Complex-Valued Processes

The simplest approach to the definition of the spectral parameters of a complex-valued process is through the spectral representation. Let $W(t)$, $-\infty < t < \infty$, be a complex-valued process with stationary increments and spectral representation

$$W(t) = \int \frac{e^{i\lambda t} - 1}{i\lambda} \, dZ_W(\lambda)$$

Then the spectrum $f_{W\ldots W, W\ldots W}(\lambda_1, \ldots, \lambda_{k+l-1})$, where there are k W's before the comma and l after, is given by

$$\operatorname{cum}\{dZ_W(\lambda_1), \ldots, dZ_W(\lambda_k), \overline{dZ_W(\lambda_{k+1})}, \ldots, \overline{dZ_W(\lambda_{k+l})}\} =$$
$$\delta(\lambda_1 + \ldots + \lambda_k - \lambda_{k+1} - \ldots - \lambda_{l+k}) f_{W\ldots W, W\ldots W}(\lambda_1, \ldots, \lambda_{k+l-1}) \, d\lambda \ldots d\lambda_{l+k}.$$

In particular, the power spectrum of a zero mean series is given by

$$E \, dZ_W(\lambda_1) \overline{dZ_W(\lambda_2)} = \delta(\lambda_1 - \lambda_2) f_{W,W}(\lambda_1) \, d\lambda_1 \, d\lambda_2$$

Sinai (1963) discusses some aspects of the spectral theory of complex-valued processes.

Summary

The axis of rotation of the Earth does not remain fixed relative to the body of the Earth. Instead it has a motion composed of a movement with period 12 months and another movement with period 14.2 months (the Chandler wobble). The 12 month component appears to result from annual fluctuations in the loading of the Earth. The period 14.2 months corresponds to the fundamental frequency of vibration of the Earth. Scientific workers are not agreed upon the cause of the vibration however.

In this paper we use harmonic analysis to examine the possibility that either major earthquakes or annual fluctuations of the atmosphere are the cause. Our computations suggest that neither of these phenomena provides the source of the energy for the vibration.

Résumé

L'axe de rotation de la Terre ne reste pas fixe par rapport au globe terrestre. Son mouvement se décompose en deux composants: un mouvement de periode 12 mois et un autre mouvement de periode 14,2 mois (le mouvement de Chandler). La première composante provient des fluctuations annuelles dans la repatition des masses de la Terre. La deuxième composante correspond à la fréquence fondamentale de vibration de la Terre. Les chercheurs scientifiques ne sont pas d'accord sur le cause de la vibration a la fréquence fondamentale.

Dans cet article, nous utilisons l'analyse harmoniques pour examiner la possibilité que cette vibration proviendre de grands tremblements de terre ou de fluctuations annuelles de l'atmosphere. Nos calculs indiquent qu'auncune de ces séries n'est la source d'energie de la vibration.

Geophys. J. R. astr. Soc. (1979) 59, 593–603

Estimation of uncertainties in eigenspectral estimates from decaying geophysical time series

B. A. Bolt *Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA*
D. R. Brillinger *Department of Statistics, University of California, Berkeley, California 94720, USA*

Received 1979 May 30; in original form 1979 February 14

Summary. The response of many dynamical systems to an impulse is a linear combination of decaying cosines. The frequencies of the cosines have generally been estimated in geophysics by periodogram analysis and little formal indication of uncertainty has been provided. This work presents an estimation procedure by the methods of complex demodulation and non-linear regression that specifically incorporates in the basic model the decaying aspect of the cosines (periodogram analysis does not). The use of plots of the instantaneous phase as a function of time is shown to greatly enhance resolution. Expressions for the variances of eigenfrequencies, amplitudes, phases and damping constants Q are derived by non-linear least-squares. The results are illustrated, for the problem of the free oscillations of the Earth, by computations with the record made at Trieste of the Chilean earthquake of 1960 May 22. Sample values are periods and standard errors of 737.79 ± 0.13 s, 506.25 ± 0.13 s and 429.60 ± 0.14 s for ${}_0T_8$, ${}_0T_{13}$ and ${}_0T_{16}$ with Q values and standard errors of 200 ± 14, 230 ± 28 and 215 ± 30, respectively.

Introduction

A basic need in the measurement of terrestrial eigenspectra is a general algorithm for simultaneously estimating eigenfrequencies, amplitudes, phases and damping coefficients. This paper provides such a method, formulated in a statistical context so that variances of each estimate can also be obtained. The method also has wider applications.

From the beginning of work on the Earth's free vibrations, the emphasis has been on estimation of the spectral eigenfrequencies (Derr 1969; Buland & Gilbert 1978), but few estimates have been accompanied by statistical uncertainties. This requirement is important because independent frequency estimates have been seen to differ by up to 0.5 per cent on occasion (e.g., 2 s for ${}_0T_{14}$, ${}_0T_{17}$) and it is difficult to know how to combine the separate estimates.

Many fewer measurements are available of the actual ground displacements in each eigenvibration (Nowroozi 1974), partly because some key recording instruments were not calibrated for impulse response, but also because some methods of spectrum estimation used could not provide the true amplitudes. New work on terrestrial eigenvibrations is stressing not only measurements of the ground amplitudes but also the damping of amplitudes. As emphasized by several authors (Jobert & Roult 1976; Anderson & Hart 1978), even the most recent estimates of the damping constant (usually given in seismology as the specific dissipation constant Q) show considerable scatter and indicate the great difficulty of precise measurements of the amplitude decay rate. Further, there are questions of whether Q depends on frequency. Progress clearly depends upon the more systematic use of statistical analysis of the time series (Bolt & Brillinger 1975).

The procedures and formula developed in this paper were motivated by the problem of spectral estimation of damped terrestrial eigenvibrations. In particular, the computer programs were tested on the time series obtained by the long-period pendulums at the Grotta Gigante, Trieste, following the 1960 Chile earthquake (Bolt & Marussi 1962). These data have provided some of the best estimates of the gravest torsional eigenfrequencies to the present time. It is hardly necessary to point out, however, that the methodology developed is of a general nature and is applicable to a wide class of geophysical time series.

Our procedure depends heavily on the ability of complex demodulation (Tukey 1961) not only to locate as precise a value of an eigenfrequency as the data permit, but often to allow an assessment of whether difficulties in resolution are arising from such physical causes as multiple energy sources or splitting of peaks due to Earth inhomogeneities and rotation. We investigate especially the use of the instantaneous phase spectrum for decisions on resolution. This is a sensitive method that seems to have received little use in the analysis of geophysical periodicities previously. We set out an informative way of comparison between demodulate estimates of the amplitudes, frequencies and damping factors of the oscillations and with estimates obtained by the technique of non-linear regression (see Draper & Smith 1966). The latter technique allows the relative uncertainties between individual calculated eigenfrequencies to be estimated. The former gives a way to select the most closely resolved modes.

The model

The impulse response, $s(t)$, of a wide variety of stable geophysical, mechanical and electromagnetic linear systems with finite dissipation is a linear combination of decaying cosine waves,

$$s(t;\theta) = \sum_{k=1}^{K} \alpha_k \exp\{-\beta_k t\} \cos\{\gamma_k t + \delta_k\}, \qquad t \geqslant 0, \tag{1}$$

where $\theta = \{\alpha_k, \beta_k, \gamma_k, \delta_k, k = 1, \ldots, K\}$ with $\alpha_k, \beta_k, \gamma_k > 0$, $0 \leqslant \delta_k < 2\pi$ and γ_k distinct. (See Lamb 1920, pp. 230–239; Whittaker 1944, pp. 230–234; Lancaster 1966, Chapter 9.) The γ_k are the eigenfrequencies of the system. The β_k determine the rate of decay of the oscillations and are often redefined as

$$\beta_k = \gamma_k/(2Q_k) \tag{2}$$

in terms of Q_k damping factors.

A traditional means of estimating the γ_k of equation (1) in geophysics has been the searching for peaks in the periodograms, or smoothed periodograms, calculated from the geophysical time series (see e.g. Zadro & Caputo 1968; Dziewonski & Gilbert 1972). The usual numerical procedure has been to calculate the amplitude Fourier spectrum only, using an FFT algorithm. A less usual method involves the fitting of a long autoregressive scheme to the digital record (Burg 1972; Bolt & Currie 1975). None of these estimation procedures have taken specific note of the presence of the damping factor β_k in equation (1), even though, as Dahlen (1978) has lately shown, the concept of damped sinusoids is valuable in theoretical discussions of terrestrial eigenvibrations and multiplets. Also, as mentioned above, their use has generally not been accompanied by the provision of formal indications of the statistical variability of the estimates.

The suggested approach is multi-stage. Assume that one of the traditional methods has been used to determine frequencies that perhaps correspond to eigenvibrations. Then complex demodulation (discussed in the next section) is carried out at the determined frequencies. Examination of the results of complex demodulation suggests whether an individual frequency is reasonable and allows initial estimation of a precise value for the frequency, decay, phase and amplitude. Finally non-linear regression, based on the Fourier transform values in the neighbourhood of a given frequency, is carried out in order to determine final estimates of the spectral parameters and their standard errors.

Complex demodulation

Given a record $X(t)$, $t = 1, \ldots, T$, the complex demodulate at frequency λ of that record is the time series $W(t, \lambda)$, $t = 1, 2, \ldots$, that results from low-pass filtering the series $X(t) \exp\{-i\lambda t\}$. The complex demodulate $W(t, \lambda)$ will be much smoother than the original time series. The technique is described in detail in Bingham, Godfrey & Tukey (1967), Brillinger (1975, p. 33), Bloomfield (1976, Chapter 6), for example.

In the present application, suppose the low-pass filter adopted has impulse response $b(t)$, transfer function $B(\lambda)$ with sufficiently small bandwidth and suppose λ is near an eigenfrequency γ_k. The demodulate may be written

$$W(t, \lambda) = \Sigma b(t - u)X(u) \exp\{-i\lambda t\}. \tag{3}$$

For the signal $s(t, \theta)$ of equation (1), the result of demodulating is

$$Z(t, \lambda) \simeq \tfrac{1}{2}B(0)\alpha_k \exp\{-\beta_k t\} \exp\{i(\gamma_k - \lambda)t + i\delta_k\}. \tag{4}$$

Standardize the low-pass filter by $B(0) = 2$ as we may. Then, from the complex demodulate, one sees the following forms for the instantaneous phase function

$$\arg Z(t, \lambda) \simeq (\gamma_k - \lambda)t + \delta_k \tag{5}$$

and for the logarithm of the instantaneous amplitude function

$$\log_e |Z(t, \lambda)| \simeq -\beta_k t + \log_e \alpha_k. \tag{6}$$

It follows that plots of $\arg W(t, \lambda)$ and $\log |W(t, \lambda)|$ against t can provide evidence of the presence of a damped periodicity in a time series of interest. Indeed, successive variations of the demodulate frequency λ lead to parameter trajectories from which α, β, γ and δ can be estimated in some optimal sense. If the plots (see Figs 1–4) of $\arg W(t, \lambda)$ and $\log_e |W(t, \lambda)|$ are made nearly linear over the record duration T, especially where the signal amplitude is large, the damped vibration is close to the adopted model. If the plot of $W(t, \lambda)$ is erratic, there is a suggestion that the record is just noise. If the plots have regular non-linear behaviour, there is some violation of the basic simple model, perhaps beating between signal and noise harmonics with nearly equal frequencies, perhaps the injection of new energy into the system by applied forces (perhaps an aftershock arrived), perhaps there is time dependent dispersion.

The slope of the logarithm of the instantaneous amplitude curve gives an estimate of the decay constant β_k; the intercept gives the log instantaneous amplitude of the oscillation at the beginning of the movement. Similarly, the intercept of the instantaneous phase plot yields the relative phase of the oscillation. In addition, it should be noted that some idea of the uncertainty of these estimates is given by the variation of the complex demodulate curves about the fitted straight lines over the selected time interval.

Non-linear regression

Consider first, data generated by a model

$$y_j = f_j(\theta) + e_j,$$

$j = 1, \ldots, J$ where the y_j are observed, where the $f_j(\theta)$ are known except for the K-dimensional parameter θ, and where the e_j are unobserved, uncorrelated random errors with mean 0 and common variance σ^2. The least squares estimate of θ is the value providing the minimum of the expression

$$\sum_{j=1}^{J} |y_j - f_j(\theta)|^2.$$

Suppose that the function $f_j(\theta)$ is differentiable with derivatives

$$g_{jk}(\theta) = \frac{\partial f_j(\theta)}{\partial \theta_k}$$

$k = 1, \ldots, K$. Collect the y_j together into the J-vector $\mathbf{y}$, the $f_j(\theta)$ into the J-vector $\mathbf{f}(\theta)$ and the $g_{jk}(\theta)$ into the $J \times K$ matrix $\mathbf{g}(\theta)$. One means of determining an extreme value of θ is through the Gauss–Newton iteration procedure

$$\theta^{n+1} = \theta^n + [\mathbf{g}(\theta^n)^* \mathbf{g}(\theta^n)]^{-1} \mathbf{g}(\theta^n)^* [\mathbf{y} - \mathbf{f}(\theta)] \tag{7}$$

$n = 0, 1, 2, \ldots$, having started with some initial value θ^0. (Other procedures are described in Chambers (1973).) Under regularity conditions this estimate will be approximately normal with mean θ and covariance matrix that may be estimated by

$$[\mathbf{g}(\theta^n)^* \mathbf{g}(\theta^n)]^{-1} \sum_j |y_j - f_j(\theta^n)|^2/(J - K) \tag{8}$$

(see Jennrich 1969).

In the present case, where the noise is not uncorrelated and K is very large, it seems appropriate to modify the above approach as follows. Suppose

$$X(t) = s(t, \theta) + \epsilon(t) \qquad t = 1, \ldots, T$$

where $s(t, \theta)$ is given by equation (1) and $\epsilon(t)$ is a stationary noise series with mean 0 and power spectrum $f_{\epsilon\epsilon}(\lambda)$. Define

$$\Delta^T(\lambda) = \sum_{t=1}^{T} \exp\{-i\lambda t\}$$

$$d_x^T(\lambda) = \sum_{t=1}^{T} X(t) \exp\{-i\lambda t\},$$

$0 \leqslant \lambda \leqslant 2\pi$, with similar definitions for $d_s^T(\lambda)$, $d_\epsilon^T(\lambda)$. By Parseval's formula

$$\sum_{t=1}^{T} |X(t) - s(t,\theta)|^2 = T^{-1} \sum_{j=0}^{T-1} \left| d_x^T\left(\frac{2\pi j}{T}\right) - d_s^T\left(\frac{2\pi j}{T}\right) \right|^2. \tag{9}$$

Minimizing the left-hand side of this expression is equivalent to minimizing the right-hand side. Now $d_s^T(\lambda)$ is a sum of the terms

$$\sum_{t=1}^{T} \alpha_k \exp\{-\beta_k t\} \cos\{\gamma_k t + \delta_k\} \exp\{-i\lambda t\} = b_k \Delta^T(\lambda - \chi_k) + \bar{b}_k \Delta^T(\lambda + \chi_k),$$

$k = 1, \ldots, K$, where $b_k = \frac{1}{2}\alpha_k \exp\{i\delta_k\}$, $\chi_k = \gamma_k + i\beta_k$. By inspection, the term in $\Delta^T(\lambda - \chi_k)$ has appreciable magnitude only for λ near γ_k. This means that the minimum of equation (9) may be obtained approximately by simultaneously minimizing the expressions

$$\sum_{I_k} \left| d_x^T\left(\frac{2\pi j}{T}\right) - b_k \Delta^T\left(\frac{2\pi j}{T} - \chi_k\right) \right|^2 \tag{10}$$

where $I_1, \ldots, I_k$ are disjoint frequency intervals making up the interval $[0, \pi]$, and $2\pi j/T$, λ_k belong to I_k. In addition, for $2\pi j/T$ in I_k,

$$d_\epsilon^T\left(\frac{2\pi j}{T}\right) \simeq d_x^T\left(\frac{2\pi j}{T}\right) - b_k \Delta^T\left(\frac{2\pi j}{T} - \chi_k\right)$$

are approximately independent complex normal variates with zero mean and common variance $2\pi T f_{\epsilon\epsilon}(\lambda)$. (See Brillinger 1975, Theorem 4.4.1. This approximation seems to work very well in practice, *ibid.*)

We proceed by computing the $d_x^T(2\pi j/T)$ using a fast Fourier transform algorithm, identifying the intervals I_k from the periodogram of the record $X(t)$ and then estimating b_k, χ_k by minimizing expression (10) using a Gauss–Newton iteration procedure. The covariance matrix of these estimates may be estimated by an expression analogous to equation (8).

If we think of β_k as fractions of T, say $\beta_k = \phi_k/T$ as seems reasonable in practice (for otherwise the signal $s(t, \theta)$ would quickly become a negligible part of $X(t)$ as t increases) then if $\hat{\alpha}_k$, $\hat{\phi}_k$, $\hat{\gamma}_k$, $\hat{\delta}_k$ denote the least squares estimates it can be shown, by direct extension of the arguments of Hannan (1973) that

$$\begin{aligned}
\operatorname{var} \hat{\alpha}_k &\sim T^{-1} 4\pi f_{\epsilon\epsilon}(\gamma_k) I_2(\phi_k) J(\phi_k)^{-1} \\
\operatorname{var} \hat{\phi}_k &\sim T^{-1} 4\pi f_{\epsilon\epsilon}(\gamma_k) \alpha_k^{-2} I_0(\phi_k) J(\phi_k)^{-1} \\
\operatorname{var} \hat{\gamma}_k &\sim T^{-3} 4\pi f_{\epsilon\epsilon}(\gamma_k) \alpha_k^{-2} I_0(\phi_k) J(\phi_k)^{-1} \\
\operatorname{cov} \{\hat{\alpha}_k, \hat{\phi}_k\} &\sim T^{-1} 4\pi f_{\epsilon\epsilon}(\gamma_k) \alpha_k^{-1} I_1(\phi_k) J(\phi_k)^{-1} \\
\operatorname{cov} \{\hat{\gamma}_k, \hat{\delta}_k\} &\sim T^{-2} 4\pi f_{\epsilon\epsilon}(\gamma_k) \alpha_k^{-2} I_1(\phi_k) J(\phi_k)^{-1},
\end{aligned} \tag{11}$$

$k = 1, \ldots, K$ with all other covariances asymptotically negligible, where

$$I_l(\phi) = \int_0^1 u^l \exp\{-2\phi u\}\, du$$

$l = 0, 1, 2$

$$J(\phi) = I_0(\phi) I_2(\phi) - I_1(\phi)^2.$$

Expressions analogous to those of equation (11) are derived for the case of $\beta_k = 0, k = 1, \ldots, K$ in Whittle (1954), Walker (1971), Hannan (1973).

In the case that a separate record of the noise process, $\epsilon(t)$, is available, it may be used to estimate $f_{\epsilon\epsilon}(\gamma)$ directly and alternate estimates of the variances of interest may be constructed through the formulas of equation (11). This should be done whenever possible as it seems that the variance estimates should be more robust than those produced by the non-linear regression. We were unable to do this in the present case of the Trieste data.

Numerical results

The steps outlined in the paper were applied to the Trieste record of the 1960 May 22 Chilean earthquake. The data were digitized at a time interval of 2 min and tides were removed. The number of data points was 2548 points. The periodogram of the record was examined for peaks. Demodulation was carried out at the peak frequencies (Bolt & Currie 1975). A representative selection of the results is described below. The coefficients $b(t)$ of equation (3) were taken proportional to $1 + \cos(\pi u/L)$ for $|u| < L$ and were 0 otherwise with $L = 200$. (For a lengthy stretch of data, it would have been advantageous to employ a fast Fourier transform in the computations (see Bingham *et al.* 1967)).

Consider first the demodulates for ${}_0T_4$ shown in Fig. 1. The instantaneous phase remains almost constant until about 25 hr from the onset. This is followed by small variations in phase until almost 40 hr when the phase increases sharply, becoming erratic at about 50 hr. There is thus evidence that there is almost a pure harmonic, at about the demodulation frequency chosen, for at least 25 hr and perhaps for another 10. The behaviour of the instantaneous amplitude is consistent with the phase information; a straight line equation (6) might well be fitted to the first 25–35 hr and a decay rate β_k measured. After this time the amplitudes become erratic with large fluctuations which suggest the level of background noise has been reached. Some variation in the fitting of the line equation (6) even to the first part of the spectrum is, however, clearly permissible and numerical fits of straight lines indicate slopes corresponding to Q values between extremes of 300 and 400 are perhaps allowable. If a longer period of recording were used, however, Fig. 1 indicates that a lower decay rate might be calculated (i.e., a false high Q). Comparison with similar plots shows

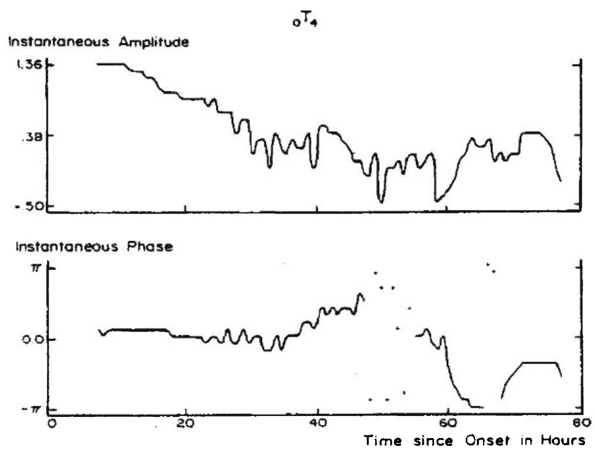


Figure 1. Results of complex demodulation for the mode ${}_0T_4$ at a demodulation period of 1303.15 s from the N–S horizontal Trieste record of the 1964 Chilean earthquake. The upper plot gives the $\log_{10}$ instantaneous ground amplitude as a function of time in hours from the beginning of the record. The lower plot shows the variation in instantaneous phase between $-\pi$ and π with time in hours.

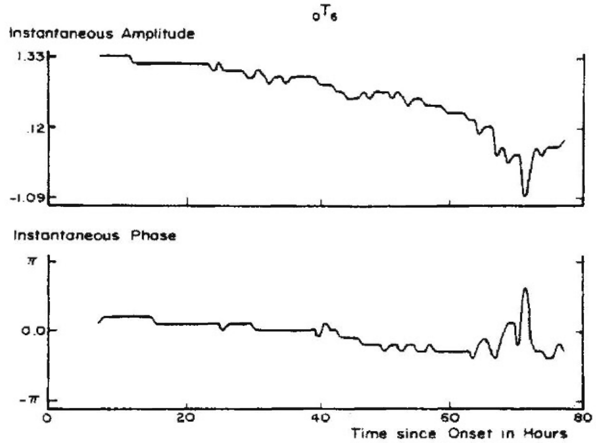


Figure 2. Complex demodulation for $_0T_6$ at a demodulation period of 925.65 s for the Trieste data.

that the $_0T_4$ mode gives one of the more stable instantaneous plots calculated from the Trieste data. In this regard it should be noted that this eigenvibration is well separated from neighbouring torsional and spheroidal modes (see Fig. 1, Bolt & Currie 1975) so that no interference is expected.

Now, consider the similarly isolated $_0T_6$ mode, demodulated in Fig. 2. Here there is even more stability of instantaneous phase and amplitude than for $_0T_4$. There is only a slight drift in phase over the first 65 hr. (This slight drift may indicate that the demodulation frequency adopted could be improved slightly.) The decay for $_0T_6$ is clearly similar to that for $_0T_4$ (note different vertical scales), and a straight line can be fitted to the instantaneous amplitude up to 60 hr with comparable precision. Overall, we would expect the estimate of the $_0T_6$ frequency to be more reliable than that of $_0T_4$.

In Fig. 3, the instantaneous spectra for the demodulates of the spheroidal mode $_0S_9$ are shown. In this case, apart from a hiatus near the beginning of the record, the phase is almost constant near -0.7π throughout the record. We thus have an assurance that we are measuring a single coherent decaying sinusoid throughout most of the recording. Further study of the instantaneous amplitude confirms this, although there is a more rapid decrease

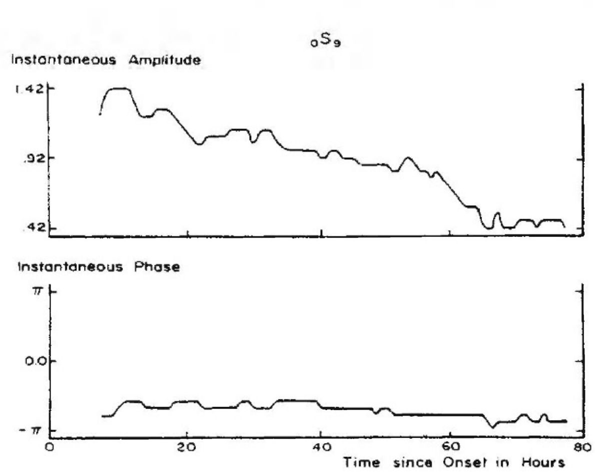


Figure 3. Complex demodulation for $_0S_9$ at a demodulation period of 634.90 s for the Trieste data.

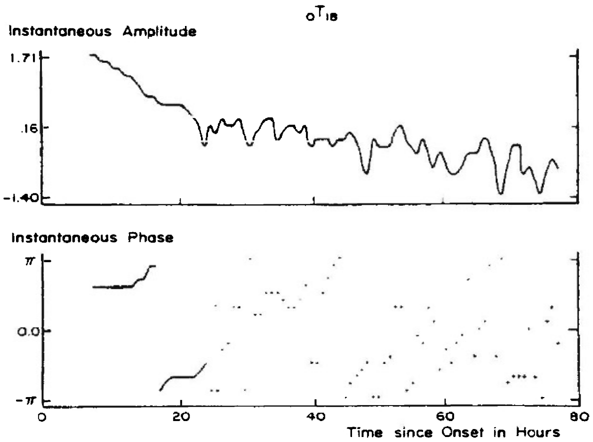


Figure 4. Complex demodulation for $_0T_{18}$ at a demodulation period of 391.45 s for the Trieste data.

in amplitude after about 60 hr. Some allowance for this change can be made in estimating the decay rate β_k. The observed Q is clearly significantly higher for $_0S_9$ than for $_0T_4$ and $_0T_6$ and appears moderately well resolved in the sense of the straight line fit equation (6). The explanation of the change in slope at 60 hr remains unknown, but presumably the effect of interfering signals (noise?) has become more important.

The fourth demodulate presented (Fig. 4) is an illustration of a mode for which the instantaneous phase plot detects major difficulties in resolution. Even 20 hr after the onset the phase angle begins to change rapidly and thereafter cannot be followed. (Note that phase moves continuously from the top to the bottom of the plot.) The conclusion is that only the first 20 hr of amplitudes should be used to estimate Q for this mode. A mean Q value of only about 190 is indicated by the slope of a fitted line. Thereafter the amplitude decays more slowly with large fluctuations. It is interesting that the spectral peak of $_0T_{18}$ is very near that of its neighbour $_0S_{17}$ and it is feasible that some cross modulation (or leakage) is occurring.

Table 1 gives the non-linear least squares estimates of the periods, Q-factors, and relative initial amplitudes and phases for the modes that complex demodulation suggested were truly present and were not multiple. The figures in brackets below give estimates of the corresponding standard errors. (The model was reparameterized to estimate eigenperiods, rather than frequencies, as these seem to be the more usual values discussed.)

The iteration scheme converged exceedingly rapidly. In the case of the modes indicated, the results presented are those obtained after 10 iterations. The case of split peaks might have been handled by fitting the sum of two decaying cosine waves within a frequency interval I_k.

Conclusions

The present paper demonstrates the advantages of the complex demodulation technique for the spectral analysis of geophysical time series composed of damped harmonic terms in the presence of noise. The discussion here focused on the critical problem of improving estimates of amplitudes, frequencies and Q values for the modes of damped eigenvibrations of the Earth. When comparison is possible (Bolt & Currie 1975), it is found that recent estimates of eigenfrequencies for some modes (e.g., $_0T_{14}$, $_0T_{17}$) differ by up to 0.5 per cent.

Table 1. Spectral estimates for the Trieste data (standard errors in parentheses).

Mode	Complex Demodulate Period (sec)	Q	Initial Amplitude (arbitrary units)	Phase (radians)
${}_oT_4$	1303.150 (.697)	347.8 (20.3)	30.64 (4.63)	.216 (.151)
${}_oT_5$	1078.816 (.371)	185.0 (23.6)	62.01 (8.07)	1.973 (.130)
${}_oS_6$	963.005 (.267)	336.2 (62.7)	23.41 (4.30)	1.094 (.184)
${}_oT_6$	925.651 (.332)	357.2 (91.4)	25.55 (6.43)	.160 (.252)
${}_oT_7$	818.377 (.427)	124.8 (16.2)	127.38 (17.35)	-2.624 (.137)
${}_oT_8$	737.791 (.129)	199.6 (14.0)	119.37 (8.74)	2.786 (.073)
${}_oS_8$	707.458 (.202)	376.2 (81.0)	46.73 (10.10)	2.764 (.217)
${}_oS_6$	659.908 (.207)	184.4 (21.3)	125.91 (15.36)	-2.921 (.122)
${}_oS_9$	634.087 (.086)	658.1 (117.3)	23.18 (3.92)	-2.155 (.169)
${}_oT_{10}$	619.202 (.134)	187.6 (15.2)	160.62 (13.77)	-3.049 (.086)
${}_oT_{13}$	506.251 (.134)	230.3 (28.1)	140.77 (17.88)	.532 (.127)
${}_2S_8$	486.827 (.118)	159.4 (12.3)	106.21 (9.33)	-.944 (.088)
${}_oT_{15}$	452.416 (.109)	172.2 (14.3)	126.21 (11.41)	.040 (.091)
${}_oT_{16}$	429.604 (.141)	214.5 (30.1)	155.35 (22.94)	2.940 (.148)
${}_oT_{18}$	391.447 (.088)	188.3 (16.0)	119.32 (10.76)	1.929 (.091)
${}_oT_{20}$	359.399 (.041)	248.7 (14.2)	84.15 (5.05)	-2.558 (.061)

However, it is difficult to combine the independent estimates because of the lack of comparable probability models. It is recommended that the present method be used so that pooling with appropriate weights can be made.

Studies of terrestrial eigenspectra have now advanced to the stage when analysis of a long record of free oscillations must provide more than a set of mean eigenfrequencies. Not only does the rotation and ellipticity of the Earth produce frequency multiplets about the central (degenerate) frequency, but lateral inhomogeneities split the peaks also. Earthquake sources of various types and at various locations generate at different seismographic stations different relative strengths in the multiplets. As well, variations in long-period noise spectra, activation of new sources, and rotation of the nodal lines, relative to the receiver, all produce fluctuations which complicate the meaning to be attached to a simple mean eigenfrequency estimate. It is demonstrated in this paper that the plots of the real and imaginary parts of the complex demodulates of each mode provide a powerful way to detect and explore such fluctuations. The eccentric behaviour is not 'swept under the rug' as occurs with most traditional methods. Already some progress in the geophysical interpretation of these plots has been made (Hansen 1978).

Formulae in the present paper enable programs to be written to compute relatively quickly the complex demodulates and, by non-linear least squares, variances of the spectral parameters. By repetition at successive steps in the demodulating frequency, the set of instantaneous amplitude and phase plots allows a decision to be made on the best eigenfrequency resolution available from the data and the quality of the damping factor Q and amplitude of ground motion that can be obtained. Clearly more experience with the method is needed before specific rules for decisions can be given.

Acknowledgments

We would like to thank Professor John Tukey for his valuable discussion of an early draft and Mr R. Hansen for helpful comments. The work was supported in part by the National Science Foundation under Grants EAR 76-00118 and NCS 76-06117.

References

Anderson, D. L. & Hart, R. S., 1978. Attenuation models of the Earth, *Phys. Earth planet. Interiors*, **16**, 289–306.

Bingham, C., Godfrey, M. D. & Tukey, J. W., 1967. Modern techniques of power spectrum estimation, *IEEE Trans.*, AU-15, 56–66.

Bloomfield, P., 1976. *Fourier Analysis of Time Series: An Introduction*, J. Wiley, New York.

Bolt, B. A. & Brillinger, D. R., 1975. Estimation of uncertainties in fundamental frequencies of decaying geophysical time series (abstract), *EOS Trans. Am. geophys. Un.*, **56**, 403.

Bolt, B. A. & Currie, R., 1975. Maximum entropy estimates of Earth torsional eigenperiods from 1960 Trieste data, *Geophys. J. R. astr. Soc.*, **40**, 107–114.

Bolt, B. A. & Marussi, A., 1962. Eigenvibrations of the Earth observed at Trieste, *Geophys. J. R. astr. Soc.*, **6**, 299–311.

Brillinger, D. R., 1975. *Time Series Data Analysis and Theory*, Holt, Rinehard & Winston, New York.

Buland, R. & Gilbert, F., 1978. Improved resolution of complex eigenfrequencies in analytically continued seismic spectra, *Geophys. J. R. astr. Soc.*, **52**, 457–470.

Burg, J. P., 1972. The relationship between maximum entropy spectra and maximum likelihood spectra, *Geophysics*, **37**, 375–376.

Chambers, J. M., 1973. Fitting nonlinear models: numerical techniques, *Biometrika*, **60**, 1–13.

Dahlen, F. A., 1978. *The Spectra of Unresolved Split Normal Mode Multiplets*, Princeton University, Princeton, N.J.

Derr, J., 1969. Free oscillation observations through 1968, *Bull. seism. Soc. Am.*, **59**, 2079–2099.

Draper, N. R. & Smith, H., 1966. *Applied Regression Analysis*, Wiley, New York.

Dziewonski, A. M. & Gilbert, F., 1972. Observations of normal modes from 84 recordings of the Alaskan earthquakes of 1964 March 28, *Geophys. J. R. astr. Soc.*, **27**, 393–446.

Hannan, E. J., 1973. The estimation of frequency, *J. appl. Prob.*, **10**, 510–519.

Hansen, R., 1978. Complex demodulation estimates of eigenspectral amplitudes and damping factors from the August, 1977 Indonesian earthquake, *EOS* (abstract), **59**, 1140.

Jennrich, R. J., 1969. Asymptotic properties of non-linear least-squares estimators, *Ann. Math. Stat.*, **40**, 633–643.

Jobert, N. & Roult, G., 1976. Periods and damping of free oscillations observed in France after sixteen earthquakes, *Geophys. J. R. astr. Soc.*, **45**, 155–176.

Lamb, H., 1920. *Higher Mechanics*, Cambridge Press, Cambridge.

Lancaster, P., 1966. *Lambda-matrices and Vibrating Systems*, Pergamon, Oxford.

Nowroozi, A. A., 1974. Characteristic periods and Q for oscillations of the Earth following intermediate-focus earthquakes, *J. Phys. Earth*, **22**, 1–23.

Tukey, J. W., 1961. Discussion emphasizing the connection between analysis of variance and spectrum analysis, *Technometrics*, **3**, 1–29.

Walker, A. M., 1971. On the estimation of a harmonic component on a time series with stationary independent residuals, *Biometrika*, 58, 21–36.

Whittaker, E. T., 1944. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Dover, New York.

Whittle, P., 1954. Some recent contributions to the theory of stationary processes, Appendix 2 of *A Study in the Analysis of Stationary Time Series* by H. Wold, Almquist, Stockholm.

Zadro, M. B. & Caputo, M., 1968. Spectral, bispectral analysis and Q of the free oscillations of the Earth, *Suppl. Nuovo Cimento*, 6, 67–81.

Statistical Inference for Stationary Point Processes[1]

by David R. Brillinger

The University of California, Berkeley

Introduction

This work is divided into three principal sections which also correspond to the three lectures given at Bloomington. The topics cover, some useful point process parameters and their properties, estimation of time domain parameters and the estimation of frequence domain parameters. The work may be viewed as an extension of some of the results in Cox and Lewis (1966, 1972) to apply to vector-valued processes and to higher order parameters. It will proceed at a heuristic level rather than formal. A formal approach may be found in Daley and Vere-Jones (1972) for example. The notation $\int f$ will be used for $\int f(x)\, d\mu(x)$, μ being Lebesgue measure. A general lemma concerning the existence of consistent estimates is given in Section IV.

I. Point Process Parameters

Consider isolated points of r different types randomly distributed along the real line, R.

---

[1] Prepared while the author was a Miller Research Professor and with the support of N.S.F. Grant GP-31411.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_29, © Springer Science+Business Media, LLC 2012

Examples that we have in mind include, the times of heart beats or earthquakes in the case $r = 1$, the times of nerve pulses released by a network of r nerve cells in the case of general r. Let $N_a(A)$ denote the number of points of type a falling in the interval $A \subset R$ and let $N_a(t) = N_a(0,t]$ for $a = 1,\ldots,r$.

1. Suppose

Prob {point of type a in $(t,t+h]$} $\sim p_a(t)h$ as $h \downarrow 0$. $p_a(t)$ provides a measure of the <u>intensity</u> with which points of type a occur near t. We can often conclude that

$$E\, N_a(t) = \int_0^t p_a(t)\, dt$$

2. Suppose, for $t_1 \neq t_2$

Prob{point of type a in $(t_1,t_1+h_1]$ and point of type b in $(t_2,t_2+h_2]$}

$$\sim p_{ab}(t_1,t_2)h_1h_2$$

as h_1, $h_2 \downarrow 0$. $p_{ab}(t_1,t_2)$ provides a measure of the intensity with which points of type a occur near t_1 and simultaneously points of type b occur near t_2.

A related useful measure is provided by

Prob{point of type a in $(t_1,t_1+h]$ | point of type b at t_2}

$$\sim p_{ab}(t_1,t_2)h/p_b(t_2)$$

as $h \downarrow 0$. The ratio $p_{ab}(t_1,t_2)/p_b(t_2)$ is seen to provide a measure of the intensity with which type a points occur near t_1, given that there is a type b point at t_2. In the case that type a points are distributed independently of type b points, $p_{ab}(t_1,t_2) = p_a(t_1)p_b(t_2)$, and the ratio becomes $p_a(t_1)$, the first order intensity. The function $p_{ab}(t_1,t_2)$ is like the second order moment function of ordinary time series; however in practise it seems to be much more useful as it has a further interpretation as a probability. Often it is true that

$$E\, N_a(t)\, N_b(t) = \int_0^t \int_0^t p_{ab}(t_1,t_2)dt_1dt_2 \quad \text{for } a \neq b$$

$$= \int_0^t \int_0^t p_{ab}(t_1,t_2)dt_1dt_2 + \int_0^t p_a(t)\, dt \qquad \text{for } a = b$$

3. Suppose next that, for $t_1,\ldots,t_k$ distinct and $\nu_1,\ldots,\nu_r$ non-negative integers with sum k

$$\text{Prob}\{\text{type a point in each of } (t_j, t_j+h_j],$$
$$j = \sum_{b<a} \nu_b + 1,\ldots, \sum_{b\le a} \nu_b \text{ and } a = 1,\ldots,r\}$$
$$\sim p^{(\nu_1)\ldots(\nu_r)}(t_1,\ldots,t_k)h_1\ldots h_k \qquad (1)$$

as $h_1,\ldots,h_k \downarrow 0$; $k = 1,2,\ldots$. (An alternate notation, consistent with the cases k = 1,2 above is

$$\text{Prob}\{\text{type } a_j \text{ point in } (t_j, t_j + h_j],\ j = 1,\ldots,k\}$$

$$\sim p_{a_1 \cdots a_k}(t_1,\ldots,t_k)\ h_1 \cdots h_k$$

as $h_1,\ldots,h_k \downarrow 0$; $k = 1,2,\ldots$.) The function $p^{(\nu_1)\cdots(\nu_r)}$ is called a product density of order k. Such a function was introduced by S. O. Rice in a particular situation and by A. Ramakrishnan in a general situation, see Srinivasan (1974). No claim is made that the probability in (1) always depends on $h_1,\ldots,h_k$ in such a direct manner. Rather it is the claim that this happens for an interesting class of examples. Brillinger (1972) gives an expression for

$$E\, N_{a_1}(t_1) \cdots N_{a_k}(t_k)$$

4. The probability generating functional of the process $\underset{\sim}{N}(t) = \{N_1(t),\ldots,N_r(t)\}$ is defined by

$$G[\xi_1,\ldots,\xi_r] = E[\exp\{\int \log \xi_1(t)\ dN_1(t) + \cdots + \int \log \xi_r(t)\ dN_r(t)\}]$$

for suitable functions $\xi_1 \ldots \xi_r$. Writing it as

$$E[\prod_{a=1}^{r} \prod_{\tau \text{ type } a \text{ point}} \{1 + (\xi_a(\tau)-1)\}]$$

and expanding, we can see that it is given by

$$\sum_{\nu_1,\ldots,\nu_r} \frac{1}{\nu_1! \cdots \nu_r!} \int (\xi_1 - 1)^{(\nu_1)} \cdots (\xi_r - 1)^{(\nu_r)}\ p^{(\nu_1)\cdots(\nu_r)} .$$

where we define

$$(\xi(t_1,\ldots t_\nu) - 1)^{(\nu)} = (\xi(t_1) - 1)\ldots(\xi(t_\nu)-1)$$

This functional is of use in computing probabilities of interest for the process. For example setting

$$\begin{aligned} \xi_a(t) &= z_a \quad \text{for } t \in A \\ &= 1 \quad \text{for } t \notin A \end{aligned}$$

and determining the coefficient of $z_1^{j_1}\ldots z_r^{j_r}$ we see that

$$\text{Prob}\{N_1(A) = j_1,\ldots,N_r(A) = j_r\}$$

$$= \frac{1}{j_1! \ldots j_r!} \sum_{\nu_1 \geq j_1} \cdots \sum_{\nu_r \geq j_r} \cdot$$

$$\frac{(-1)^{\nu_1 - j_1 + \ldots + \nu_r - j_r}}{(\nu_1 - j_1)! \ldots (\nu_r - j_r)!} \cdot$$

$$\int_A p_{\nu_1+\ldots+\nu_r}^{(\nu_1)\ldots(\nu_r)} \tag{2}$$

We may likewise determine <u>conditional product densities</u> such as

$$p^{(\gamma_1)\cdots(\gamma_r)}(t_1,\ldots,t_k \mid N_1(A)=j_1,\ldots,N_r(A)=j_r)$$

$$= \frac{1}{(j_1-\nu_1)!\cdots(j_r-\gamma_r)!} \sum_{\nu_1 \ge j_1} \cdots \sum_{\nu_r \ge j_r} \frac{(-1)^{\nu_1-j_1+\cdots+\nu_r-j_r}}{(\nu_1-j_1)!\cdots(\nu_r-j_r)!} \int_{A^{\nu_1-\gamma_1+\cdots+\nu_r-\gamma_r}}$$

$$p^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_{\gamma_1};\ldots;t_{\gamma_1+1},\ldots,t_{\gamma_1+\gamma_2};\ldots:$$

$$\ldots)/(2)$$

These conditional product densities are useful in statistical inference. They provide likelihood functions and also allow the investigation of the distribution of statistics conditionally **on the** observed number of points. (Were $N(A) = 0$, one wouldn't want to claim much.)

The integrated product densities give the factorial moments of the process. For example, if $N_{(\nu)} = N(N-1)\ldots(N-\nu+1)$, then

$$E\, N_1(A)_{(\nu_1)}\cdots N_r(A)_{(\nu_r)} = \int_{A^k} p^{(\nu_1)\cdots(\nu_r)}$$

Also of use are the <u>cumulant densities</u>, $q^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_k)$ given by

$$\log G[\xi_1,\ldots,\xi_r] = \sum_{\nu_1,\ldots,\nu_r} \frac{1}{\nu_1!\cdots\nu_r!} \int (\xi_1-1)^{(\nu_1)}\cdots(\xi_r-1)^{(\nu_r)} q^{(\nu_1)\cdots(\nu_r)} \quad (3)$$

They measure the degree of dependence of increments of the process at different t_j.

Certain other conditional product densities are of use. We mention

$$\text{Prob}\{\text{type a point in each of } (t_j, t_j+h_j],\ j = \sum_{b<a} \nu_b + 1, \ldots, \sum_{b\le a} \nu_b \text{ and } a = 1,\ldots,r \,|\, N_1\{0\} = 1\}/(h_1 \ldots h_k)$$

$$\sim p^{(\nu_1)\ldots(\nu_r)}(t_1,\ldots,t_k \mid N_1\{0\} = 1)$$

$$= p^{(\nu_1+1)(\nu_2)\ldots(\nu_r)}(0,t_1,\ldots,t_k)/p_1(0)$$

and for $\tau_1,\ldots,\tau_k \le t$

$$\text{Prob}\{\text{type 1 point in } (t,t+h] \mid \nu_1 \text{ points of type 1, } \nu_2 \text{ points of type 2, } \ldots \text{ at } \tau_1, \tau_2, \ldots, \tau_k \text{ respectively}\}/h$$

$$\sim p^{(\nu_1+1)(\nu_2)\ldots(\nu_r)}(t,\tau_1,\tau_2,\ldots,\tau_k)/p^{(\nu_1)\ldots(\nu_r)}(\tau_1,\ldots,\tau_k)$$

If all points up to t are included, this becomes the complete intensity

$$\lim_{h\downarrow 0} \text{Prob}\{\text{type 1 point in } (t,t+h] \mid \underset{\sim}{N}(u),\ u \le t\}$$

5. Certain probabilities and moments are of special interest. We list some of these.

(i) the renewal functions

$$U_{ab}(t) = E\{N_a(t) \mid N_b\{0\} = 1\} \qquad \text{for } t > 0$$

$$= \int_0^t p_{ab}(u,0)\, du \,/\, p_b(0) \qquad a,b=1,\ldots,r.$$

The renewal density is $p_{ab}(t,0)/p_b(0)$.

(ii) the forward recurrence time distribution is given by

$$\text{Prob}\{\text{event before or at } t\}$$
$$= \text{Prob}\{\text{time of next event from 0 is} \le t\}$$
$$= 1 - \text{Prob}\{N(t) = 0\}$$
$$= 1 - \sum_{\nu \ge 0} \frac{(-1)^\nu}{\nu!} \int_{(0,t]^\nu} p^{(\nu)}$$

(iii) the survivor function (or distribution of lifetime)

$$\text{Prob}\{\text{time of next event from 0 is} > t \mid N\{0\} = 1\}$$
$$= \text{Prob}\{N(t) = 0 \mid N\{0\} = 1\}$$
$$= p(0)^{-1} \sum_{\nu \ge 0} \frac{(-1)^\nu}{\nu!} \int_{(0,t]^\nu} p^{(\nu+1)}(0,\ldots$$
$$= 1 - F(t) \text{ say.}$$

(iv) the hazard function or force of mortality

$$\mu(t) = f(t)/(1 - F(t))$$

$$\sim \text{Prob}\{\text{point in } (t,t+h] \mid N\{0\} = 1, N(t) = 0\}/h$$

where $F(t)$ is given in (iii) and $f(t)$ is its derivative.

(v) the variance time curve

$$\text{var } N(t) = E\, N(t)(N(t) - 1) + E\, N(t) - (E\, N(t))^2$$
$$= \int_0^t \int_0^t p^{(2)}(t_1,t_2)dt_1\, dt_2 + \int_0^t p(t)\, dt - \left(\int_0^t p(t)dt\right)^2$$

(vi) the Palm functions

$$q_1(j_1,\ldots,j_r\, ;\, t) = \text{Prob}\{N_1(t) = j_1,\ldots,N_r(t) = j_r \mid N_1\{0\} = 1\}$$
$$= \frac{1}{j_1!\cdots j_r!p_1(0)} \sum_{\nu_1 \geq j_1} \cdots \sum_{\nu_r \geq j_r} \frac{(-1)^{\nu_1-j_1+\cdots+\nu_r-j_r}}{(\nu_1-j_1)!\cdots(\nu_r-j_r)!}$$
$$\int_{(0,t]^{\nu_1+\cdots+\nu_r}} p^{(\nu_1+1)(\nu_2)\cdots(\nu_r)} (0,\ldots$$

6. We next indicate the values of a few of these parameters for some examples of interest.

Example 1. The Poisson process with mean intensity $p(t)$. The numbers of points in disjoint intervals $I_1,\ldots,I_k$ are independent Poisson variates with means $P(I_1),\ldots,P(I_k)$ respectively where $P(I) = \int_I p(t)\, dt$. Here

$$p^{(k)}(t_1,\ldots,t_k) = p(t_1)\cdots p(t_k)$$

and so

$$G[\xi] = \exp\{\int(\xi(t) - 1)\, p(t)\, dt\}$$
$$\text{Prob}\,\{N(A) = j\} = \frac{1}{j!} P(A)^j \exp\{-P(A)\}$$
$$p^{(k)}(t_1,\ldots,t_k \mid N(A) = j) = \frac{j!}{(j-k)!} \frac{p(t_1)}{P(A)} \cdots \frac{p(t_k)}{P(A)}$$

If $P(t) = \int_0^t p(t)\, dt$ and $N'(s)$, $s \in R_+$ is a Poisson process with mean intensity 1, then the general process may be represented as

$$N(t) = N'(P(t))$$

Example 2. The doubly stochastic Poisson process. Suppose $\{x_1(t),\ldots,x_r(t)\}$, $t \in R_+$, is a process with non-negative sample paths, moments

$$m^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_k) = E\{x_1(t_1)\cdots x_1(t_{\nu_1}) \\ x_2(t_{\nu_1+1})\cdots x_r(t_k)\}$$

and moment generating functional

$$M[\theta_1,\ldots,\theta_r] = E[\exp\{\int\theta_1(t)x_1(t)dt + \ldots + \\ \int\theta_r(t)x_r(t)dt\}]$$

Suppose after a realization of this process is obtained, independent Poissons with mean intensities $x_1(t),\ldots,x_r(t)$ are generated. Then

$$p^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_k) = m^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_k)$$

$$G[\xi_1,\ldots,\xi_r] = M[\xi_1-1,\ldots,\xi_r-1]$$

$$= E[\exp\{\int(\xi_1(t)-1)x_1(t)dt+\ldots\}]$$

If $X_a(t) = \int_0^t x_a(t)dt$, and $N_1'(s),\ldots,N_r'(s)$ are independent Poissons with mean intensities 1, then this process may be represented as

$$N_1'(X_1(t)),\ldots,N_r'(X_r(t)) \quad .$$

This process seems to be useful for checking out general formulas that have been developed, such as (2) and (3), among other things.

Example 3. The cluster process. Suppose $N_1'(t),\ldots,N_r'(t)$ is a primary process of cluster centers with probability generating functional $G[\xi_1,\ldots,\xi_r]$. Suppose that secondary points are generated in independent clusters centered at the points of $\underset{\sim}{N}'$. Suppose that the p.g.f. for cluster points of type a centered at t is $G_a[\xi|t]$. Then the p.g.f. of the overall process is

$$\begin{aligned} G[\xi_1,\ldots,\xi_r] &= E\{\prod_{j,k}\xi_1[\sigma_j^1+\tau_{jk}^1]\ldots\prod_{j,k}\xi_r[\sigma_j^r+\tau_{jk}^r]\} \\ &= E\{\prod_j G_1[\xi_1|\sigma_j^1]\ldots\prod_j G_r[\xi_r|\sigma_j^r]\} \\ &= G[G_1[\xi_1|\cdot],\ldots,G_r[\xi_r|\cdot]] \end{aligned}$$

If $r = 2$, and the first component is the primary process and the second component corresponds to clusters of one member, then we have a process of the character of the $G/G/\infty$ queue.

Example 4. The renewal process. Here the points correspond to the partial sums of a random walk with positive steps. Suppose $r = 1$, $t_1 < t_2 < \ldots < t_k$, then

$$p^{(k)}(t_1,\ldots,t_k) = p^{(1)}(t_1)\frac{p^{(2)}(t_2,t_1)}{p^{(1)}(t_1)}\frac{p^{(2)}(t_3,t_2)}{p^{(1)}(t_2)} \cdots \frac{p^{(2)}(t_k,t_{k-1})}{p^{(1)}(t_{k-1})}$$

where $p^{(1)}$ and $p^{(2)}$ satisfy renewal equations, see p. 35 in Srinivasan (1974).

Example 5. Zero crossing processes. Expressions may be set down for the product densities of point processes corresponding to the zeros of random functions, see Leadbetter (1972).

7. We now turn to a consideration of the case in which the process is stationary, that is, probability distributions are invariant under translations of t. This means for example,

$$p_a(t) = p_a$$

$$p_{ab}(t_1,t_2) = p_{ab}(t_1-t_2)$$

$$p^{(\nu_1)\cdots(\nu_r)}(t_1,\ldots,t_k) = p^{(\nu_1)\cdots(\nu_r)}(t_1-t_k,\ldots,t_{k-1}-t_k)$$

and if S^u denotes the shift transformation, $S^u\xi(t) = \xi(t+u)$, then

$$G[S^u\xi_1,\ldots,S^u\xi_r] = G[\xi_1,\ldots,\xi_r].$$

As the process has stationary increments, it has a spectral representation

$$N_a(t) = \int_{-\infty}^{\infty} [(\exp\{it\lambda\}-1)/(i\lambda)]\, dZ_a(\lambda)$$

for a = 1,...,r. We may define <u>cumulant spectra</u> of order k by

$$\delta(\lambda_1+\ldots+\lambda_k)\ f^{(\nu_1)\ldots(\nu_r)}(\lambda_1,\ldots,\lambda_{k-1})d\lambda_1\ldots d\lambda_k$$
$$= \text{cum}\{dZ_1(\lambda_1),\ldots,dZ_1(\lambda_{\nu_1}),\ldots,dZ_r(\lambda_k)\}$$

with $\delta(\bullet)$ the Dirac delta function. Alternately, making use of product densities, we might define the <u>power spectra</u> by

$$f_{aa}(\lambda) = (2\pi)^{-1}[p_a + \int_{-\infty}^{\infty}\{p_{aa}(u) - p_a^2\}\exp\{-i\lambda u\}d\lambda]$$

$-\infty<\lambda<\infty$, a = 1,...,r, since

$$\text{cov}\{dN_a(t+u),\ dN_a(t)\} = [p_a\delta(u) + \{p_{aa}(u)-p_a^2\}]dt\, du$$

and <u>cross-spectra</u> by

$$f_{ab}(\lambda) = (2\pi)^{-1}\int_{-\infty}^{\infty}\{p_{ab}(u)-p_ap_b\}\ \exp\{-i\lambda u\}du$$

$-\infty<\lambda<\infty$, $1 \le a \ne b \le r$, since

$$\text{cov}\{dN_a(t+u),\ dN_b(t)\} = \{p_{ab}(u)-p_ap_b\}dt\ du$$

The functions $q_{ab}(u) = p_{ab}(u) - p_ap_b$ tend to 0 as $|u| \to \infty$ and are integrable for many processes (processes whose distant increments are only weakly dependent.) In this connection we set down the mixing condition,

Assumption I. $\underset{\sim}{N}(t)$, $t \in R$, is an r vector-valued stationary point process satisfying (1), whose cumulant densities of (3) satisfy

$$\int \cdots \int_{R^{k-1}} |q^{(\nu_1)\cdots(\nu_r)}(u_1,\ldots,u_{k-1})| du_1 \ldots du_{k-1} < \infty$$

for $\nu_1,\ldots,\nu_r = 0,1,2,\ldots,\nu_1+ \ldots + \nu_r \geq 2$.

The second-order spectra of the process, $f_{ab}(\lambda)$, possess many of the same properties as the spectra of ordinary time series. There are however some differences, we mention that

$$\lim_{|\lambda|\to\infty} f_{aa}(\lambda) = (2\pi)^{-1} p_a$$

for mixing point processes instead of the 0 limit for mixing ordinary time series.

The spectral representation may be used to relate the point process to the associated ordinary time series

$$\underset{\sim}{X}(t) = h^{-1}\underset{\sim}{N}(t-\tfrac{h}{2},\ t+\tfrac{h}{2}) = \int \exp\{i\lambda t\}[(\sin h\lambda/2)/(h\lambda/2)]d\underset{\sim}{Z}(\lambda)$$

$t \in R$. This shows, for example, that the cross-spectrum of the a-th and b-th components of $\underset{\sim}{X}(t)$ is

$$[(\sin h\lambda/2)/(h\lambda/2)]^2 f_{ab}(\lambda)$$

8. A key indicator of the appearance of the process of points of type 1, say, is provided by

$$h^{-1}\ N_1(t,t+h] \qquad h \text{ small}$$

the empirical intensity with which points of type 1 are seen to occur near t. Models for the process may usefully involve models for this variate. A simple statement says

$$\text{Prob}\{\text{point of type 1 in } (t,t+h]\} \sim p_1 h$$

for h small. A more complicated statement is

$$\text{Prob}\{\text{point of type 1 in } (t,t+h] \mid \text{point of type a at } \tau\}$$
$$\sim p_{1a}(t-\tau)h/p_a$$

In the case that the process 1, near t, is independent of the process a, near τ, this last is $\sim p_1 h$, the marginal intensity. This happens often as $|t-\tau| \to \infty$. An even more complicated statement involves

$$\text{Prob}\{\text{point of type 1 in } (t,t+h] \mid \nu_1 \text{ points of type 1, } \nu_2 \text{ points of type 2, } \ldots \text{ at } \tau_1, \tau_2, \ldots, \tau_k \text{ respectively}\}$$
$$\sim p^{(\nu_1+1)(\nu_2)\ldots(\nu_r)}(t-\tau_k, \tau_1-\tau_k, \ldots, \tau_{k-1}-\tau_k)h/$$
$$p^{(\nu_1)\ldots(\nu_r)}(\tau_1-\tau_k, \ldots, \tau_{k-1}-\tau_k)$$

Suppose r = 2. A useful simple model here is;

$$\text{Prob}\{\text{point of type 1 in } (t,t+h] \mid N_2(u), -\infty<u<\infty\}$$
$$\sim \{\mu + \int a(t-u)\ dN_2(u)\}h \qquad (4)$$
$$\sim \{\mu + \sum_j a(t-\tau_j)\}h$$

where the τ_j are the times of the events of the second process. This model allows the intensity, near t, of points of type 1 to be affected in a direct manner by points of type 2. If the system is causal, then $a(u) = 0,\ u < 0$. The second process may excite or inhibit the first process depending on the sign of $a(u)$.

The model implies, for example,

$$p_1 = \mu + p_2 \int a(u)\ du \qquad (5)$$

showing that μ may be interpreted as the intensity with which type 1 points would occur where $p_2 = 0$. Also

$$p_{12}(t) = \mu p_2 + p_2 a(t) + \int a(t-u)\ p_{22}(u) du \qquad (6)$$

If

$$A(\lambda) = \int a(u)\ \exp\{-i\lambda u\} du$$

then (5) and (6) lead to

$$p_1 = \mu + p_2 A(0)$$

$$f_{12}(\lambda) = A(\lambda)\ f_{22}(\lambda)$$

suggesting how the parameters μ, $A(\lambda)$ might be identified. If $p_{22}(u)$ is constant, as in the Poisson case, then (6) leads to

$$p_{12}(t)/p_2 = c + a(t)$$

and $a(t)$ may be measured directly.

As an example of the model (4) we mention the $G/G/\infty$ queue with N_1 referring to the process of exit times, N_2 to the process of entry times, $a(-u)$ referring to the density of service times and $\mu = 0$. Clearly, here

$$\text{Prob}\{\text{customer leaves in the interval } (t,t+h] \mid N_2(u), -\infty < u < \infty\} \sim \{\sum_j a(t-\tau_j)\}h$$

An interesting problem is that of measuring the degree of association of two point processes. A measure suggested by the preceding model is the <u>coherence</u>

$$|R_{12}(\lambda)|^2 = |f_{12}(\lambda)|^2/(f_{11}(\lambda)f_{22}(\lambda))$$

see Brillinger (1974a). This parameter also appears as a measure of the degree of linear predictability of the process N_1 by the process N_2. It satisfies $0 \le |R_{12}(\lambda)|^2 \le 1$. Other measures of association could be based on the nearness of the function $p_{12}(u) - p_1p_2$ to 0.

We mention next the <u>self-exciting processes</u> introduced by Hawkes, see Hawkes (1972). For r = 1, these satisfy

$$\text{Prob}\{\text{point in } (t,t+h] \mid N(u),\ u \le t\}$$
$$\sim \{\mu + \int_{-\infty}^{t} a(t-u)\ dN(u)\}h$$
$$\sim \{\mu + \sum_{\tau_j \le t} a(t-\tau_j)\}h$$

If we have more than one process, then we could also set up multivariate linear models and define partial parameters. As another extension, we could consider non-linear models such as

$$\text{Prob}\{\text{point of type 1 in } (t,t+h] \mid N_2(u),\ -\infty<u<\infty\}$$
$$\sim\{a_0 + \int a_1(t-u)\ dN_2(u) + \iint_{u \ne v} a_2(t-u,t-v)\ dN_2(u)\ dN_2(v)\}h \quad .$$

More details concerning such extensions may be found in Brillinger (1974b).

9. We end by mentioning that some, possibly unexpected, relationships exist between certain of the parameters that have been defined. These are the Palm-Khinchin relations,

$$\text{Prob}\{N(t) \le j\} = p \int_t^{\infty} \text{Prob}\{N(u) = j \mid N\{0\} = 1\}du$$
$$= 1 - p\int_0^t \text{Prob}\{N(u) = j \mid N\{0\} = 1\}du$$

$$\text{Prob}\{N(t) > j \mid N\{0\} = 1\} = 1 + D^+\{p^{-1} \sum_{j=0}^{j} (j+1-k) \cdot \text{Prob}\{N(t) = k\}\}$$

$$E\{N(t)(N(t)-1) \ldots (N(t) - k)\} =$$

$$= (k+1)\ p \int_0^t E\{N(u)\ (N(u) - 1)\ \ldots\ (N(u) - k + 1)\ |\ N\{0\} = 1\}\ du$$

Such relationships are discussed in Cramér, Leadbetter and Serfling (1971).

In this first section of the paper we have sought to provide a framework within which stationary point processes may be handled when the only element of statistical independence is asymptotic.

II. Estimation of Time Domain Parameters for Stationary Processes

We consider the estimation of certain time domain parameters given a realization of a process $\underset{\sim}{\mathbf{N}}(t)$ over the interval $(0,T]$, i.e. given the observed times of events in $(0,T]$. We begin with the first order mean intensities p_a, $a = 1,\ldots,r$.

1. Obvious estimates of the p_a, $a = 1,\ldots,r$, are the

$$\hat{p}_a = N_a(T)/T$$

$a = 1,\ldots,r$. In connection with these we have,

<u>Theorem 1</u>. Suppose the process satisfies Assumption I. Then $[\hat{p}_1,\ldots,\hat{p}_r]$ is asymptotically

$$N_r([p_1,\ldots,p_r]\ ;\ 2\pi\ T^{-1}[f_{ab}(0)])$$

as $T \to \infty$.

This theorem, as are those given later, is proved in the final section of the paper. The estimates are asymptotically normal. The asymptotic variance of $\hat{p}_a$ is $2\pi T^{-1} f_{aa}(0)$. Were increments of the process uncorrelated, this would be $T^{-1} p_a$. We will see how to estimate $f_{aa}(\lambda)$ next section. Were T large, we might set T = JU and take

$$(TU(J-1))^{-1} \sum_{j=0}^{J-1} (N_a(jU,(j+1)U] - N(0,T]/J)^2 \sim 2\pi T^{-1} f_{aa}(0) \chi^2_{J-1}/(J-1) .$$

The ratio $2\pi f_{aa}(0)/p_a$ is useful in describing certain aspects of the process N_a. If it is greater than 1, the process is said to be <u>clustered</u> or <u>underdispersed</u>. If it is less than 1, the process is called <u>overdispersed</u>.

2. In the second order case we are interested in estimating

$p_{ab}(u) \sim$ Prob{type a in $(t+u, t+u+h_1]$ and type b in $(t,t+h_2]\}/(h_1 h_2)$ for $u \neq 0$ and

$p_{ab}(u)/p_b \sim$ Prob{type a in $(t+u,t+u+h] \mid$ type b at $t\}/h$ for $u \neq 0$.

It seems natural to base estimates of these on

$$J^T_{ab}(u) = \#\{(j,k) \text{ such that } u - \beta < \tau^a_j - \tau^b_k < u + \beta \text{ and } \tau^a_j \neq \tau^b_k\} \qquad (7)$$

for some small bin width $2\beta > 0$. On the CDC 6400, this statistic can be computed about twice as fast as a direct convolution based on $N(T)$ values.

In connection with this variate we have,

Theorem 2. Suppose the process $\underset{\sim}{N}$ satisfies Assumption I and that $p_{ab}(\cdot)$ is a continuous function for $a,b = 1,\ldots,r$. Suppose $J_{ab}^T(u)$ is given by (7) with $\beta = \beta_T$ depending on T. Suppose $u_k^T \to u_k$ with $|u_k^T - u_{k'}^T| \geq 2\beta_T$ for $1 \leq k < k' \leq K$. Then as $T \to \infty$, (i) if $\beta_T = L/T$, L fixed, the variates $J_{a_1b_1}^T(u_1^T),\ldots,J_{a_Kb_K}^T(u_K^T)$ are asymptotically independent Poissons with means $2\beta_T T\, p_{a_kb_k}(u_k)$, $k = 1,\ldots,K$ and (ii) if $\beta_T \to 0$, but $\beta_T T \to \infty$, the variates are asymptotically independent normals with variances $2\beta_T T\, p_{a_kb_k}(u_k)$, $k = 1,\ldots,K$.

The two asymptotic distributions are consistent for large $\beta_T T$, because a Poisson variate with large mean is approximately normal. The result in (i) is not unexpected because we are counting "rare" events. It is surprising that such a general result is so simple however.

The theorem leads us to estimate $p_{ab}(u)$ by

$$\hat{p}_{ab}(u) = J_{ab}^T(u)/(2\beta_T T)$$

and to approximate the distribution of this variate by

$$(2\beta_T T)^{-1}\, P(2\beta_T T p_{ab}(u)) \text{ or } N(p_{ab}(u),\ (2\beta_T T)^{-1} p_{ab}(u)),$$

where $P(\mu)$ here denotes a Poisson distribution with mean μ. This estimate should prove reasonable so long as $|u| \ll T$. In the case that u has noticeable magnitude compared to T it might be better to replace $J^T_{ab}(u)$ by $T\, J^T_{ab}(u)\,/(T - |u|)$ or by

$$J^T_{ab}(u) + p_a p_b |u| 2\beta_T \qquad (8)$$

The use of the variate of (8) is suggested by the usual estimate of the autocovariance function of an ordinary time series. Its construction is based on the observation that $q_{ab}(u) \to 0$ as $|u| \to \infty$ for many processes. It should have better overall mean-squared error properties for such processes.

We remark that we are here essentially carrying out histogram construction. Considerations of that topic are relevant here. For example, we might choose to construct a rootogram based on $\sqrt{J^T_{ab}(u)}$ to get stable variance. (If there may be some cells with low counts, we might follow Tukey and use $\sqrt{2 + 4\, J^T_{ab}(u)}$). The variate $\sqrt{\hat{p}_{ab}(u)}$ will have approximately stable variance of $(8\beta_T T)^{-1}$. The theorem likewise leads us to estimate $p_{ab}(u)/p_b$ by $J^T_{ab}(u)/(2\beta_T N_b(T))$ and to approximate the distribution of this estimate by

$$(2\beta_T T\, p_b)^{-1}\, P(2\beta_T T\, p_{ab}(u)) \text{ or } N(p_{ab}(u)/p_b,$$
$$(2\beta_T T\, p_b^2)^{-1} p_{ab}(u)).$$

The variance of $\sqrt{J^T_{ab}(u)/(2\beta_T N_b(T))}$ will be approximately stable and may be estimated by $(8\beta_T N_b(T))^{-1}$.

The above results may be used to set

approximate confidence intervals and multiple confidence intervals for the estimates. In the case that the increment of the process N_a is independent of the increment of the process N_b, u time units away, $p_{ab}(u)/p_b = p_a$. We may examine this hypothesis by plotting

$$\sqrt{J^T_{ab}(u)/(2\beta_T N_b(T))} \ , \ \sqrt{\hat{p}_a} \ , \ \sqrt{J^T_{ab}(u)/(2\beta_T N_b(T))} \ \pm \ (2\beta_T N_b(T))^{-\frac{1}{2}}$$

on the same graph for example. This sort of graph is useful in checking for some degree of association between the process N_a and the process N_b.

What we have been doing may be viewed as estimating the probability density function of the times between a events and b events from the observed differences

$$\tau^a_j - \tau^b_k \ ; \ 0 < \tau^a_j \ , \ \tau^b_k \le T.$$

Cox (1965) suggested that one could also consider "window estimates." Let W(u) be bounded and absolutely integrable. Let $W^T(u) = W(u/\beta_T)$ for the sequence of scale factors β_T, T = 1,2,... . It is now natural to base estimates on

$$J^T_{ab}(u) = \sum_{0 < \tau^a_j \neq \tau^b_k \le T} W^T(u - \tau^a_j + \tau^b_k)$$
$$= \iint_{0<\tau\neq\sigma\le T} W^T(u-\sigma+\tau)\, dN_a(\sigma)\, dN_b(\tau)$$

(The previous $J^T_{ab}(u)$ corresponds to W(u) = 1 for

$|u| < 1$.) The variances of the asymptotic distributions of (ii) of Theorem 2 are now replaced by $\beta_T T \int W(u)^2 du\ p_{ab}(u_k)$, $k = 1,\ldots,K$. By direct computation we see that

$$E\ J_{ab}^T(u) = \int_{-T}^{T} (T - |\rho|) W^T(u - \rho)\ p_{ab}(\rho) d\rho$$

$$\sim \beta_T (T - |u|) \{p_{ab}(u) \int W(\rho) d\rho - \beta_T p'_{ab}(u) \int \rho W(\rho) d\rho + \beta_T^2 p''_{ab}(u) \int \rho^2 W(\rho) d\rho / 2 + \ldots\}$$

suggesting that bias may become a problem when $p_{ab}(\rho)$ varies substantially in the neighborhood of u or when u is of appreciable magnitude compared to T. We have already discussed one modification to handle this last case.

The asymptotic distribution determined in Theorem 2 is an unconditional one. In practise the worker may feel that the conditional distribution, conditional on the observed $N_a(T)$, $N_b(T)$ is the appropriate one. In I.4 we set down the form of product densities in the conditional case. It should be possible to make use of these to determine the form of the large sample conditional distribution.

Cox and Lewis (1972) discuss some aspects of the problem of estimating second-order product densities for a vector-valued process.

3. In the k-th order case we might consider the statistic

$$J^T_{a_1\cdots a_k}(u_1,\ldots,u_{k-1}) = \int_0^T \underset{\neq}{\cdots} \int_0^T W^T(u_1-\sigma_1+\sigma_k,\ldots,$$

$$u_{k-1}-\sigma_{k-1}+\sigma_k)\, dN_{a_1}(\sigma_1)\cdots dN_{a_k}(\sigma_k) \quad (9)$$

where $W^T(u_1,\ldots,u_{k-1}) = W(u_1/\beta_T,\ldots,u_{k-1}/\beta_T)$ and the $\neq$ in (9) indicates that the range of integration is over distinct σ_j.

Theorem 3. Suppose the process $\underset{\sim}{N}$ satisfies Assumption I and that $p_{a_1\cdots a_k}(\cdot)$ is continuous at $(u_1,\ldots,u_{k-1})$. Then as $T \to \infty$, (i) if $\beta_T^{k-1}T = L$, L fixed, if $W(u_1,\ldots,u_{k-1}) = 1$ for $|u_j| < 1$, the variate of (9) is asymptotically Poisson with mean $(2\beta_T)^{k-1}T\, p_{a_1\cdots a_k}(u_1,\ldots,u_{k-1})$ and (ii) if $\beta_T \to 0$, but $\beta_T^{k-1}T \to \infty$, the variate is asymptotically normal with mean

$$T\int_{-T}^{T}\cdots\int_{-T}^{T} W^T(u_1-\rho_1,\ldots,u_{k-1}-\rho_{k-1})\,p_{a_1\cdots a_k}$$

$$(\rho_1,\ldots,\rho_{k-1})\,d\rho_1\cdots d\rho_{k-1} \qquad (10)$$

and variance $\beta_T^{k-1}T \int W^2\, p_{a_1\cdots a_k}(u_1,\ldots,u_{k-1})$.

The integral of (10) may be expected to be near

$$\beta_T^{k-1}\, T^{k-1} \int W\, p_{a_1\cdots a_k}(u_1,\ldots,u_{k-1})$$

suggesting the consideration of the estimate

$$\hat{p}_{a_1\cdots a_k}(u_1,\ldots,u_{k-1}) = J^T_{a_1\cdots a_k}(u_1,\ldots,u_{k-1})/$$

$$(\beta_T^{k-1}\, T^{k-1} \int W\,)$$

4. Let A denote the interval (0,T] and suppose that the points observed in A are: γ_1 of type 1 at $t_1,\ldots, t_{\gamma_1}$; ... ; γ_r of type r at $\ldots, t_k$. Then, using the expressions of Section I, the likelihood function here is B/C where

$$B = \sum_{\nu_1 \geq \gamma_1} \cdots \sum_{\nu_r \geq \gamma_r} \frac{(-1)^{\nu_1-\gamma_1+\ldots+\nu_r-\gamma_r}}{(\nu_1-\gamma_1)!\ldots(\nu_r-\gamma_r)!}$$

$$\int_{A^{\nu_1-\gamma_1+\ldots+\nu_r-\gamma_r}} p^{(\nu_1)\ldots(\nu_r)}$$

$$(t_1,\ldots,t_{\gamma_1};\ldots:t_{\gamma_1+1},\ldots,t_{\gamma_1+\gamma_2};\ldots:\ldots$$

and

$$C = \frac{1}{\gamma_1!\ldots\gamma_r!} \sum_{\nu_1 \geq \gamma_1} \cdots \sum_{\nu_r \geq \gamma_r} \frac{(-1)^{\nu_1-\gamma_1+\ldots+\nu_r-\gamma_r}}{(\nu_1-\gamma_1)!\ldots(\nu_r-\gamma_r)!}$$

$$\int_{A^{\nu_1+\ldots+\nu_r}} p^{(\nu_1)\ldots(\nu_r)}$$

Let us consider the approximate value of the likelihood function, B/C, for large T. In the case of large T

$$\int_{A^{\nu_1-\gamma_1+\ldots+\nu_r-\gamma_r}} p^{(\nu_1)\ldots(\nu_r)}(t_1,\ldots t_{\gamma_1};\ldots$$

$$\sim p^{(\gamma_1)\ldots(\gamma_r)}(t_1,\ldots,t_{\gamma_1+\ldots+\gamma_r})$$

$$T^{\nu_1-\gamma_1+\ldots+\nu_r-\gamma_r}\, p_1^{\nu_1-\gamma_1} \ldots p_r^{\nu_r-\gamma_r}$$

suggesting that for large T, the likelihood function is approximately

$$\gamma_1! \ldots \gamma_r! \; p^{(\gamma_1)\ldots(\gamma_r)}(t_1, \ldots, t_{\gamma_1+\ldots+\gamma_r})$$
$$(Tp_1)^{-\gamma_1} \ldots (Tp_r)^{-\gamma_r}$$

4. In this section we will propose estimates of the parameters described in Section I.5 in the case that a realization of a stationary process is available for the time interval (0,T].

(i) We begin with the renewal function,

$$U_{ab}(t) = E\{N_a(t) \mid N_b\{0\} = 1\} = \int_0^t p_{ab}(u)\,du\,/p_b$$

A natural estimate to consider is

$$\hat{U}_{ab}(t) = \#\{t \geq \tau_j^a - \tau_k^b > 0\}/N_b(T)$$
$$= \int_0^{T-t} \int_0^t dN_a(u+w)\; dN_b(u)/N_b(T)$$

To determine the asymptotic distribution of $\hat{U}_{ab}(t)$ we will need the joint asymptotic distribution of $\#\{\cdot\}$ and $N_b(T)$. It is fairly clear that under Assumption I, the variate is asymptotically normal with asymptotic variance that is $O(T^{-1})$. However the form of the asymptotic variance seems very messy. In practise one would probably have to estimate it by segmenting the data.

(ii) Let us next estimate the survivor function

$$1 - F(t) = \text{Prob}\{N(t) = 0 \mid N\{0\} = 1\}$$
$$= \text{Prob}\{\tau_{i+1} - \tau_i > t\}$$
$$= 1 - \text{Prob}\{\tau_{i+1} - \tau_i \leq t\}$$

This last suggests the estimate

$$\hat{F}(t) = \#\{\tau_{i+1} - \tau_i \leq t \; ; \; i = 1,\ldots, N(T)-1\}/N(T)$$

This estimate is based on the interarrival times $X_i = \tau_{i+1} - \tau_i$. The process X_i, $i = 0,\pm 1,\ldots$ is stationary. If it is mixing in some sense then $1 - \hat{F}(t)$ will be asymptotically normal, see Deo (1973), for example. This last suggests the interesting problem of relating a mixing condition for a stationary point process to some mixing condition for the corresponding process of interarrival times.

(iii) The following is a plausible estimate for the hazard function, with β_T a small positive number,

$$\hat{\mu}(t) = \frac{\#\{t-\beta_T < \tau_{i+1} - \tau_i < t + \beta_T \; ; \; i = 1,\ldots,N(T) - 1\}}{2\,\beta_T \#\{\tau_{i+1} - \tau_i > t \; ; \; i = 1,\ldots,N(T) - 1\}}$$

(iv) Next consider the estimation of the forward recurrence time distribution

$$\begin{aligned} G(t) &= 1 - \operatorname{Prob}\{N(t) = 0\} \\ &= p \int_0^t (1 - F(u))\, du \\ &= p[(1 - F(t))t] + p\int_0^t u \, dF(u) \end{aligned}$$

where we use a Palm-Khinchin relation from Section I.9 and integrate by parts. The last relation suggests the estimate

$$\hat{G}(t) = \hat{p}[(1 - \hat{F}(t))t] + \hat{p} \sum_{\tau_{i+1}-\tau_i \le t} (\tau_{i+1} - \tau_i)/(N(T)-1)$$

$$\sim t\ \#\{\tau_{i+1} - \tau_i > t\}/T + \sum_{\tau_{i+1}-\tau_i \le t} (\tau_{i+1} - \tau_i)/T$$

III. Estimation of Frequency Domain Parameters

1. We begin with a discussion of first order statistics. Suppose T = JU, J an integer. Set

$$d_a^U(\lambda, j) = \int_{jU}^{(j+1)U} \exp\{-i\lambda t\}\, dN_a(t) \qquad j = 0, \ldots, J-1$$

$$= \sum_{jU < \tau_a \le (j+1)U} \exp\{-i\lambda\tau_a\}$$

$$= \int \exp\{-i(\lambda-\alpha)(j+\tfrac{1}{2})\}(\sin(\lambda-\alpha)U/2)/((\lambda-\alpha)/2)\, dZ_a(\alpha)$$

using the spectral representation at the last step. In the case that J = 1, U = T, we shall write $d_a^T(\lambda)$. We have,

Theorem 4. Let the process $\underset{\sim}{N}(t)$ satisfy Assumption I. Suppose $\lambda \ne 0$. Then $\underset{\sim}{d}^U(\lambda, j)$, $j = 0, \ldots, J-1$ are asymptotically independent r variate complex normal with mean $\underset{\sim}{0}$ and covariance matrix $2\pi U[f_{ab}(\lambda)]$ as $T \to \infty$. Also variates at frequencies of the form $2\pi u/U$, are asymptotically independent for u distinct positive integers.

2. Suppose we are interested in estimating the second order spectrum $f_{ab}(\lambda)$. Various procedures

suggest themselves, based either on the expression

$$E\{dZ_a(\lambda)\ dZ_b(\mu)\} = \delta(\lambda-\mu)\ f_{ab}(\lambda)\ d\lambda\ d\mu \quad \lambda,\mu \neq 0$$

or the expression

$$\{p_a + \int_{-\infty}^{\infty} \{p_{aa}(u) - p_a^2\} \exp\{-i\lambda u\}\ du\}/(2\pi) = f_{aa}(\lambda)$$

$$\{\int_{-\infty}^{\infty} \{p_{ab}(u) - p_a p_b\} \exp\{-i\lambda u\}\ du\}/(2\pi) = f_{ab}(\lambda)\ a \neq b$$

<u>Procedure I.</u> Set $\underset{\sim}{I}^U(\lambda,j) = (2\pi U)^{-1} \underset{\sim}{d}^U(\lambda,j)\overline{\underset{\sim}{d}^U(\lambda,j)}^{\tau}$ for $\lambda \neq 0$ and consider the estimate

$$\underset{\sim}{f}^U(\lambda) = J^{-1} \sum_{j=0}^{J-1} \underset{\sim}{I}^U(\lambda,j)$$

From Theorem 4, as $T \to \infty$, but J remains fixed $\underset{\sim}{f}^U(\lambda)$ tends to $J^{-1}W_r^C(J, \underset{\sim}{f}(\lambda))$ where W_r^C denotes the complex Wishart.

<u>Procedure II.</u> Set $\underset{\sim}{I}^T(\lambda) = (2\pi T)^{-1} \underset{\sim}{d}^T(\lambda)\overline{\underset{\sim}{d}^T(\lambda)}^{\tau}$. For $2\pi s_j/T$ distinct, $\neq 0$, non-negative and all $\to \lambda$ set

$$\underset{\sim}{f}^T(\lambda) = J^{-1} \sum_{j=0}^{J-1} \underset{\sim}{I}^T(2\pi s_j/T).$$

From Theorem 4, as $T \to \infty$, $\underset{\sim}{f}^T(\lambda)$ tends to $J^{-1}W_r^C(J, \underset{\sim}{f}(\lambda))$

Both of the above estimates are asymptotically normal if the limiting conditions are as $T \to \infty$, $J \to \infty$, but $J/T \to 0$.

In the above procedures we sometimes choose to weight the periodogram ordinates unequally. For example in Procedure II we might take

$$\underset{\sim}{f}^T(\lambda) = \frac{2\pi}{T} \sum_{s\neq 0} W^T(\lambda - \frac{2\pi s}{T})\ \underset{\sim}{I}^T(\frac{2\pi s}{T})$$

with $W^T(\alpha) = B_T^{-1}W(\alpha/B_T)$ where $\int W = 1$. If $B_T \to 0$, $B_T T \to \infty$ as $T \to \infty$, this estimate is asymptotically normal, see Brillinger (1972).

Procedure III. Let $\hat{p}_a$, $\hat{p}_{ab}(u)$ be given by the expressions of II.1, II.2 respectively. Let $w^T(u) = w(B_T u)$ be a convergence factor. Set

$$\begin{aligned} f_{ab}^T(\lambda) &= \{2\beta_T \sum_j \{\hat{p}_{ab}(2\beta_T j) - \hat{p}_a\hat{p}_b\}\exp\{-i\lambda 2\beta_T j\} \\ &\qquad w^T(2\beta_T j)\}/(2\pi) \qquad a \neq b \\ &= \{\hat{p}_a + 2\beta_T \sum_j \{\hat{p}_{aa}(2\beta_T j) - \hat{p}_a^2\}\exp\{-i\lambda 2\beta_T j\} \\ &\qquad w^T(2\beta_T j)\}/(2\pi) \qquad a = b \end{aligned}$$

Because of the periodicities involved, it only makes sense to compute this estimate for $|\lambda| \leq \pi/\beta_T$. The choice of bin width $2\beta_T$ is seen to show itself in the Nyquist frequency π/β_T. This estimate is asymptotically normal under conditions including B_T, $\beta_T \to 0$, $B_T T \to \infty$ as $T \to \infty$. This estimate is the one computed most rapidly. It has the disadvantage of possibly leading to negative power spectrum estimates and coherences bigger than 1, even if $W(\alpha) \geq 0$.

Procedure IV. Compute the spectrum of the ordinary process

$$X(t) = h^{-1}N(t - \tfrac{h}{2}, t + \tfrac{h}{2}) \quad t=0,\pm h,\pm 2h,\ldots$$

but remember that

$$f_{XX}(\lambda) = \sum_j f_{NN}(\lambda + \frac{2\pi j}{h})(\sin(\lambda + \frac{2\pi j}{h}))^2/(\lambda + \frac{2\pi j}{h})^2$$

Problems of aliasing clearly arise here.

Tapering and prefiltering play essential roles in the estimation of the spectra of ordinary time series. It is not entirely obvious how to apply these techniques in the point process case (with the exception of tapering for Procedures I and II.)

If the complete intensity

$$\lambda(t)h \sim \text{Prob}\{\text{point in } (t,t+h)] \mid N(u),\ u \leq t\}$$

exists and can be evaluated, then with

$$\Lambda(t) = \int_0^t \lambda(t)\, dt$$

the transformation $N(t) \to N(\Lambda(t))$ carries N over into a Poisson process with unit intensity, and constant power spectrum. (This transformation is analagous to the conditional probability integral transformation to uniform variates in the case of ordinary time series.) For the doubly stochastic Poisson process $\lambda(t) = x(t)$.

Prefiltering procedures carried out entirely in the frequency domain, for ordinary time series, clearly have point process analogs. For example, if we can think of a $g(\lambda)$ near $f(\lambda)$, then we might form the estimate

$$g(\lambda)\ \frac{2\pi}{T} \sum_{s \neq 0} W^T(\lambda - \frac{2\pi s}{T})\ g(\frac{2\pi s}{T})^{-1}\ I^T(\frac{2\pi s}{T})$$

Detrending can be very important. Lewis (1972) contains important advice on these matters.

3. We next turn to a brief discussion of the estimation of the parameters of the model

$$\text{Prob}\{\text{type 1 event in } (t,t+h] \mid N_2(u),\ u \le t\} \sim (\mu + \int a(t-u)\ dN_2(u))h$$

as $h \downarrow 0$. If $p_{22}(u)$ is not constant, then we estimate $a(u)$, a time domain parameter by going through the frequency domain. We have the relations

$$\begin{aligned} A(\lambda) &= \int a(u) \exp\{-i\lambda u\}du \\ p_1 &= \mu + A(0)p_2 \\ f_{12}(\lambda) &= A(\lambda)\ f_{22}(\lambda) \\ a(u) &= (2\pi)^{-1} \int A(\alpha) \exp\{iu\alpha\}d\alpha \end{aligned}$$

suggesting the estimates

$$\begin{aligned} \hat{A}(\lambda) &= f_{12}^T(\lambda)/f_{22}^T(\lambda) \\ \hat{\mu} &= \hat{p}_1 - \hat{A}(0)\ \hat{p}_2 \end{aligned}$$

$$\hat{a}(u) = (2\pi)^{-1} B_T \sum_k \hat{A}(kB_T) \exp\{iukB_T\}\ v^T(kB_T)$$

where $v^T(\alpha) = v(C_T\alpha)$ is a convergence factor. More details on this procedure may be found in Brillinger (1974 a).

4. On occasion we may be led to model the process in some manner involving a finite dimensional parameter θ. We would then like to be able to estimate θ. Sometimes such a model will lead to a tractible form

for the second-order spectra. For example, suppose we have a cluster process with primary process Poisson and the secondary process independent exponentials from the cluster centers, then the power spectrum of the process has the form

$$f(\lambda;\theta) = a^2\frac{(b^2 + \lambda^2)}{(c^2 + \lambda^2)}$$

involving the three dimensional parameter $\theta = (a,b,c)$. We now describe one method of estimating θ. Related methods are given in Whittle (1953), Walker (1964), Hawkes and Adamopoulos (1973).

Let the true value of θ be θ'. Suppose

$$\lim_{|\lambda| \to \infty} f(\lambda;\theta) = \mu(\theta)$$

and $\mu(\theta') = p/(2\pi)$ where p is the mean intensity of the process. $\mu(\theta')$ may be estimated by $\hat{\mu} = \hat{p}/(2\pi)$.

The periodograms $I^T(2\pi s/T)$, $s = 1,2,\ldots$ are asymptotically independent exponentials with means $f(2\pi s/T;\ \theta')$, $s = 1,2,\ldots$. The scaled variates

$$I^T(2\pi s/T)/\hat{\mu} \qquad s = 1,2,\ldots$$

are therefore asymptotically independent exponentials with means

$$f(2\pi s/T;\ \theta')/\mu(\theta') \qquad s = 1,2,\ldots$$

This result suggests our setting down the following approximate "log likelihood" function

$$-\sum_{s=1}^{S_T}\left\{\log\frac{f(2\pi s/T;\theta)}{\mu(\theta)}+\frac{\mu(\theta)}{f(2\pi s/T;\theta)}\frac{I^T(2\pi s/T)}{\hat{\mu}}\right\} \quad (11)$$

and taking as an estimate of θ, the value $\hat{\theta}$ that maximizes (11).

In the theorem below we set $g(\lambda;\ \theta) = f(\lambda;\ \theta)/\mu(\theta)$ and

$$\Lambda^T(\theta) = -\frac{2\pi}{T}\sum_{s=1}^{S_T}\left\{\log g(2\pi s/T;\ \theta)+\frac{I^T(2\pi s/T)/\hat{\mu}}{g(2\pi s/T;\ \theta)} - 1\right\} \quad (12)$$

The $\hat{\theta}$ maximizing (11) also maximizes (12).

<u>Theorem 5</u>. If (a) the process $N(t)$, $-\infty < t < \infty$, has mean intensity $\mu(\theta')$ and power spectrum $f(\lambda;\ \theta')$, (b) $f(\lambda;\theta)$, $\theta \in \Theta \subset R^L$, is non-negative and

$$\mu(\theta) = \lim_{|\lambda|\to\infty} f(\lambda;\ \theta)$$

exists, (c) with $g(\lambda;\ \theta) = f(\lambda;\ \theta)/\mu(\theta)$,

$$\Lambda(\theta) = -\int\{\log g(\lambda;\ \theta) + \frac{g(\lambda;\ \theta')}{g(\lambda;\ \theta)} - 1\}\ d\lambda$$

exists as a Lebesgue integral, has a unique maximum at θ' and is such that

$$\max_{\theta''\in U} \Lambda(\theta'') \to \Lambda(\theta)$$

as the neighborhood U of θ shrinks to $\{\theta\}$, (d) $\Lambda^T(\theta) \xrightarrow{P} \Lambda(\theta)$ at θ' and uniformly near other θ (e) $\theta\in\Theta$ maximizing (11) is bounded in probability, then $\hat{\theta} \to \theta'$ in probability as $T \to \infty$, Condition (d) is satisfied for processes satisfying Assumption I, provided $g(\lambda;\theta)$ is a sufficiently regular function of λ.

We next turn to the large sample distribution of $\hat{\theta}$. To this end set,

$$\Lambda_j^T(\theta) = \frac{\partial}{\partial\theta_j}\Lambda^T(\theta) \qquad \Lambda_{jk}^T(\theta) = \frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda^T(\theta) \qquad (13)$$

Because of (c) above, generally $\Lambda_j(\theta') = \partial\Lambda(\theta')/\partial\theta_j = 0$.

Theorem 6. Suppose the conditions of Theorem 5 are satisfied. Suppose also (f) the derivatives of (13) exist, (g) $\Lambda_{jk}^T(\zeta^T) \to A_{jk}$ for any sequence ζ^T of variates tending to θ' in probability, (h) with $\underset{\sim}{A} = [A_{jk}]$, $\sqrt{T}\{\Lambda_1^T(\theta'),\dots,\Lambda_L^T(\theta')\} \to N_L(\underset{\sim}{0}, \underset{\sim}{A}+\underset{\sim}{B})$, then $\hat{\theta}$ is asymptotically normal with mean θ' and covariance matrix $T^{-1}\underset{\sim}{A}^{-1}(\underset{\sim}{A}+\underset{\sim}{B})\underset{\sim}{A}^{-1}$.

For processes satisfying Assumption I and $g(\lambda;\theta)$ a sufficiently regular function of λ we have

$$A_{jk} = \int_0^\infty \frac{\partial \log g(\alpha;\theta')}{\partial\theta_j'}\frac{\partial \log g(\alpha;\theta')}{\partial\theta_k'}\,d\alpha$$

$$B_{jk} = 2\pi\int_0^\infty\int_0^\infty \frac{\partial \log g(\alpha;\theta')}{\partial\theta_j'}\frac{\partial \log g(\beta;\theta')}{\partial\theta_k'}\frac{f^{(4)}(\alpha,-\alpha,-\beta;\theta')}{f(\alpha;\theta')\,f(\beta;\theta')}\,d\alpha\,d\beta$$

The above procedure provides us with a further estimate, $\hat{f}(\lambda) = f(\lambda;\hat{\theta})$ of the power spectrum. Under the conditions of the theorem, this estimate will be asymptotically normal with mean $f(\lambda;\theta')$ and variance

$$\sum_{j,k} \frac{\partial f(\lambda;\theta')}{\partial\theta'_j} \frac{\partial f(\lambda;\theta')}{\partial\theta'_k} \operatorname{cov}\{\hat{\theta}_j, \hat{\theta}_k\}$$

In the case of a vector-valued process, instead of maximizing (11) we would maximize

$$-\sum_{s=1}^{S} \{\log \operatorname{Det} \underset{\sim}{g}(2\pi s/T;\theta) + \operatorname{tr}(\underset{\sim}{H}^T(2\pi s/T)\, \underset{\sim}{g}(2\pi s/T;\theta))\}$$

where

$$g(\lambda;\theta)_{jk} = f(\lambda;\theta)_{jk}/\sqrt{\mu_j(\theta)\,\mu_k(\theta)}$$

$$H^T(\lambda)_{jk} = I^T(\lambda)_{jk}/\sqrt{\hat{\mu}_j\hat{\mu}_k}$$

5. We mention briefly that the parameters of a self-exciting process may be estimated via a frequency domain analysis. Such a process is defined by a relationship

$$E\{dN(t) \mid N(u),\ u \le t\} = (\mu + \int_{-\infty}^{t} a(t-u)dN(u))\ dt$$

where μ, $a(u) \ge 0$; $\int a(u)\ du < 1$; $a(u) = 0$ for $u < 0$. Let

$$A(\lambda) = \int_0^\infty a(u) \exp\{-i\lambda u\}du$$

For this process $\mu = p[1 - A(0)]$ and

$$f(\lambda) = p/(2\pi|1 - A(\lambda)|^2)$$

Because $A(\lambda)$ is the Fourier transform of a one-sided function, the problem of estimating $A(\lambda)$ from $f^T(\lambda)$, is seen to involve the factorization of $f^T(\lambda)$. Rice (1973) carried out this empirically and found the asymptotic distribution of the estimate. This procedure also provides a further spectral estimate, namely $\hat{f}(\lambda) = \hat{p}/(2\pi|1-\hat{A}(\lambda)|^2)$.

6. We next turn to the problem of estimating the variance time curve given by var N(t) as a function of t. Using the spectral representation, we see that

$$V(t) = \operatorname{var} N(t) = \int_{-\infty}^{\infty}\left(\frac{\sin \alpha t/2}{\alpha/2}\right)^2 f(\alpha)\, d\alpha$$
$$= tp + \int_{-\infty}^{\infty}\left(\frac{\sin \alpha t/2}{\alpha/2}\right)^2 \left(f(\alpha) - \frac{p}{2\pi}\right) d\alpha$$

The following type of estimate is considered by Torres-Melo (1974),

$$V(t) = t\hat{p} + B\left[t^2\left(f^T(0) - \frac{\hat{p}}{2\pi}\right) + 2\sum_{s=1}^{S}\left(\frac{\sin Bst/2}{Bs/2}\right)^2 \left(f^T(Bs) - \frac{\hat{p}}{2\pi}\right)\right]$$

He finds the asymptotic distribution of this estimate.

7. Product densities may be estimated in a similar manner to the variance time curve. We have

$$p(u) = \int_{-\infty}^{\infty}\left(f(\alpha) - \frac{p}{2\pi}\right)\exp\{iu\alpha\}\, d\alpha + p^2$$

suggesting the estimate

$$\hat{p}(u) = B\left[\left(f^T(0) - \frac{\hat{p}}{2\pi}\right) + 2\sum_{s=1}^{S}\left(f^T(Bs) - \frac{\hat{p}}{2\pi}\right)\cos Bs\right]$$

This estimate would undoubtedly be improved by the insertion of convergence factors.

Finally we remark that we may sometimes wish to estimate the spectral measure

$$F(\lambda) = \int_0^{\lambda} f(\alpha)\, d\alpha$$

The obvious estimate is

$$\hat{F}(\lambda) = B \sum_{Bs \leq \lambda} f^T(Bs)$$

IV. Proofs

1. Proof of Theorem 1. The joint factorial cumulant of $N_{a_1}(T),\ldots,N_{a_k}(T)$ is

$$\int_0^T \cdots \int_0^T q_{a_1\ldots a_k}(t_1,\ldots,t_k)dt_1\ldots dt_k = O(T)$$

in view of Assumption I. The ordinary joint cumulant of these same variates is a sum of multiples of lower order factorial cumulants. It follows that it too is $O(T)$ as $T \to \infty$. This means that the standardized joint cumulants of order k of these variates are $O(T^{1-k/2}) \to 0$ as $T \to \infty$ for $k > 2$, and so the variates are asymptotically jointly normal.

2. Proof of Theorem 2. The variate $J^T_{ab}(u)$ may be represented as

$$\int_G dN_a(\sigma)\, dN_b(\tau)$$

where G is the set $\{u - \beta_t < \sigma - \tau < u + \beta_T, \sigma \neq \tau\}$. It follows from this representation, Assumption I and the rules of Leonov and Shiryaev (1959) that the joint factorial moment of order k of $J_{ab}^T(u)$ is of order $O(\beta_T^k T)$. An ordinary cumulant of order k, c_k, is connected to corresponding factorial cumulants, $c_{(k)}$, through

$$c_k = \sum_{j=1}^{k} S_j^k \, c_{(j)}$$

where S_j^k is a Stirling number. If $\beta_T = L/T$, then $E\, J_{ab}^T(u^T) \to 2L\, p_{ab}(u)$ as $T \to \infty$, when $u^T \to u$. It follows, that in this case the cumulant of order k of $J_{ab}^T(u^T) \to 2L\, p_{ab}(u)$ and so the variate is asymptotically Poisson. In the case $\beta_T T \to \infty$, the standardized joint cumulant of order k is $O(\beta_T T)^{1-k/2} \to 0$ for $k > 2$. It follows that the variate is asymptotically normal. The indicated asymptotic independence follows on evaluating joint second-order cumulants.

3. Theorem 3 is proved in the same manner that Theorem 2 is proved.

4. Theorem 4 is proved by evaluating the joint cumulants of the d_a^U. A related result, Theorem 4.2, is proved in Brillinger (1972).

5. Before proving Theorem 5, we prove a lemma of some independent interest.

Lemma 1. If (i) Θ is locally compact, complete, separable, metric, (ii) $(\Omega, \mathcal{G}, P)$ is a probability space with Ω complete, separable, metric, (iii) $Q^T(\theta,\omega)$ is real-valued, Borel measurable for $(\theta,\omega) \in \Theta \times \Omega$ and all T, (iv) $Q(\theta)$ is real-valued, lower semi-continuous, $Q(\theta) > Q(\theta')$ for $\theta \neq \theta'$, (v)

$Q^T(\theta',\omega) = Q(\theta') + o_p(1)$, $Q^T(\theta,\omega) \geq Q(\theta) + o_p(1)$, $\theta \neq \theta'$ as $T \to \infty$, (vi) given $\varepsilon, \eta > 0$, $\theta_1 \neq \theta'$, there exists U_1 a neighborhood of θ_1 and there exists T_0 such that

$$\text{Prob}\{[Q^T(\theta_1,\omega) - \inf_{\theta \in U_1} Q^T(\theta,\omega)] \leq -\varepsilon\} < \eta \text{ for } T \geq T_0$$

(vii) for each ω and T there exists $\hat{\theta}$ such that

$$Q^T(\hat{\theta},\omega) = \inf_{\theta \in \Theta} Q^T(\theta,\omega)$$

(viii) given $\eta > 0$, there exists a compact set $C \subseteq \Theta$ and T_0 such that $\text{Prob}\{\hat{\theta} \notin C\} < \eta$, for $T > T_0$, then $\hat{\theta} = \theta' + o_p(1)$.

Proof. The measurability of $\hat{\theta}$ results from Theorem 2 of Brown and Purves (1973). Let $U \subseteq C$ be an open neighborhood of θ'. From (iv) there exists $\gamma > 0$ such that $Q(\theta_1) - Q(\theta') \geq 3\gamma$ for $\theta \in C\backslash U$. Suppose $\theta_1 \in C\backslash U$. Then from (v)

$$\lim_{T \to \infty} \text{Prob}\{Q^T(\theta_1,\omega) - Q^T(\theta',\omega) \leq 2\gamma\} = 0$$

From this and (v) there exists a neighborhood U_1 of θ_1 such that

$$\lim_{T \to \infty} \text{Prob}\{\inf_{\theta \in U_1} Q^T(\theta,\omega) - Q^T(\theta',\omega) \leq \gamma\} = 0 \tag{14}$$

Using the fact that C is compact, select a finite number of points θ_s, $s = 1,\ldots,N$ with neighborhoods U_s, $s = 1,\ldots,N$ covering $C\backslash U$. From (14)

$$\lim_{T \to \infty} \mathrm{Prob}\{\inf_{\theta \in C \setminus U} Q^T(\theta,\omega) - Q^T(\theta',\omega) \leq \gamma\} = 0 \quad (15)$$

Now from (vii)

$$\mathrm{Prob}\{\hat{\theta} \notin U \text{ or } \Theta \setminus C\} \leq \mathrm{Prob}\{\inf_{\theta \in C \setminus U} Q^T(\theta,\omega) - Q^T(\theta',\omega) \leq \gamma\}$$

From (15) this last tends to 0. From (viii), $\mathrm{Prob}\{\hat{\theta} \in \Theta \setminus C\}$ tends to 0. This gives the result.

Theorem 5 now follows from this Lemma.

6. Theorem 6 follows from the relation

$$\Lambda_j^T(\theta) = 0 = \Lambda_j^T(\theta') + \sum_k (\theta_k - \theta_k') \Lambda_{jk}^T(\alpha)$$

with α between θ and θ'.

REFERENCES

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. pp. 483-513 in Proc. Sixth Berkeley Symp. Math. Stat. Prob., Vol. I. (eds. L. M. LeCam, J. Neyman, E. L. Scott). Berkeley, University of California Press.

BRILLINGER, D. R. (1974a). Cross-spectral analysis of processes with stationary increments including G/G/∞ queue. Ann. Prob., 2, 815-827.

BRILLINGER, D. R. (1974b). The identification of point process systems. Special Invited Lecture presented to the Institute of Mathematical Statistics at Edmonton.

COX, D. R. (1965). On the estimation of the intensity function of a stationary point process. J. R. Statist. Soc., B, 27. 332-337.

COX, D. R. and LEWIS, P. A. W. (1966). The Statistical Analysis of Series of Events. London, Methuen.

COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. pp. 401-448 in Proc. Sixth Berkeley Symp. Math. Stat. Prob., Vol. II. (eds. L. M. LeCam, J. Neyman, E. L. Scott). Berkeley, University of California Press.

CRAMÉR, H., LEADBETTER, M. R. and SERFLING, R. J. (1971). On distribution function - moment relationships in stationary point processes. Zeit. Wahrschein., 18. 1-8.

DALEY, D. J. and VERE-JONES, D. (1972). A summary of the theory of point processes. pp. 299-383 in STOCHASTIC POINT PROCESSES (ed. P. A. W. Lewis)

New York, Wiley.

DEO, C. M. (1973). A note on empirical processes of strong-mixing sequences. Ann. Prob., 1. 870-875.

HAWKES, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. pp. 261-271 in Stochastic Point Processes (ed. P. A. W. Lewis). New York, Wiley.

HAWKES, A. G. and ADAMOPOULOS, L. (1973). Cluster models for earthquakes - regional comparisons. Bul. Inter. Statist. Inst., 39.

LEONOV, V. P. and SHIRYAEV, A. N. (1959). On a method of calculation of semi-invariants. Theory Prob. Appl., 4. 319-329.

LEWIS, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes. pp. 1-54 in Stochastic Point Processes (ed. P. A. W. Lewis). New York, Wiley.

RICE, J. A. (1973). Statistical analysis of self-exciting point processes and related linear models. Ph.D. Thesis, University of California, Berkeley.

SRINIVASAN, S. K. (1974). Stochastic Point Processes. New York, Hafner.

TORRES-MELO, L. (1974). Stationary point processes. Ph.D. Thesis, University of California, Berkeley.

WALKER, A. M. (1964). Asymptotic properties of least squares estimates of parameters of the spectrum of a stationary nondeterministic time series. J. Austral. Math. Soc., 4. 363-384.

WHITTLE, P. (1953). Estimation and information in stationary time series. Ark. Math. Astron. Fys., 2. 423-434.

The Annals of Probability
1975, Vol. 3, No. 6, 909–929

SPECIAL INVITED PAPER

THE IDENTIFICATION OF POINT PROCESS SYSTEMS[1]

BY DAVID R. BRILLINGER

The University of California, Berkeley

A point process system is a random operator assigning a nonnegative integer-valued measure to a random nonnegative integer-valued measure. We define certain parameters for such a system and discuss the problem of estimating these parameters. We also consider the related problem of measuring the degree of association of two point processes.

1. Introduction and summary. A (stochastic) *point process* M is a random nonnegative integer-valued measure. If a point process M influences an apparatus $\mathscr{S}$ (perhaps real, perhaps conceptual, typically incorporating stochastic features), to give rise to another point process N, we write

$$N = \mathscr{S}[M]$$

and say that the point process N is the output of the *system* $\mathscr{S}$ operating on the input process M. We write $M(A)$ to denote the measure of the time interval A for a realization of the input process and $N(A)$ the corresponding measure for N. In practice $M(A)$ refers to the number of occurrences in A of some phenomenon of interest and $N(A)$ to the corresponding number of occurrences of some second phenomenon. We illustrate with two examples, one specific, the other more vague.

EXAMPLE 1. Let M have single points (corresponding to isolated occurrences) located at σ_j, $j = 0, \pm 1, \cdots$ and suppose that γ_j are real-valued random variables. Then

$$N(A) = \#\{j : \sigma_j + \gamma_j \text{ in } A\}$$

(i.e., $N(A)$ denotes the number of points σ_j which when moved by γ_j lie in the set A) defines a point process system. This particular system is called a *random translation* or *motion*.

EXAMPLE 2. The pulse discharges of many nerve cells have large amplitudes and are of short duration, so that they can be conveniently described as a point process. If we take two nerve cells that have a certain physiological configuration (e.g. proximity, or electrically connected), then it may be the case that the

Received January 3, 1975.

[1] Presented at the 1974 Annual Meeting of the Institute of Mathematical Statistics held at Edmonton, Alberta, August 13–16. Part of the research carried out as a Miller Research Professor, part with the support of NSF Grant GP-31411.

AMS 1970 *subject classifications.* Primary 60G10, 10K35, 62M10, 93B30.

Key words and phrases. Identification, system, point process, stationary.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_30, © Springer Science+Business Media, LLC 2012

point process M of pulses from one cell influence the point process N of pulses emitted by another cell. Beyond postulating that $N = \mathscr{S}[M]$, we may have little notion of the system operator $\mathscr{S}$ until investigation is undertaken. We discuss such a problem in Section 7.

We say that the system is *deterministic* if $\mathscr{S}$ incorporates no random feature. Its input, M, may of course still be random. We say that the system is *time invariant* when the bivariate process (M, N) is stationary for stationary M.

By the problem of the *identification* of a point process system we shall mean that of determining characteristics of the system from observations of inputs and corresponding outputs. In the case that the system, $\mathscr{S}$, is stochastic, the most that we can hope for is to determine average quantities or parameters that characterize the statistical properties of $\mathscr{S}$. Complete identification is not possible in general.

In Section 2 we define certain parameters of stochastic point processes. In Section 3 we set down a number of useful parameters for point process systems and indicate how they might be estimated. In Section 4 we discuss the related problem of measuring the degree of association of two point processes. In Section 5 we consider the identification of systems having multidimensional input or output. The problem of identification is sometimes taken to be that of determining an estimate of a finite dimensional parameter that characterizes the behavior of a process or system. In Section 6 we present one approach to this problem. The final Section, 7, presents some results concerning the identification of the point process system corresponding to a nerve cell with a single input nerve fibre.

We do not discuss the interesting problems of "on line" (or recursive) identification, of the identification of systems with feedback, nor of special procedures for realizable systems. We do not give specific references to well-known results. These may be found in Bartlett (1963), Cox and Lewis (1966), Lewis (1972).

2. Stochastic point process parameters. Before discussing specific identification procedures, we must first introduce certain parameters that describe stochastic point processes. We will restrict ourselves to parameters of stationary bivariate processes with isolated points. (In the point process literature, such processes are referred to as orderly.)

Let (M, N) be a stationary bivariate point process on the real line with differential increments at time t given by $\{dM(t), dN(t)\} = \{M(t, t + dt], N(t, t + dt]\}$. The *mean intensity*, p_M, of the process M is defined by

$$E\{dM(t)\} = p_M\, dt\,. \tag{2.1}$$

Because the points of the process have been assumed to be isolated, expression (2.1) may be interpreted as

$$\text{Prob}\,\{M \text{ point in } (t, t + dt]\}\,.$$

The mean intensity, p_N, of the N process is defined in a similar manner.

The *second-order cross product density* at lag u, $p_{NM}(u)$, is defined by

$$E\{dN(t+u)\,dM(t)\} = p_{NM}(u)\,du\,dt\,, \qquad u \neq 0\,. \tag{2.2}$$

Expression (2.2) may also be interpreted as giving

$$\text{Prob}\{N \text{ point in } (t+u, t+u+du] \text{ and } M \text{ point in } (t, t+dt]\}\,.$$

The other second-order product densities, $p_{MM}(u)$, $p_{NN}(u)$ are defined through (2.2) by equating M and N.

These parameters may be used to define the *conditional mean intensity*

$$E\{dN(t+u)\,|\,M\{t\} = 1\} = p_{NM}(u)\,du/p_M\,, \qquad u \neq 0\,, \tag{2.3}$$

which may be interpreted as

$$\text{Prob}\{N \text{ point in } (t+u, t+u+du]\,|\,M \text{ event at } t\}\,.$$

As $|u| \to \infty$, the increments $dN(t+u)$ and $dM(t)$ are tending to become independent for many processes. This phenomenon leads to the definition of the *cross-covariance density*

$$q_{NM}(u) = p_{NM}(u) - p_N p_M\,, \qquad u \neq 0 \tag{2.4}$$

which tends to 0 as $|u| \to \infty$. The autocovariance densities, $q_{MM}(u)$, $q_{NN}(u)$ are defined similarly.

Provided M points and N points do not occur simultaneously we can write

$$dC_{NM}(u)\,dt = \text{Cov}\{dN(t+u), dM(t)\} = q_{NM}(u)\,du\,dt\,.$$

However in the case of the components themselves we must write

$$dC_{MM}(u)\,dt = \text{Cov}\{dM(t+u), dM(t)\} = (\delta(u) + q_{MM}(u))\,du\,dt$$
$$dC_{NN}(u)\,dt = \text{Cov}\{dN(t+u), dN(t)\} = (\delta(u) + q_{NN}(u))\,du\,dt\,,$$

where $\delta(\cdot)$ is the Dirac delta function, to take account of the singularity at $u = 0$.

The *cross-spectrum* of the two process at frequency λ, $f_{NM}(\lambda)$, is now defined by

$$\begin{aligned} f_{NM}(\lambda) &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\, dC_{NM}(u) \\ &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\, q_{NM}(u)\,du \end{aligned} \tag{2.5}$$

for $-\infty < \lambda < \infty$, provided the integral exists. The *power spectrum* of the process M, $f_{MM}(\lambda)$, is defined by

$$\begin{aligned} f_{MM}(\lambda) &= (2\pi)^{-1} \int \exp\{-iu\lambda\}\, dC_{MM}(u) \\ &= (2\pi)^{-1} p_M + (2\pi)^{-1} \int \exp\{-iu\lambda\} q_{MM}(u)\,du \end{aligned} \tag{2.6}$$

with a similar definition for $f_{NN}(\lambda)$.

We may continue in the previous manner and define higher-order parameters such as the third-order product density

$$p_{MMM}(u, v) = E\{dM(t+u)\,dM(t+v)\,dM(t)\}/dt\,du\,dv\,, \qquad u \neq v,\ u \neq 0,\ v \neq 0\,, \tag{2.7}$$

the third order cumulant density

$$q_{MMM}(u, v) = \text{cum}\{dM(t+u), dM(t+v), dM(t)\}/dt\,du\,dv\,, \quad u \neq v,\ u \neq 0,\ v \neq 0\,, \tag{2.8}$$

and even higher-order spectra, see Brillinger (1972).

The parameters defined in (2.1), (2.2), (2.3), (2.7) have the advantage, over corresponding parameters defined in the case of ordinary time series, of possessing a further interpretation as probabilities.

Given a segment $\{M(0, t], N(0, t]\}$, $0 < t \leqq T$, of a realization of an M, N process satisfying some regularity conditions, each of the parameters defined above may be estimated consistently as $T \to \infty$, and the asymptotic distributions of the estimates are known, see Cox and Lewis (1966, 1972) and Brillinger (1972, 1975b). Estimates of third-order densities are given in Brillinger (1975a). In this section, like Bartlett (1963), we have eschewed the mathematical problems about existence of mean intensities, autocovariance density functions, etc. Lewis (1972) contains papers concerned with these issues.

3. System parameters and system identification. Suppose that we are dealing with a time invariant system with input process M and output process N. A key element of the character of such a system is provided by

$$\text{Prob}\{N \text{ point in } (t, t+dt] \mid M\} \sim E\{dN(t) \mid M\}\,.$$

In connection with it we suppose

$$\lim_{h \downarrow 0} \text{Prob}\{N \text{ point in } (t, t+h] \mid M\}/h = \mu_M(t) \tag{3.1}$$

for given input process $M \in \mathscr{N}$. Let us discuss plausible forms for $\mu_M(t)$ for a succession of input processes.

(i) Suppose we take as input to the system $M(\cdot) \equiv 0$, that is no input events. Then we might be willing to assume that $\mu_M(t)$ exists and is equal to a constant,

$$\mu_M(t) = s_0\,. \tag{3.2}$$

The system is here assumed to be emitting points at rate s_0.

(ii) Next, suppose we take as input to the system, M corresponding to a single event at time σ. Then we might alter (3.2) to

$$\mu_M(t) = s_0 + s_1(t-\sigma) = s_0 + \int s_1(t-u)\,dM(u)\,. \tag{3.3}$$

$s_1(t)$ represents the effect, on the output intensity, of inputting a single point at time 0. For example, in a service system with service time density $g(t)$, we have (3.3) with $s_0 = 0$, $s_1(t) = g(t)$.

(iii) Suppose next we take as input to the system, M corresponding to points at times σ_1 and σ_2. Were there no interaction of the two points we might be prepared to write (3.1) as

$$\mu_M(t) = s_0 + s_1(t-\sigma_1) + s_1(t-\sigma_2) = s_0 + \int s_1(t-u)\,dM(u)\,. \tag{3.4}$$

For example if the service system has 2 or more servers, then (3.4) holds with $s_0 = 0$, $s_1(t) = g(t)$.

If there were an interaction, then we might write (3.1) as

$$(3.5)\qquad \begin{aligned}\mu_M(t) &= s_0 + s_1(t - \sigma_1) + s_1(t - \sigma_2) + s_2(t - \sigma_1, t - \sigma_2) \\ &= s_0 + \int s_1(t-u)\,dM(u) + \iint_{u \neq v} s_2(t-u, t-v)\,dM(u)\,dM(v)\end{aligned}$$

where the function $s_2(\cdot)$ gives the effect of the interaction. If the service system above has but 1 server, then (3.1) has the form

$$\mu_M(t) = g(t - \sigma_1) + \int_{\sigma_2}^{t} g(v - \sigma_1) g(t - v)\,dv\,, \qquad \sigma_1 < \sigma_2 < t\,,$$

which is of the form of (3.5).

(iv) It is now evident that we may proceed in a recursive manner building up a succession of models for (3.1) of the form

$$(3.6)\qquad \begin{aligned}\mu_M(t) = s_0 + \textstyle\sum_{k=1}^{K} \int \cdots \int_{u_1,\cdots,u_k;\text{distinct}} s_k(t - u_1, \cdots, \\ t - u_k)\,dM(u_1) \cdots dM(u_k)\end{aligned}$$

where the function $s_K(t - \sigma_1, \cdots, t - \sigma_K)$ may be interpreted as the interaction effect at time t when the input process consists of K events at times $\sigma_1, \cdots, \sigma_K$. The expansion of (3.6) is a point process analog of the Volterra expansions considered in Wiener (1958) for Gaussian processes.

We shall say that the system is *linear* when $K = 1$ in (3.6), that is

$$(3.7)\qquad \lim_{h \downarrow 0} \text{Prob}\,\{N \text{ event in } (t, t+h] \mid M\}/h = s_0 + \int s_1(t-u)\,dM(u)\,.$$

By analogy with the terminology of the ordinary time series case, we might call $s_1(\cdot)$ in (3.7), the *average impulse response* of the system. We remark that (3.7) is an average property of the system, not a sample path property. We say that the system is *realizable* or *causal*, when $s_1(u) = 0$ for $u < 0$.

EXAMPLE 3. *The $G/G/\infty$ queue.* Suppose that the jth customer of a service facility arrives at time σ_j and experiences service time γ_j, $j = 0, \pm 1, \cdots$. Suppose that the γ_j are random variables with density function $g(u)$. Then, symbolically,

$$dN(t) = (\textstyle\sum_j \delta(t - \sigma_j - \gamma_j))\,dt\,,$$

and so

$$\begin{aligned}E\{dN(t) \mid M\} &= (\textstyle\sum_j \int \delta(t - \sigma_j - \gamma) g(\gamma)\,d\gamma)\,dt \\ &= (\textstyle\sum_j g(t - \sigma_j))\,dt \\ &= (\int g(t-u)\,dM(u))\,dt\,.\end{aligned}$$

This is of the form of (3.7) with $s_0 = 0$, $s_1(u) = g(u)$ and an example of the random translation of Example 1.

EXAMPLE 4. *A Hawkes' process.* Suppose the system may be described by

$$(3.8)\qquad \mu_M(t) = \mu + \int_{-\infty}^{t} a(t-u)\,dN(u) + \int_{-\infty}^{t} b(t-u)\,dM(u)$$

and that it generates a stationary process $N(\cdot)$ when a stationary $M(\cdot)$ is taken as input. Expression (3.8) leads to

$$\mu_M(t) = \nu + \int_{-\infty}^{t} c(t-u)\, dM(u)$$

where $p_N = \mu + A(0)p_N + B(0)p_M = \nu + C(0)p_M$, $C(\lambda) = (1 - A(\lambda))^{-1}B(\lambda)$; $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ being the Fourier transforms of $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ respectively; see Hawkes (1972) for further details and references.

EXAMPLE 5. For some $\alpha, \Delta > 0$

$$\mu_M(t) = \alpha M(t-\Delta, t]$$

The output intensity is here assumed to be proportional to the number of input points in the immediately previous time interval of length Δ. This model has the form of (3.7) with $s_0 = 0$,

$$\begin{aligned} s_1(u) &= \alpha\,, \qquad 0 \leqq u < \Delta \\ &= 0\,, \qquad \text{otherwise.} \end{aligned}$$

We now turn to the problem of identifying the linear system (3.7). Provided the process $N(\cdot)$ is well-defined, the relationship (3.7) leads to the equalities,

$$\begin{aligned} p_N &= s_0 + p_M \int s_1(u)\, du \\ p_{NM}(t) &= s_0 p_M + s_1(t)p_M + \int s_1(t-u)p_{MM}(u)\, du \\ q_{NM}(t) &= s_1(t)p_M + \int s_1(t-u)q_{MM}(u)\, du \\ f_{NM}(\lambda) &= S_1(\lambda)f_{MM}(\lambda) \end{aligned} \tag{3.9}$$

where $S_1(\cdot)$ is the Fourier transform of $s_1(\cdot)$. These relations suggest the estimates

$$\begin{aligned} \hat{S}_1(\lambda) &= \hat{f}_{NM}(\lambda)\hat{f}_{MM}(\lambda)^{-1} \\ \hat{s}_0 &= \hat{p}_N - \hat{p}_M \hat{S}_1(0) \\ \hat{s}_1(t) &= (2\pi)^{-1} \int \hat{S}_1(\lambda) \exp\{it\lambda\}\, d\lambda \end{aligned}$$

where $\hat{p}_M, \hat{p}_N, \hat{f}_{MM}(\lambda), \hat{f}_{NM}(\lambda)$ are estimates of $p_M, p_N, f_{MM}(\lambda), f_{NM}(\lambda)$. Details of this estimation procedure may be found in Brillinger (1974). An example of its use with neurophysiological data is given in Section 7 of this paper.

An alternate identification procedure that may be of use in certain situations is the following. Suppose that it is known that $s_1(u)$ vanishes for $|u| > \Delta$. Suppose that the input points are spaced farther than 2Δ apart. Then the individual terms of

$$\mu_M(t) = s_0 + \int s_1(t-u)\, dM(u) = s_0 + \sum_j s_1(t - \sigma_j)$$

do not interfere. This suggests that $s_0 + s_1(u)$ can be estimated, reasonably, by an expression such as

$$\#\{|\tau_k - \sigma_j - u| < \beta\}/(2\beta M(0, T])$$

for some small β, where the τ_k denote the times of observed output events from the system. This estimate is suggested by first principles. It is also suggested by the second equation of expression (3.9) as $p_{MM}(u) = 0$ for $|u| \leqq 2\Delta$ here.

Even when the model (3.7) is not satisfied, the function $s_1(\cdot)$ satisfying (3.9) is of some interest. It provides the best linear mean-squared error predictor of the process N based on M. The relations (3.9), most especially the third, suggest that the simplest way to identify the system is to take Poisson noise as input to the system, for then $q_{MM}(u) = 0$ identically, and so $s_1(t) = q_{NM}(t)/p_M$. Finally we remark that (3.8) gives an answer to the interesting question of what sort of input behavior is most likely to lead to an output point, say at 0. We see that the increments $dM(t)$ should mimic the shape of $a(-t)$.

The above discussion indicates that, provided one has sufficient data, a linear point process system may be identified fairly directly. Unfortunately things are not so nice in the nonlinear case. Consider the model (3.6) with $K = 2$. It is convenient to set it down in an alternate form. With $M'(u) = M(u) - up_M$, we write it as

$$(3.10) \qquad \mu_M(t) = r_0 + \int r_1(t-u)\,dM'(u) + \iint_{u \neq v} r_2(t-u, t-v)\,dM'(u)\,dM'(v)\,.$$

Supposing $r_2(u, v) = r_2(v, u)$, expression (3.10) leads to

$$\begin{aligned} p_N &= r_0 + \iint r_2(-u, -v) q_{MM}(u-v)\,du\,dv \\ q_{NM}(t) &= r_1(t)p_M + \int r_1(t-u) q_{MM}(u)\,du + 2\int r_2(t, t-v) q_{MM}(v)\,dv \\ &\quad + \iint r_2(t-u, t-v) q_{MMM}(u, v)\,du\,dv \\ q_{NMM}(s, t) &= r_1(s-t) q_{MM}(0) + r_1(s) q_{MM}(t) + \int r_1(s-u) q_{MMM}(u, t)\,du \\ &\quad + 2r_2(s, s-t)p_M{}^2 + 2\int r_2(s-u, s-t) q_{MM}(u)\,du \\ &\quad + 2\int r_2(s, s-t-v) q_{MM}(v)\,dv \\ &\quad + 2\iint r_2(s-u, s-t-v) q_{MM}(u) q_{MM}(v)\,dv\,du \\ &\quad + 2\int r_2(s-u, s) q_{MMM}(u, t)\,du + 2\int r_2(s-u, s-t) q_{MMM}(u, t)\,du \\ &\quad + \iint r_2(s-u, t-v) q_{MMMM}(u, v, t)\,du\,dv\,. \end{aligned}$$

It is not at all apparent how we could make direct use of these relationships without making further assumptions. We do note that if $q_{MM}(u)$, $q_{MMM}(u, v)$, $q_{MMMM}(u, v, w)$ are all identically 0, as would be the case for a process with independent increments, such as the Poisson, then the relationships give

$$\begin{aligned} r_0 &= p_N \\ r_1(u) &= q_{NM}(u)/p_M \\ r_2(u, v) &= q_{NMM}(u, u-v)/(2p_M{}^2) \end{aligned}$$

and the functions r_1 and r_2 may be identified directly. The above discussion suggests that we should probe a point process system with Poisson noise whenever possible. Unfortunately in practise this is often not possible because the noise generating device has a "dead time", that is a nonnegligible minimum interval between points. Other procedures for identifying polynomial systems involving ordinary time series are given in Brillinger (1970). It is not presently clear if these may be adapted to the point process case usefully. The Fourier-Hermite orthogonal polynomials discussed there for Gaussian processes (and introduced into the context of system identification by Wiener (1958)) could be replaced by the Poisson–Charlier polynomials (see Hida (1970)) for Poisson noise.

An alternate nonlinear model for the conditional intensity (3.1) is the multiplicative model

$$\mu_M(t) = \beta \prod_j b(t - \sigma_j)$$
$$= \exp\{\alpha + \int a(t-u)\, dM(u)\}$$

with $\alpha = \log \beta$, $a(u) = \log b(u)$. If we expand the exponential, then we see that this corresponds to the model (3.6) with $K = \infty$. In the case that M is Poisson, this model leads to the relationships

$$p_N = \exp\{\alpha + p_M \int [b(u) - 1]\, du\}$$
$$p_{NM}(u) = p_M p_N b(u)\,.$$

Another nonlinear model of some interest is provided by

$$\mu_M(t) = \alpha \qquad \text{if} \quad M(t - \Delta, t] \geqq k$$
$$= 0 \qquad \text{otherwise}$$

for some $\Delta > 0$. An output point occurs here only if there are at least k input events in the previous time interval of length Δ.

So far we have only discussed models for the first-order system parameter (3.1). A related second-order parameter is the following,

$$\mu_M(s, t) = \lim_{h\downarrow 0} \text{Prob}\,\{N \text{ points in } (s, s + h_1] \text{ and } (t, t + h_2] \mid M\}/(h_1 h_2)\,. \tag{3.11}$$

This parameter would be especially useful were input points stimulating pairs of output points. From what has been said already we might consider modelling (3.11) by

$$r_0(s - t) + \int r_1(s - u, t - u)\, dM'(u)$$

where $M'(u) = M(u) - up_M$, $r_0(-u) = r_0(u)$, $r_1(s, t) = r_1(t, s)$. This model leads to the relationships

$$p_N = \lim_{u\to\infty} r_0(u)$$
$$p_{NN}(u) = r_0(u)$$
$$p_{NNM}(s, t) - p_{NN}(s - t)p_M = p_M r_1(s, t) + \int r_1(s - u, t - u) q_{MM}(u)\, du\,.$$

Denoting the Fourier transform of r_1 by R_1 and of the left hand side of the last expression by P, we see that

$$R_1(\lambda_1, \lambda_2) = P(\lambda_1, \lambda_2)/(2\pi f_{MM}(\lambda_1 + \lambda_2))\,.$$

We end this section by mentioning that there is a growing literature concerning a martingale approach to point processes. (See Segall et al. (1975), Segall and Kailath (1975), Boel et al. (1973), Van Schuppen and Wong (1974), Dolvio (1974), for example.) It makes use of the Doob–Meyer decomposition of submartingales and results of Doléans-Dadé and Meyer (1970) among other things. It is concerned with formalizing representations of the form

$$N(0, t] = A(t) + w(t)$$

where $A(t)$, $w(t)$ are respectively a predictable increasing process and a zero mean martingale on some σ-algebras $\mathscr{B}_t \subset \sum_{\{N(s);s\leqq t\}}$. The cases where $A(t)$ is differentiable are analagous to our assumption of (3.1). The topics covered in the literature include: detection, control, forecasting, likelihood ratios and the representation of martingales in the basic process.

4. The measurement of association. A problem of some interest is the measurement of the degree of interdependence of two point processes. This involves addressing ourselves to the question of whether the input to a point process system affects the output at all and if it does to what degree?

We begin by noting that

$$\begin{aligned}\operatorname{corr}\{dN(t+u), dM(u)\} &= (p_{NM}(u)\,dt\,du - p_N p_M\,dt\,du)/(p_N\,dt\,p_M\,du)^{\frac{1}{2}} \\ &\propto p_{NM}(u) - p_N p_M\,.\end{aligned}$$

This remark suggests our considering the measure $p_{NM}(u) - p_N p_M$. This particular measure may also be interpreted as

$$\frac{\text{Prob}\,\{dN(t+u)=1 \ \text{ and } \ dM(t)=1\} - \text{Prob}\,\{dN(t+u)=1\}\,\text{Prob}\,\{dM(t)=1\}}{dt\,du}\,.$$

An equivalent measure is

$$\frac{p_{NM}(u)}{p_M} - p_N = (\text{Prob}\,\{dN(t+u) = 1 \mid dM(t) = 1\} - \text{Prob}\,\{dN(t+u) = 1\})/dt\,.$$

Both of these measures are 0 in the case of independence.

The problem can also be viewed as one of looking for association in the 2×2 table:

$dN(t+u)$	$dM(t)$		Totals
	0	1	
0	1	$p_M\,dt$	1
1	$p_N\,du$	$p_{NM}(u)\,dt\,du$	$p_N\,du$
Totals	1	$p_M\,dt$	

A variety of measures of association have been suggested for 2×2 tables, see pages 536–540 in Kendall and Stuart (1961). In the present context, these lead to

(i) the *cross-product ratio*

$$\alpha(u) = p_{NM}(u)/[p_N p_M]\,,$$

(ii) *Yule's coefficient of association*

$$Q(u) = [p_{NM}(u) - p_N p_M]/[p_{NM}(u) + p_N p_M],$$

(iii) *Yule's coefficient of colligation*

$$Y(u) = [(p_{NM}(u))^{\frac{1}{2}} - (p_N p_M)^{\frac{1}{2}}]/[(p_{NM}(u))^{\frac{1}{2}} + (p_N p_M)^{\frac{1}{2}}],$$

(iv) *Pearson's* ϕ^2

$$\phi^2(u) = [p_{NM}(u) - p_N p_M]^2/[p_N p_M].$$

The "null" values of these measures occur in the case that $p_{NM}(u) = p_N p_M$.

An alternate manner in which to proceed is to look at the degree of correlation of certain combinations of the values of the process. For example if we set

$$d_M{}^T(\lambda) = \textstyle\int_0^T \exp\{-i\lambda t\}\, dM(t), \qquad d_N{}^T(\lambda) = \textstyle\int_0^T \exp\{-i\lambda t\}\, dN(t)$$

then

$$\begin{aligned} \lim_{T\to\infty} |\text{corr}\,\{d_M{}^T(\lambda), d_N{}^T(\lambda)\}|^2 &= \lim_{T\to\infty} \frac{|\text{Cov}\,\{d_M{}^T(\lambda), d_N{}^T(\lambda)\}|^2}{\text{Var}\, d_M{}^T(\lambda)\ \text{Var}\, d_N{}^T(\lambda)} \\ &= |f_{MN}(\lambda)|^2/|f_{MM}(\lambda) f_{NN}(\lambda)|^2 \\ &= |R_{MN}(\lambda)|^2. \end{aligned} \tag{4.1}$$

This last measure is called the *coherence* of the two processes at frequency λ. Its values lie between 0 and 1, with 0 occurring in the case of independence.

5. Multidimensional systems. So far we have been considering the case in which the system has a single input and a single output. In many interesting situations, the input and output processes are multidimensional. No great difficulties appear in extending the linear system of (3.7) to the multidimensional case. Specifically, we might postulate

$$\lim_{h\downarrow 0} E\{\mathbf{N}(t, t+h]/h \,|\, \mathbf{M}\} = \mathbf{s}_0 + \textstyle\int \mathbf{s}_1(t-u)\, d\mathbf{M}(u) \tag{5.1}$$

with the process $\mathbf{M}$ being r dimensional, the process $\mathbf{N}$ being s dimensional, $\mathbf{s}_0$ being an s vector and $\mathbf{s}_1(\cdot)$ being an $s \times r$ matrix. If $\mathbf{S}_1$ denotes the Fourier transform of $\mathbf{s}_1$, if $\mathbf{f}_{MM}(\lambda)$ denotes the spectral density matrix of the process $\mathbf{M}$ and if $\mathbf{f}_{NM}(\lambda)$ denotes the cross-spectral density matrix of the two processes, then the relation (5.1) leads to the equality $\mathbf{f}_{NM}(\lambda) = \mathbf{S}_1(\lambda)\mathbf{f}_{MM}(\lambda)$, showing that the system may be identified through estimating spectral density matrices.

In the multidimensional case we may be interested in certain partial parameters. Consider a univariate process M and a bivariate process $\mathbf{N}$ with component N_1 and N_2, corresponding to M being input to two systems $\mathscr{S}_1$ and $\mathscr{S}_2$ with outputs N_1 and N_2 respectively. In practise, the outputs N_1 and N_2 may appear to be related. However this association may only be due to the fact that the two systems had the same input M and not due to any further connection. Partial spectra provide a tool for checking into this possibility. Consider the model

$$\begin{aligned} dN_1(t) &= (\mu_1 + \textstyle\int a_1(t-u)\, dM(u))\, dt + d\varepsilon_1(t) \\ dN_2(t) &= (\mu_2 + \textstyle\int a_2(t-u)\, dM(u))\, dt + d\varepsilon_2(t) \end{aligned}$$

where ε_1 and ε_2 are processes with stationary increments. This model leads to the relationships

$$\begin{aligned} f_{\varepsilon_j\varepsilon_k}(\lambda) &= f_{N_jN_k\cdot M}(\lambda) \\ &= f_{N_jN_k}(\lambda) - f_{N_jM}(\lambda)f_{MN_k}(\lambda)/f_{MM}(\lambda) \end{aligned}$$

for $j, k = 1, 2$. In the case that the processes ε_1 and ε_2 are uncorrelated the partial cross-spectrum $f_{N_1N_2\cdot M}(\lambda)$ and consequently the partial coherence

$$|R_{N_1N_2\cdot M}(\lambda)|^2 = |R_{\varepsilon_1\varepsilon_2}(\lambda)|^2 = |f_{\varepsilon_1\varepsilon_2}(\lambda)|^2/[f_{\varepsilon_1\varepsilon_1}(\lambda)f_{\varepsilon_2\varepsilon_2}(\lambda)] \tag{5.2}$$

will be identically 0 allowing an examination of the hypothesis through estimates of these functions. An example of the checking of such a hypothesis for some neurophysiological data is given in Section 7.

6. Finite parameter models. On occasion we may find ourselves in a situation where a system of interest is characterized by a finite dimensional parameter θ. Suppose that in such a situation we may derive the form of the spectral density matrix assuming stationary input and output processes and that it is given by

$$\begin{bmatrix} f_{MM}(\lambda; \theta) & f_{MN}(\lambda; \theta) \\ f_{NM}(\lambda; \theta) & f_{NN}(\lambda; \theta) \end{bmatrix}.$$

Suppose further that

$$\lim_{|\lambda|\to\infty} f_{MM}(\lambda; \theta) = \mu_M(\theta), \qquad \lim_{|\lambda|\to\infty} f_{NN}(\lambda; \theta) = \mu_N(\theta).$$

Set

$$\begin{aligned} g_{MM}(\lambda; \theta) &= f_{MM}(\lambda; \theta)/\mu_M(\theta), \qquad g_{NN}(\lambda; \theta) = f_{NN}(\lambda; \theta)/\mu_N(\theta) \\ g_{NM}(\lambda; \theta) &= f_{NM}(\lambda; \theta)/(\mu_N(\theta)\mu_M(\theta))^{\frac{1}{2}}. \end{aligned}$$

Let $\hat{p}_M = M(0, T]/T$ and $\hat{p}_N = N(0, T]/T$, then under regularity conditions (see Brillinger (1975 b)) the variate $\mathbf{h}^T(\lambda) = \{d_M{}^T(\lambda)/(\hat{p}_M)^{\frac{1}{2}}, d_N{}^T(\lambda)/(\hat{p}_N)^{\frac{1}{2}}\}$ is asymptotically bivariate complex normal with mean $\mathbf{0}$ and covariance matrix

$$T\begin{bmatrix} g_{MM}(\lambda; \theta) & g_{MN}(\lambda; \theta) \\ g_{NM}(\lambda; \theta) & g_{NN}(\lambda; \theta) \end{bmatrix} = T\mathbf{g}(\lambda; \theta)$$

for $\lambda \neq 0$. This suggests setting down the following approximate "log likelihood" function

$$-\sum_{s=1}^{S} \left\{\log \operatorname{Det} \mathbf{g}\left(\frac{2\pi s}{T}; \theta\right) + \operatorname{tr}\left(\mathbf{J}^T\left(\frac{2\pi s}{T}\right)\mathbf{g}\left(\frac{2\pi s}{T}; \theta\right)^{-1}\right)\right\} \tag{6.1}$$

where $\mathbf{J}^T(\lambda) = \mathbf{h}^T(\lambda)^\tau\overline{\mathbf{h}^T(\lambda)}$, and then estimating θ by $\hat{\theta}$, the value maximizing expression (6.1). This procedure is a point process version of a procedure suggested by Whittle (1953, 1961) for ordinary time series. Under regularity conditions (see Brillinger (1975 b))it may be shown that the estimate $\hat{\theta}$ is consistent and asymptotically normal with mean θ and covariance matrix $2\pi T^{-1}\mathbf{A}^{-1}(\mathbf{A} + \mathbf{B})\mathbf{A}^{-1}$

where **A**, **B** are matrices with entries

$$A_{jk} = \int_0^{2\pi S/T} \operatorname{tr}\left(\frac{\partial \mathbf{g}(\alpha)}{\partial \theta_j} \mathbf{g}(\alpha)^{-1} \frac{\partial \mathbf{g}(\alpha)}{\partial \theta_k} \mathbf{g}(\alpha)^{-1} \right) d\alpha$$

$$B_{jk} = \int \sum_a \sum_b \sum_c \sum_d C_{abj}(\alpha) C_{cdk}(\beta) g_{abcd}(\alpha, -\alpha, -\beta)\, d\alpha\, d\beta$$

with $C_{abj}(\alpha)$ the entry in row a, column b of the matrix

$$\mathbf{g}(\alpha)^{-1} \frac{\partial \mathbf{g}(\alpha)}{\partial \theta_j} \mathbf{g}(\alpha)^{-1} .$$

Estimates constructed in the above manner cannot be expected to be efficient as they are based only upon first and second order parameters and statistics. It would be interesting to construct a procedure involving third order parameters as well.

7. Some examples based on neurophysiological data. The field of neurophysiology is an excellent source of problems and data relating to point process systems. The paper by Bryant et al. (1973) is a good example of recent quantitative work in the field. The data discussed below were provided to this worker by those authors.

When a microelectrode is inserted into a nerve cell, a changing voltage may be recorded. Figure 1 is an example of such records for two neighboring cells, (L10, L3), of the sea slug (*Aplysia californica*). Here, and in many cases, the records are made up of pulses of large amplitude and short duration. Consequently the times of the pulses may reasonably be thought of as realizations of point processes. Figure 2 provides estimates of certain of the parameters mentioned in this paper with M referring to the times at which the cell L10 of a sea slug fired and N referring to the corresponding times at which L3 fired. In all there were 2548 M events and 1532 N events corresponding to mean rates of $\hat{p}_M = 2.21$ and $\hat{p}_N = 1.33$ events/sec. respectively. A and B are estimates of $(p_{MM}(u)/p_M)^{\frac{1}{2}}$, $(p_{NN}(u)/p_N)^{\frac{1}{2}}$, $0 \leqq u \leqq 12.5$ sec. respectively. The construction of such estimates is described in Brillinger (1975b). The square roots are taken, because the estimates then have stable variance, *ibid.* The graphs have dips near 0 because of the cells' dead times. (There is a refractory period, after a nerve cell has fired, during which it cannot fire again.) The horizontal lines of A and B are at the levels $(\hat{p}_M)^{\frac{1}{2}}$, $(\hat{p}_N)^{\frac{1}{2}}$ respectively corresponding to estimates of the level for processes with orthogonal increments. The L10 cell was here stimulated

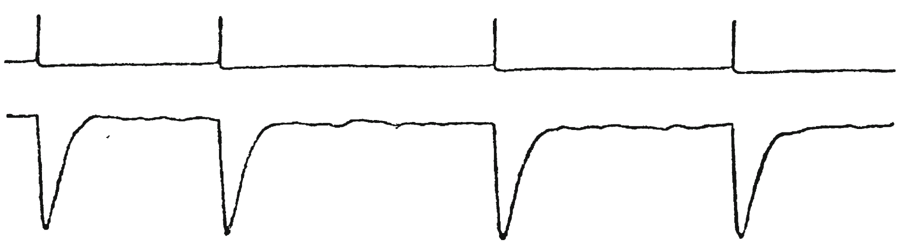


FIG. 1.

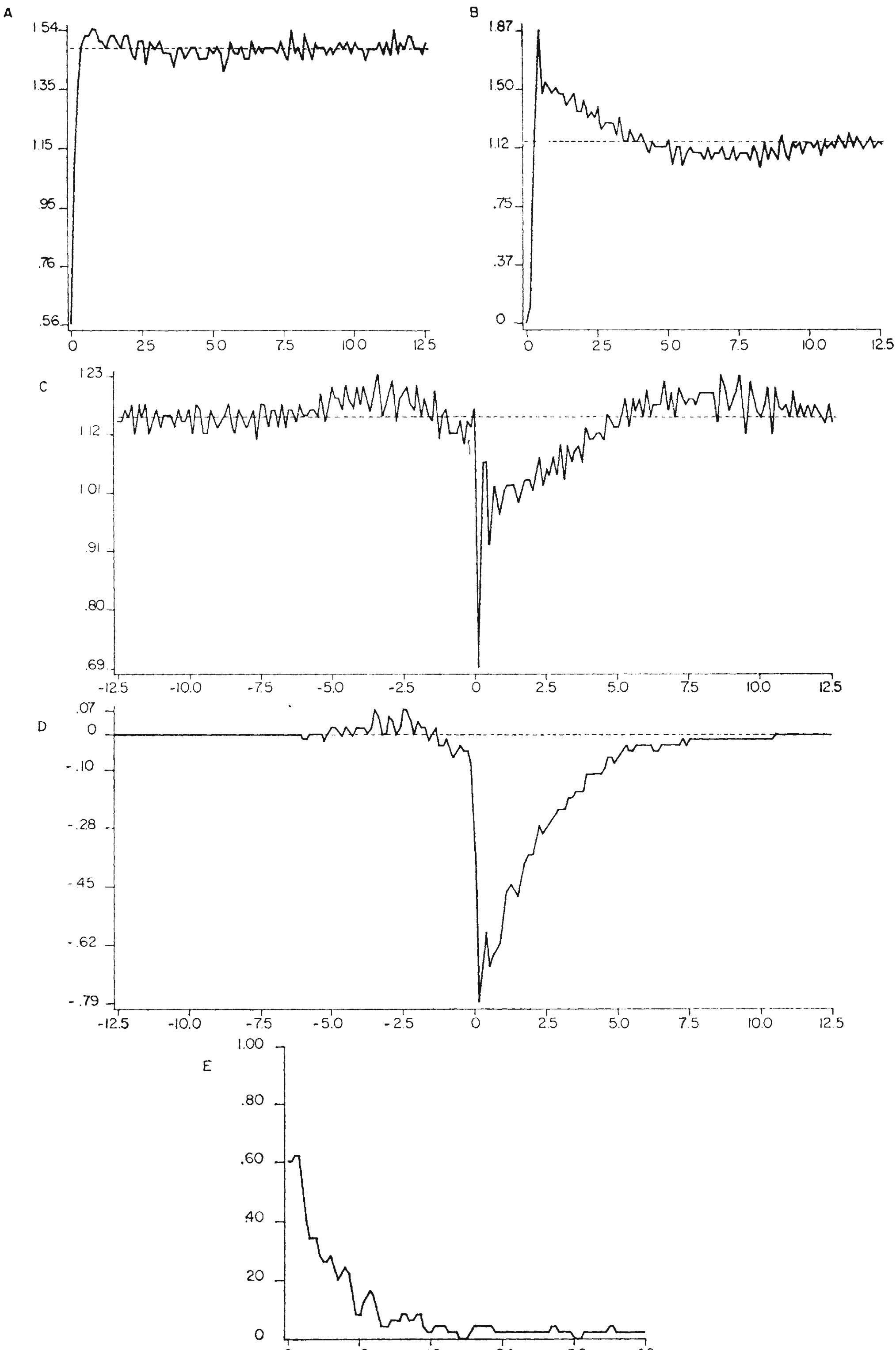

A
1.54
1.35
1.15
.95
.76
.56
0
2.5
5.0
7.5
10.0
12.5
B
1.87
1.50
1.12
.75
.37
0
0
2.5
5.0
7.5
10.0
12.5
C
1.23
1.12
1.01
.91
.80
.69
-12.5
-10.0
-7.5
-5.0
-2.5
0
2.5
5.0
7.5
10.0
12.5
D
.07
0
-.10
-.28
-.45
-.62
-.79
-12.5
-10.0
-7.5
-5.0
-2.5
0
2.5
5.0
7.5
10.0
12.5
E
1.00
.80
.60
.40
20
0
0
.8
1.6
2.4
3.2
4.0

FIG. 2.

to fire in as Poisson a manner as possible. C is an estimate of $(p_{NM}(u)/p_M)^{\frac{1}{2}}$. The horizontal line is at the level $(\hat{p}_N)^{\frac{1}{2}}$ corresponding to unassociated M and N processes. The graph suggests that there is a drop in the rate of N events for up to 5 sec. after the occurrence of an M event. D provides an estimate of the average impulse response function, $s_1(\cdot)$, of (3.7). The estimate suggests that $s_1(u)$ is near 0 for $u < 0$, in accordance with the neurophysiologists' understanding of the relationship between the cells, and it suggests that the rate of L3 pulses drops for a period after the arrival of an L10 pulse. E is an estimate of the coherence function, $|R_{MN}(\lambda)|^2$, of (4.1). The estimate is significantly different from 0.0, at the 95 per cent level, for 94 of the 100 points plotted. The apparent coherence at low frequencies is surprisingly large, considering that coherence is a measure of degree of linear association and the system is nonlinear here. (Other such coherences may be found in Figure 3.) Graphs C and D are here so similar because the input is near Poisson.

Figure 3 presents some of the results of an analysis of the sort described in Section 5 for a three cell network, (L10, L3, L2), of the sea slug. In the notation of that section, M corresponds to L10, N_1 to L3 and N_2 to L2. Graphs A, B, C are estimates of the coherences $|R_{MN_1}(\lambda)|^2$, $|R_{MN_2}(\lambda)|^2$, $|R_{N_1N_2}(\lambda)|^2$ respectively. The horizontal line corresponds to the 95 per cent point of the null distribution in each case. These graphs suggests that the three cells are intercorrelated. The neurophysiologists suspected, for these particular cells, that L10 was driving both L3 and L2 and that there was no direct path between L3 and L2. Graph

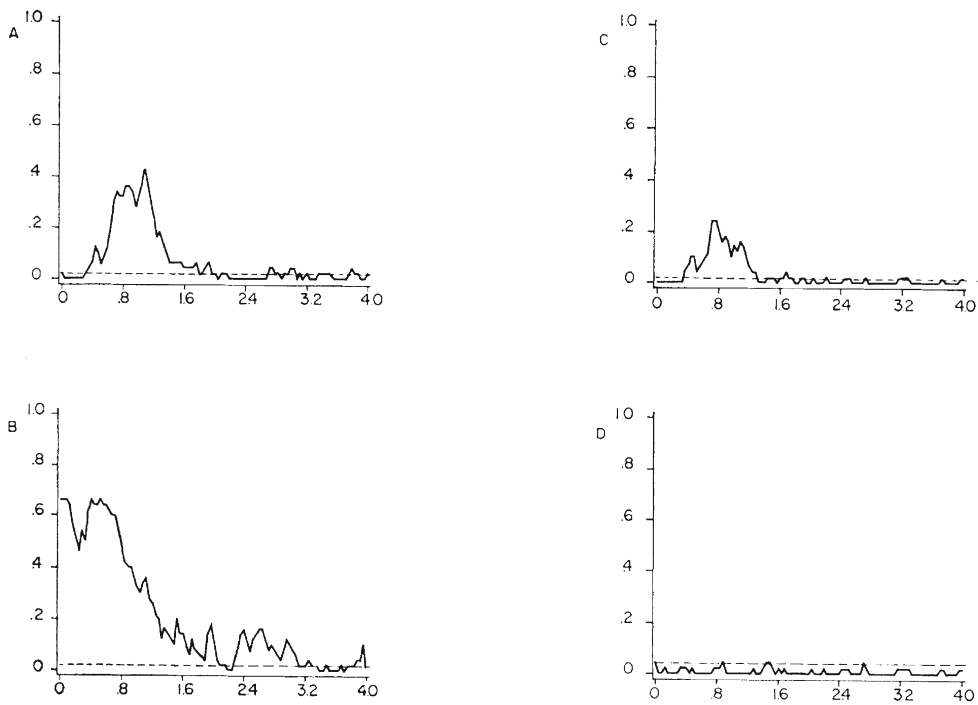


FIG. 3.

D is in accord with the suspicion. It is an estimate of the partial coherence, $|R_{N_1N_2 \cdot M}(\lambda)|^2$, of (5.2). The horizontal line corresponds to the 95 per cent point of the null distribution. There is no suggestion that the partial coherence is not 0.0.

8. Acknowledgment. I would like to thank Professor J. P. Segundo and Dr. H. L. Bryant, Jr. of the U.C.L.A. Brain Research Institute for many stimulating conversations on the topics of nerve cell network analysis and point processes and for providing the data analyzed in the previous section. Dr. Daryl Daley made helpful remarks leading to improvement in the exposition of the paper.

REFERENCES

BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264–296.

BOEL, R., VARAIYA, P. and WONG, E. (1973). Martingales on jump processes: I. Representation results, II. Applications. Memoranda ERL-M 407, 409. Electronics Research Laboratory, Berkeley.

BRILLINGER, D. R. (1970). The identification of polynomial systems by means of higher order spectra. *J. Sound Vibration* **12** 301–313.

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 483–513, Univ. of California Press.

BRILLINGER, D. R. (1974). Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *Ann. Probability* **2** 815–827.

BRILLINGER, D. R. (1975a). Estimation of product densities. *Proc. Computer Science and Statist. Eighth Annual Symp. on the Interface.* Health Sciences Computing Facility, U.C.L.A. 431–438.

BRILLINGER, D. R. (1975b). Statistical inference for stationary point processes. *Stochastic Processes and Related Topics* **1** 55–99 (M. L. Puri, ed.). Academic, New York.

BRYANT, H. L. JR., RUIZ MARCOS, A. and SEGUNDO, J. P. (1973). Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs. *J. Neurophysiology* **36** 205–225.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

COX, D. R. and LEWIS, P. A. W. (1972). Multivariate point processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 401–448, Univ. of California Press.

DOLÉANS-DADÉ, C. and MEYER, P. (1970). Integrales stochastiques par rapport aux martingales locales. *Seminaire de Probabilites* IV. Lecture Notes in Mathematics **124** Springer-Verlag, New York, 77–107.

DOLVIO, F. (1974). Counting processes and integrated conditional rates: a martingale approach with application to detection. Ph. D. thesis, Univ. Michigan.

HAWKES, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. *Stochastic Point Processes.* (P. A. W. Lewis, ed.). Wiley, New York.

HIDA, T. (1970). *Stationary Stochastic Processes.* Princeton Univ. Press.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics* **2**. Griffin, London.

LEWIS, P. A. W. (1972). *Stochastic Point Processes.* Wiley, New York.

SEGALL, A., DAVIS, M. H. A. and KAILATH, T. (1975). Nonlinear filtering with counting observations. *IEEE Trans. Information Theory* **IT-21** 143–149.

SEGALL, A. and KAILATH, T. (1975). The modelling of randomly modulated jump processes. *IEEE Trans. Information Theory* **IT-21** 135–142.

VAN SCHUPPEN, J. H. and WONG, E. (1974). Transformation of local martingales under a change of law. *Ann. Probability* **2** 879–888.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Astron. Fys.* **2** 423–434.

WHITTLE, P. (1961). Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.* **33** 105–130.

WIENER, N. (1958). *Nonlinear Problems in Random Theory.* M. I. T. Press, Cambridge.

DEPARTMENT OF STATISTICS,
UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA 94720

DISCUSSION ON PROFESSOR BRILLINGER'S PAPER

D. R. COX (*Imperial College, London*) My comments concern the statistical aspects of Dr. Brillinger's interesting paper. First, when it is required to study the dependence of a process $\{N\}$ on an explanatory process $\{M\}$, there are often strong arguments for arguing conditionally on the observed process $\{m\}$. In particular, assumptions about $\{M\}$ itself are avoided; even its stationarity is not required so long as the interrelations are time-invariant.

Secondly, some qualification seems desirable of. Dr. Brillinger's blanket recommendation that $\{M\}$ should, where possible, be chosen to be Poisson. Will not much depend on the constraints on observation and on the nature of the interrelations? For instance, one can envisage situations where it would be more informative to take $\{M\}$ as a regular sequence of widely spread points, supplemented, perhaps, by some pairs of points close together to examine linearity.

Thirdly, an alternative to the study of interrelations is via the modulation of simple models for $\{N\}$ (Cox, 1972). In this the intensity of the $\{N\}$ process is modified by a factor depending on relevant aspects of the $\{M\}$ process. Two advantages of this approach are that in certain cases likelihood functions can be obtained and that simple relations, nonlinear in Dr. Brillinger's special sense, can be accommodated; for example, the backward recurrence time in the $\{M\}$ process may be particularly relevant. An advantage of Dr. Brillinger's approach is that special assumptions about $\{N\}$ are avoided.

REFERENCES

COX, D. R. (1972). The statistical analysis of dependences in point processes. In *Stochastic Point Processes*, P. A. W. Lewis, ed. Wiley, New York.

P. Z. MARMARELIS (*California Institute of Technology*) Professor Brillinger's well-written paper on the identification of point process systems fulfills, among others, a long-standing need for such work in the field of neurophysiological system analysis. I expect that many applications of these techniques on point process systems (certainly on neural systems) will come to fruition following Brillinger's work.

MEASURING THE ASSOCIATION OF POINT PROCESSES: A CASE HISTORY

DAVID R. BRILLINGER

1. Introduction. Modern applied statistics typically involves elements of computation, probability theory, statistical theory and collaboration with specialists in the subject matter of some substantive field. In this article I shall describe part of a continuing experience of collaboration with two neurophysiologists from U.C.L.A., H. L. Bryant Jr. and J. P. Segundo. In formal terms, the problem considered is one of measuring the degree of association of points of two different sorts distributed along a straight line in an irregular manner. In real terms, the problem is one of investigating the behavior of a simple nerve cell network in a sea slug (*Aplysia californica*). The paper discusses a summary measure of association that has proved useful in assessing whether two nerve cells are behaving in a related manner or are behaving independently. The experiments by means of which the data were collected are described in Bryant, Ruiz Marcos and Segundo **[4]**, as are the results of preliminary statistical analyses. The paper **[4]** is representative of the extent to which quantification is now occuring in the life sciences.

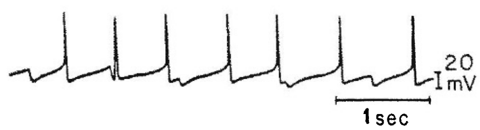


FIG. 1. A typical record of the changing voltage level of a nerve cell.

2. Some neurophysiology. The nerve cell (or neuron) is the basic unit of the animal involved in the transmission of information. Described schematically, it consists of a central cell body (or soma), branches (called dendrites) carrying impulses to the body and a long outgrowth (the axon) conducting impulses from the body. One way information is transmitted through the dendrites and axon is through changes in electrical activity. Figure 1 is an example of the changing voltage recorded when a microelectrode is inserted into a nerve cell. The record is seen to be made up of pulses of large amplitude compared to their duration. Because of its appearance, such a record is often called a spike train.

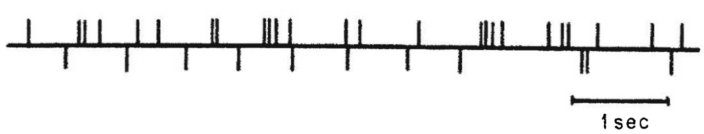


FIG. 2. A record of the times of spikes of two simultaneously firing nerve cells.

The junction whereby one neuron may influence another is called the synapse. When a pulse reaches the terminal point of an axon it provokes the release of a transmitter substance which alters the permeability of the dendrite of the next cell to certain ions. The resulting flow of ions generates a small electric current which moves down the dendrite to the soma. If the junction is excitory, the spike activity of the second cell is increased, if inhibitory it is decreased. Figure 2 is an example of the times of spikes for two nearby cells, the times for one cell corresponding to spikes above the line and for the other corresponding to spikes below. In practise, given two neurons, it may not be known whether either is influencing the other and it may be of interest to determine if there is some influence or association. Doing this by eye from records such as those of Figure 2 can be very difficult. Researchers have therefore been led to compute summary values from the records (see Griffith and Horn **[9]** for example) and this is the concern of the present paper.

The data discussed is recorded simultaneously on cells L3 and L10 of the sea hare. This

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_31, © Springer Science+Business Media, LLC 2012

particular animal and these particular cells were used because the cells may be identified in different specimens and consequently experiments may be repeated. The experimental methods are described in detail in [4]. Further information concerning neurons and synapses may be found in Eccles [7].

3. Some probability theory. Commonly in his work on applied problems, a statistician brings the apparatus of probability theory into use. This involves his asking the experimentalists and himself whether or not it is reasonable to talk about random outcomes and probabilities of events connected with outcomes. The statistician seeks to bring probability theory into a problem because it provides a precise means of defining parameters and models and it allows him to interpret and assess various manipulations of experimental data. Not all problems of data analysis require the introduction of probability theory, but many seem to benefit from its appearance—among the latter are problems concerning nerve cell spike trains.

The branch of probability theory concerned with entities like irregular spike trains is that of stochastic point processes. A stochastic point process is a random, non-negative, integer-valued measure. If I is an interval of the real line and ω is a random element, then the values of this measure may be denoted by $N(I, \omega)$, with $N(I, \omega)$ denoting the number of points in the interval I for the realization corresponding to ω. Here the atoms of the measure $N(I, \omega)$ correspond to the times of spikes of a particular spike train. Repeating the experiment would most likely yield a different set of spike times and consequently a different measure $N(I, \omega')$. In this sense N is a random measure. (We remark that in many problems one can suppress the dependence of N on ω, however, it is an essential element of the approach.) Point processes were considered recently in the MONTHLY by Chung [5] and are discussed in Cox and Lewis [6] and in a volume [11] edited by Lewis, for example.

For nerve cell trains, it is appropriate to assume that the point process is without multiple points; that is, the spike times are isolated, separated by positive distances. Because the spikes proceed from no inherent origin, it also seems appropriate to assume that the point process is stationary in time in the sense that the probability distribution of the random vector

$$\{N(I_1, \omega), \cdots, N(I_k, \omega)\}$$

is the same as that of the shifted vector

$$\{N(I_1 + t, \omega), \cdots, N(I_k + t, \omega)\}$$

for all t and $k = 1, 2, \cdots$, where $I + t$ denotes the interval $(a + t, b + t)$ if $I = (a, b)$.

Important parameters of a stationary point process N include the **mean intensity**, p_N, and the **second-order product density**, $p_{NN}(u)$, given by

$$p_N = \lim_{h \downarrow 0} \text{Prob}\{\text{point in the interval } (t, t+h)\}/h \tag{1}$$

and

$$p_{NN}(u) = \lim_{h,h' \downarrow 0} \text{Prob}\{\text{point in } (t + u - h, t + u + h) \text{ and point in } (t - h', t + h')\}/(4hh') \tag{2}$$

$-\infty < u < \infty$, respectively when these limits exist.

In fact we shall be concerned with two different types of points, say M points and N points, with $M(I, \omega)$ referring to the number of M points in the interval I and $N(I, \omega)$ the number of N points in the interval I. We denote the mean intensity of M points by p_M and the second-order product density of M points by $p_{MM}(u)$. We also define a **cross-product density**, $p_{MN}(u)$, by

$$p_{MN}(u) = \lim_{h,h' \downarrow 0} \text{Prob}\{M \text{ point in } (t + u - h, t + u + h) \text{ and } N \text{ point in } (t - h', t + h')\}/(4hh'). \tag{3}$$

The parameters in (1), (2), (3) do not depend on t because the process is stationary.

The first thing that one tends to notice when examining a spike train is whether there are a lot of spikes or only a few. The mean intensity of the process gives information in this connection. Expression (1) implies that the probability of there being an N point in a small interval of length h is approximately $p_N h$. The next thing that one tends to notice is the relative positioning of pairs of spikes of a single train or from one train to another. Expressions (2) and (3) give information in this connection. From expression (3), for example, we have

$$\text{Prob}\{M \text{ point in } (t+u-h, t+u+h) \text{ and } N \text{ point in } (t-h', t+h')\} \sim p_{MN}(u)4hh' \tag{4}$$

for h, h' non-negative and small. Using the definition of conditional probability and (1), this implies that

$$\text{Prob}\{M \text{ point in } (t+u-h, t+u+h) \text{ given an } N \text{ point at } t\} \sim 2h\, p_{MN}(u)/p_N. \tag{5}$$

In the case that the M points are distributed independently of the N points, the probability referred to in expression (5) is just $\text{Prob}\{M \text{ point in } (t+u-h,\ t+u+h)\}$ and so

$$p_{MN}(u)/p_N = p_M \quad \text{or} \quad p_{MN}(u) = p_M p_N \tag{6}$$

for all u. This last suggests that the function $p_{MN}(u)$, and related functions such as

$$\frac{p_{MN}(u)}{p_M p_N} \quad \text{or} \quad \sqrt{\frac{p_{MN}(u)}{p_M p_N}} \tag{7}$$

might prove useful measures of the degree of association of points of the M process with points of the N process. They are identically 1.00 in the case of independence.

We remark that, since we have assumed the points of the processes to be isolated, we can replace the probabilities of expressions (1)–(3) by expected values, for example we could write for (3)

$$p_{MN}(u) = \lim_{h,h' \downarrow 0} E\{N(t+u-h, t+u+h)\, N(t-h', t+h')\}/(4hh').$$

4. Some statistical theory. The preceding section described a mathematical idealization that could be of use in examining the degree of relationship of two given spike trains. The idealization suggested the definition of parameters p_M, p_N, $p_{MN}(u)$ based on the probabilities of certain events. In order to make concrete use of these parameters we need to have some idea of their values for the spike trains at hand.

Statistical theory has long been concerned with the problem of estimating the probability of an event given experimental results. In elementary situations one estimates the probability of an event A by n_A/n, where n_A denotes the number of times the event A occurred out of n times when it might have occurred. Let us use this approach to construct estimates of $p_M, p_N, p_{MN}(u)$.

Suppose that spike trains M and N are observed throughout the time interval $(0, T)$. Let $s_1 < s_2 < \cdots < s_{M(T)}$ be the observed times of M spikes and $t_1 < t_2 < \cdots < t_{N(T)}$ be the observed times of N spikes where we have observed $M(T)$ M spikes and $N(T)$ N spikes in all. Let h be small and imagine the interval $(0, T)$ divided into T/h intervals of length h. The number of times the event "M spike in small interval of length h" occurred is $M(T)$. It might have occurred T/h times. This suggests estimating $p_M h$ by $M(T)/(T/h)$ and so estimating p_M by

$$\hat{p}_M = M(T)/T. \tag{8}$$

Likewise we could estimate p_N by $\hat{p}_N = N(T)/T$.

Next we consider the estimation of $p_{MN}(u)$. For small h, let

(9) $n(u,h)$ = the number of s_i such that $t_j + u - h < s_i < t_j + u + h$ for some j,

then the probability of expression (5) may be estimated by $n(u,h)/N(T)$. This suggests the estimation of $p_{MN}(u)$ by

$$\hat{p}_{MN}(u) = \frac{n(u,h)\hat{p}_N}{N(T)2h} = \frac{n(u,h)}{2hT}. \tag{10}$$

This estimate may be used in turn to construct estimates of the functions of expression (7) if so desired. We must not forget, however, that we have not measured $p_{MN}(u)$ exactly. Rather we have constructed an expression that should be near it, especially in the case that T is large. We must also remember that were we to repeat the experiment, almost certainly, we would obtain a different value for $\hat{p}_{MN}(u)$. This last variation is called sampling fluctuation. Figure 3 is a graph of $\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)$ for the nerve cell data described in Section 2. Here $N(T) = 1232$, $M(T) = 816$, $\hat{p}_N = 1.77$ events/sec. and $\hat{p}_M = 1.17$ events/sec.. The Figure suggests that the probability of occurrence of an M event is depressed for about 5 seconds after the occurrence of an N event. However, before we can come to a belief that the outputs of the two nerve cells are in fact related, we must first have some confidence that the deviation of the function from the value 1.00 is not due simply to sampling fluctuations.

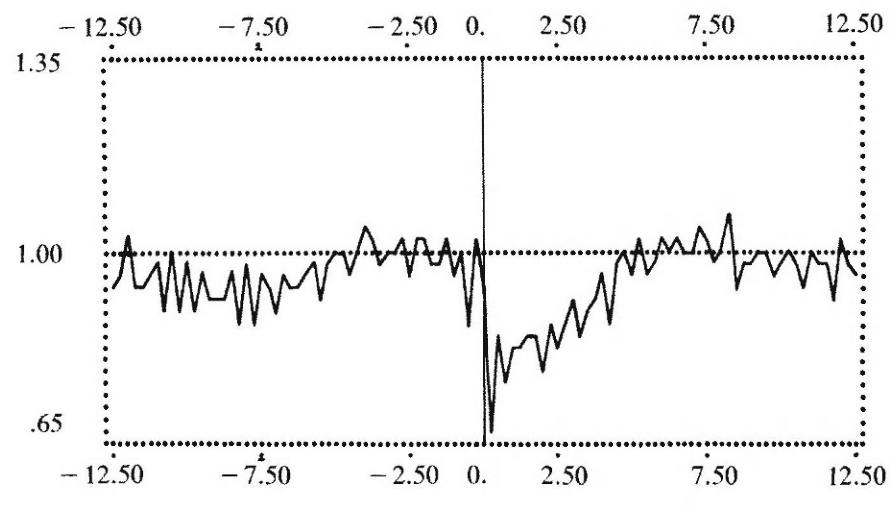


FIG. 3. An example of $\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)$ for cells L3 and L10 of *Aplysia californica*. The horizontal axis gives u in seconds.

The estimate (10) was proposed in Griffith and Horn [9]. Its direct computation involves the comparison of $M(T) \cdot N(T)$ values. In many of the experiments referred to $M(T)$ and $N(T)$ are both about 1000, so fairly clearly a high speed computer must be used in its computation. Hugh Bryant, Jr. has noted that if the spike times are recorded by increasing time, with a simple indicator to say whether a spike was an M or an N, then there exists a direct algorithm for computing (9) with one pass through the data. Suppose the data is denoted (u_j, α_j), $j = 1, 2, \cdots$ where $u_1 < u_2 < u_3 \cdots$ and u_j is an s_k if $\alpha_j = 0$, u_j is a t_k if $\alpha_j = 1$. The algorithm is the following: (1) initialize $n(lh, h)$ to 0 for $l = 0, \pm 1, \pm 3, \cdots$, (2) for $j = 1, 2, \cdots$ and $k = j+1, j+2, \cdots$ if $\alpha_j = 1$, $\alpha_k = 0$ compute $l = [(u_k - u_j)/(2h)]$ and set $n(lh, h) = n(lh, h) + 1$ or if $\alpha_j = 0$, $\alpha_k = 1$ compute $l = -[(u_k - u_j)/(2h) + 1/2]$ and set $n(lh, h) = n(lh, h) + 1$. (Here $[x]$ means the integral part of the number x.)

5. Some more probability theory. Figure 3 is a graph of an estimate of $p_{MN}(u)/(p_M p_N)$ rather than the function itself. Part of the irregular nature of the figure is undoubtedly due to sampling fluctuations. Before we can come to a reasonable decision that the two spike trains are related, with a spike of the N train associated with an apparent depression in the rate of M spikes for example, we must assess the magnitude of sampling fluctuations. The key random variate appearing in $\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)$ is $n(u,h)$ given by expression (9). Let us attempt to approximate the distribution of this variate for large T.

$n(u,h)$ is a counting variate. As h is small it is counting rare events. Now in many situations, counts of rare events are approximately Poisson (see for example Feller [8] p. 282). Volkonski and Rozanov [12] demonstrate the related result that if $N^T(I,\omega)$, $T=1,2,\cdots$ is a sequence of point processes with mean intensities $p_N^T \to 0$ as $T\to\infty$, then under a further regularity condition, the sequence of processes with rescaled time, $N^T(I/p_N^T,\omega)$, $T=1,2,\cdots$ tends to a Poisson process. Perhaps $n(u,h)$ here is approximately Poisson with mean $2hTp_{MN}(u)$.

Supposing $h=L/T$, with L constant, we define a sequence of processes $N^T(I,\omega)$ by saying that $N^T(\cdot,\omega)$ has a spike at t, if $N(\cdot,\omega)$ has a spike at t and if $M(\cdot,\omega)$ has a spike in the interval $(t+u-L/T, t+u+L/T)$. The mean intensity of the process $N^T(\cdot,\omega)$ is $\sim p_{MN}(u)\,2L/T\to 0$. In Brillinger [2] it is shown that the result of Volkonski and Rozanov [12] may be applied to conclude that for large T the process $N^T(IT,\omega)$ is approximately Poisson and in particular $n(u,T)\sim N^T((0,T),\omega)$ is approximately Poisson with mean $2hTp_{MN}(u)$.

6. Some statistical inference. A common procedure that a statistician employs to communicate an interval of plausible values for an unknown parameter, in the light of data collected, is a confidence interval. For example, a 95 per cent confidence interval has the formal interpretation that 95 is the long run percentage of such intervals that actually contained the true parameter value. At this point we could use Table 40 of Biometrika Tables, based on the Poisson distribution, to construct a confidence interval for $p_{MN}(u)$ and by division through by $\hat{p}_M\hat{p}_N$ a confidence interval for $p_{MN}(u)/(p_Mp_N)$.

A less troublesome way in which to proceed is to take advantage of the fact that if P is a Poisson variate with mean μ, then $\sqrt{P}$ is approximately a normal variate with mean $\sqrt{\mu}$ and standard deviation 1/2 (see pp. 88–96 in Kendall and Stuart [10]). The application of this square root transformation has two advantages; tables of the normal distribution are widely available for constructing confidence intervals, and the approximate standard deviation does not depend on the unknown parameter. All of this suggests a consideration of the estimate $\sqrt{\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)}$ and of the approximation of its distribution by a normal with mean $\sqrt{p_{MN}(u)/(p_Mp_N)}$ and standard deviation $1/(2\sqrt{2hTp_Mp_N})$. Approximate 95 per cent confidence intervals are constructed by adding and subtracting $1.96/(2\sqrt{2h\hat{p}_M\hat{p}_N})$ along the curve.

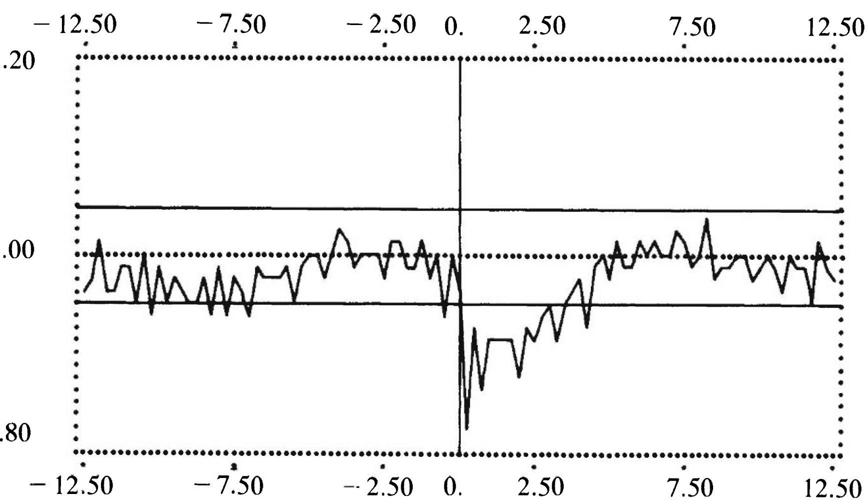


FIG. 4. $\sqrt{\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)}$ and 95 per cent confidence limits about the level 1.00.

In the case of independence of the two processes $p_{MN}(u)/(p_M p_N) = 1$ and this hypothesis may be easily checked into by plotting horizontal lines at the levels $1.00 \pm 1.96/(2\sqrt{2hT\hat{p}_M\hat{p}_N})$. This has been done in Figure 4, for the data of Figure 3. This new Figure is strongly suggestive of the association of a reduced rate of M spikes for a period after the occurrence of an N spike.

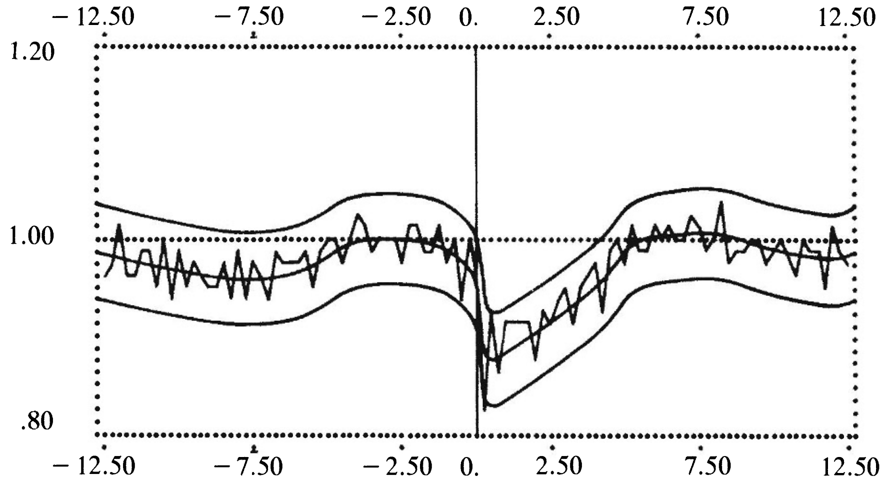


FIG. 5. $\sqrt{\hat{p}_{MN}(u)/(\hat{p}_M\hat{p}_N)}$ and 95 per cent confidence limits about a smoothed version of these values.

We might have chosen to indicate sampling fluctuations in the manner of Figure 5 where we have plotted $\pm 1.96/(2\sqrt{2hT\hat{p}_M\hat{p}_N})$ limits around a heavily smoothed version of the estimate. The points where the upper line is below 1.00 might be assessed significant. Figure 5 is also suggestive of a direction of causation for the two cells, namely the N spikes seem to be associated only with later M spikes (i.e., at positive u). This last is consistent with the neurophysiologists' understanding of the relation of the two particular cells for which this data was collected.

7. Final remarks. In this article I have sought to describe some of the stages involved in a modern applied statistics problem. These include: (i) the experimenter collects interesting data, (ii) the experimenter recognizes relevent scientific parameters to estimate, (iii) the experimenter consults a statistician as to whether or not his estimates are significant, (iv) the statistician suggests means of assessing sampling fluctuations and possibly suggests transformations in order that the data be more simply described and (v) the experimenter and statistician collaborate to determine and fit a statistical model and to design future experiments to confirm that model. The distinction between these stages is not always apparent nor is it clear whose ideas are whose. The experimenter must learn a fair amount of statistical methodology and the statistician must learn a fair amount of the experimenter's subject matter before real progress can be made.

Previous to my involvement in this work and the carrying out of the research leading to **[3]**, another applied statistician, Peter Lewis, was involved. He suggested, **[4]**, some clever and widely applicable alternate procedures for assessing the significance of the estimate (10). These were: (i) look at the variations of $n(u, h)$ for $|u|$ large enough that the interactions of the two cells are independent, (ii) substitute for the N train, a spike train of $N(T)$ independent spikes and examine the variation of $n(u, h)$ here, (iii) split the observed trains into J (say 20) pairs of shorter records and examine the variation of $n(u, h)$ when computed for the shorter records. The easiest of these procedures to carry out is (i). When done, it does lead to an estimate of the variation of the same order of magnitude as that of this paper.

Reference **[3]** is a joint paper carrying the analysis of the data described in this paper considerably further forward. I should like to thank Hugh Bryant, Jr. and Jose Segundo for the great pleasure I have derived from working with them on this problem.

Research supported in part by NSF Grant GP–31411.

References

1. Biometrika Tables for Statisticians, Vol. I, Third Edition (eds. E. S. Pearson and H. O. Hartley) Cambridge Univ. Press, Cambridge, 1966.

2. D. R. Brillinger, Estimation of the second-order intensities of a bivariate stationary point process, J. Royal Statistical Soc. B December (1975).

3. D. R. Brillinger, H. L. Bryant, Jr. and J. P. Segundo, The identification of neuronal interactions, To appear in J. Neurophysiol. (1975).

4. H. L. Bryant, Jr., A. Ruiz Marcos and J. P. Segundo, Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs, J. Neurophysiol., 36 (1973) 205–225.

5. K. L. Chung, Crudely stationary counting processes, this MONTHLY, 79 (1972) 867–877.

6. D. R. Cox and P. A. W. Lewis, The Statistical Analysis of Series of Events, Methuen, London, 1966.

7. J. Eccles, The synapse, Scientific American, 212 (1965) 56–66.

8. W. Feller, An Introduction to Probability Theory and its Applications, Vol. I, 3rd ed., Wiley, New York, 1968.

9. J. S. Griffith and G. Horn, Functional coupling between cells in the visual cortex of the unrestrained cat. Nature, 199 (1963) 876, 893–895.

10. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 3, Griffin, London, 1966.

11. Stochastic Point Processes (ed. P. A. W. Lewis), Wiley, New York, 1972.

12. V. A. Volkonskii and Yu. A. Rozanov, Some limit theorems for random functions, I. Theory Probability and its Applications, 4 (1959) 178–197.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, BERKELEY, CA 94720.

ADDENDUM TO "STRONG DERIVATIVES AND INVERSE MAPPINGS"

(This MONTHLY, 81(1974) 969–981)

ALBERT NIJENHUIS

8. Further comments. An excellent discussion of strong derivatives and other properties of functions appears in Gleason's lectures [**8**], which also make reference to [7].

A treatment of the subject of the first four sections of the present paper, ostensibly finite dimensional but general in spirit, appears in [**9**]. Their use of the "homeomorphism theorems" 5.1.6–7, is very clarifying.

A correction: formula (4.7) should be

$$\varphi(y,z) = x_0 + \varphi^*(\lambda^{-1}(y - y_0), z). \tag{4.7}$$

Additional References

7. E. Asplund and L. Bungart, A first course in integration, Holt, Rinehart and Winston, New York, 1966.

8. L. K. Durst (ed.), Proceedings of the CUPM Geometry Conference, Part II: Geometry in other subjects, Math. Assoc. of Amer., 1967.

9. J. M. Ortega and W. C. Rheinboldt, Iterative solution of nonlinear equations in several variables, Academic Press, New York, 1970.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PA 19174.

Biol. Cybernetics 35, 213–220 (1979)

Biological Cybernetics

Empirical Examination of the Threshold Model of Neuron Firing*

David R. Brillinger and José P. Segundo

Department of Statistics, University of California, Berkeley, California, USA, and
Department of Anatomy, University of California, Los Angeles, California, USA

Abstract. An elementary model of neuronal activity involves temporal and spatial summation of postsynaptic currents that are elicited by presynaptic spikes and that, in turn, elicit postsynaptic potentials at a trigger zone; when the potential at the trigger zone exceeds a "threshold" level, a postsynaptic spike is generated. This paper describes three methods of estimating the "summation function", that is, the function of time that converts the synaptic current into potential at the trigger zone: namely, maximum likelihood, cross-correlation analysis and cross-spectral analysis. All three methods, when applied to input-output data collected on various neurons of *Aplysia californica*, give comparable results. As estimated, the summation function involved in the explored cells has an early positive-going swing that is large and brief. In the cell L5, but not in R2, there was also a late negative-going swing of longer duration.

1. Introduction

A classical analytical model for a neuron firing impulses when subjected to presynaptic influences involves i) linear summation of synaptic currents and exponential decay to provide membrane potential and ii) firing when the latter exceeds a threshold with resetting of membrane potential to initial value. The model is referred to as "leaky integrator" and is discussed in Holden (1976) for example.

In the experiments to be discussed, current is inserted directly into the soma of a living cell. Let the process taking the somatic current, (at the insertion site), over to potential at the trigger zone and the evolution of the latter be assumed linear and described by the summation function $a(t)$. The function $a(t)$ describes the course that the potential would follow after a current impulse. The linearity assumption requires that the effects of current pulses at different times be additive, and will be made precise shortly. Let $B(t)$ denote the time elapsed at time t since the neuron last fired. (The backward recurrence time at t.) Let $X(t)$ denote the level of the input current at time t. Then $U(t)$, the membrane potential at the trigger zone at time t, may be represented as

$$U(t)=\int_0^{B(t)} a(u)X(t-u)du\,. \tag{1.1}$$

Let $\theta(t)$ denote the threshold potential at time t. A naive model for the firing times of the neuron is to take them to be the times at which $U(t)$ crosses $\theta(t)$ in a particular direction with $\theta(t)$ assumed constant, for example.

The analyses to be discussed were carried out on a digital computer and so it is necessary to replace expression (1.1) with the sampled version

$$U_t=\sum_{u=0}^{B_t-1} a_u X_{t-u}\,, \tag{1.2}$$

$t=0, \pm 1, \pm 2, \ldots$, with $B_t=t-\pi_t$ where π_t is the time of the spike immediately preceding t. In discrete time the spike train, or the synaptic currents it evokes, can be identified with a series of 0's and 1's, 1 if a spike occurred at the corresponding time, and may be represented by

$$Y_t=H(U_t-\theta_t)\,, \tag{1.3}$$

with θ_t the threshold at time t and $H(u)$ the Heavyside function that equals 1 for $u \geqq 0$ and 0 otherwise.

It is of interest to estimate the summation function a_t itself, the overall threshold level and other characteristics such as Prob$\{Y_t=1|U_t\}$, the probability of firing as a function of potential, from a stretch of data (X_t, Y_t),

* Prepared with the partial support of the National Science Foundation Grant MCS 77-22986 to DRB and NSF and NIH grants to JPS

0340-1200/79/0035/0213/$01.60

$t=0,\ldots,T-1$. This specific estimation problem, involving summation back to the previous firing, does not appear to have been considered previously. Knox (1974) and Rubio and Holden (1975) have investigated models involving summation of all previous input.

2. Maximum Likelihood Fit

Suppose that the threshold series θ fluctuates about a constant level θ, specifically that the successive values of θ_t are independent normal variates with mean θ and variance 1. (Variance 1 is no restriction as the a_u may be scaled arbitrarily.) Then, for example,

$$\begin{aligned}\text{Prob}\,\{Y_t=1|U_t\} &= \text{Prob}\,\{U_t-\theta_t \geqq 0\}\\ &= \Phi(U_t-\theta)\,,\end{aligned} \tag{2.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a normal variate with mean 0 and variance 1 from the assumption re θ_t. Conditional on the X_t values at hand, the probability of observing the stretch Y_t, $t=0,1,\ldots,T-1$ is

$$\prod_{t=0}^{T-1} \Phi(U_t-\theta)^{Y_t}[1-\Phi(U_t-\theta)]^{1-Y_t}\,. \tag{2.2}$$

As a function of the unknown parameters of the model (here the a_u and θ) and evaluated at the observed values, expression (2.2) is called the likelihood function of the data. A traditional means of estimating the unknown parameters of a statistical model is the take as estimates the values of the parameters that maximize the likelihood [see for example Rao (1965)]. Examples of such estimates will be provided later in the paper.

Once estimates $\hat{a}_u$ are at hand, the potential at time t may be estimated by

$$\hat{U}_t=\sum_{u=0}^{B_t-1} \hat{a}_u X_{t-u} \tag{2.3}$$

and the conditional probability of (2.1) estimated in turn. Examples of this will also be presented later.

3. A Cross-spectral Approach

The maximum likelihood fitting procedure although effective, is a fairly time consuming procedure for the model of this paper and the lengthy data sets at hand. In consequence it is worth investigating alternate fitting procedures.

Suppose that the threshold level, θ, is high enough that the B_t are generally large and that the a_t die off reasonably rapidly as t increases. Then the potential U_t, of (1.2), may be approximated by

$$V_t=\sum_{u=0}^{\infty} a_u X_{t-u} \tag{3.1}$$

and the spike train may be approximated by $H(V_t-\theta_t)$. Suppose further, and this is a key assumption, that the input series X_t is stationary Gaussian. Finally suppose the threshold series θ_t is statistically independent of the series X_t.

Because of the Gaussianity of the series X_t one may write

$$X_{t-v}=m_v+b_vV_t+\varepsilon_t \tag{3.2}$$

with m_v, b_v constants

$$\begin{aligned}b_v &= \text{cov}\,\{V_t,X_{t-v}\}/\text{var}\,V_t\\ &= \sum_{u=0}^{\infty} a_u c_{XX}(v-u)/\sigma_{VV}\end{aligned} \tag{3.3}$$

and, what is crucial, with ε_t statistically independent of V_t (see Brillinger, 1977). Here $c_{XX}(u)=\text{cov}\,\{X_{t+u},X_t\}$. It follows that

$$\begin{aligned}c_{YX}(v) &= \text{cov}\,\{Y_t,X_{t-v}\}\\ &= \text{cov}\,\{H(V_t-\theta_t),X_{t-v}\}\\ &= b_v\,\text{cov}\,\{H(V_t-\theta_t),V_t\}\\ &= L\sum_{u=0}^{\infty} a_u c_{XX}(v-u)\end{aligned} \tag{3.4}$$

with L a constant [which may be evaluated as

$\varphi(\theta/\sqrt{\sigma_{\varepsilon\varepsilon}+\sigma_{VV}})$,

$\varphi(\cdot)$ the normal density]. What is of interest here is that the final result of (3.4) is, up to a scale factor, the identification relationship of a linear time invariant system. [This sort of result has been pointed out in de Boer and Kuyper (1968), Korenberg (1973), Brillinger (1977) under varying levels of assumptions.]

Given the stretch of data (X_t, Y_t), $t=0,1,\ldots,T-1$ the summation function a_u, may be estimated (up to the unknown scale factor L) via cross-spectral analysis [see for example Chap. 8 of Brillinger (1975)]. Given the estimates $\hat{a}_u$, the fitted potential may be computed via expression (2.3) and the firing probability of (2.1) estimated in turn.

In fact the relationship (3.4) holds for an exceedingly broad class of instantaneous operators on the seies V_t. Whatever threshold character the operator may have remains to be inferred from further calculations, such as estimating the probability of (2.1) as presented later.

4. A Cross-correlation Approach

Suppose that the signal, X_t, driving the neuron has been taken to be Gaussian white noise. Then the relationship (3.4) will continue to hold, but further the autocovariance function of the input will be given by $c_{XX}(0)=\sigma_{XX}$, $c_{XX}(u)=0$ for $u \neq 0$. The relationship (3.4) hence reduces to

$$c_{YX}(v)=L\sigma_{XX}a_v . \tag{4.1}$$

There is no need then to carry out a cross-spectral analysis to estimate the summation function a_u (up to a scale factor). An estimate of the cross-covariance function of the output series with the input provides a direct estimate.

It may be remarked that the theoretical justification of the estimates of this and of the previous section are based firmly on how well the experimenter can generate input signals with specified stochastic properties and the degree to which one can replace B_t of (1.2) by ∞. On the other hand the maximum likelihood approach allows arbitrary input characteristics and deals with the scientifically relevant model. Empirical calculations to be presented later suggest that in some circumstances the Gaussian assumption is not crucial to the cross-spectral and cross-correlation approaches nor do great difficulties result from the replacement of B_t by ∞.

5. Experimental Methods

Experiments were performed on the identified neurons referred to as R2 and L5 of the abdominal ganglion of *Aplysia californica*, isolated with nerves and connectives, immersed in artificial sea water (ASW, pH 7.6), and maintained at 17°C using a servo-controlled Peltier device. To facilitate multiple-electrode penetration of certain cells the ganglia were typically placed in a 1% solution of Pronnase for 10–15 min in order to soften the connective tissue capsule.

Cells were impaled with separate stimulating and recording electrodes (KCl, 5–10 MΩ). The stimulating electrode was used to inject a continuously varying current. Intracellular recordings were obtained as previously described (Bryant et al., 1973). The stimulating current and corresponding transmembrane potential were stored on analog tape for later computer processing.

In the experiments to be discussed in this paper the input current injected was modulated as either Gaussian or uniform white noise. The specific input employed was produced by either a Hewlett-Packard noise generator (Model 8057A) if Gaussian or the computer if uniform. Pure white noise with an everywhere constant power spectrum is not physically realizable, and bandlimited is the most that can be achieved. The bandpass of the input employed was determined by the frequency on an external clock (Bryant and Segundo, 1976).

6. Statistical Methods

The analog recordings were digitized (at either 32 or 50 Hz) and the output series, Y_t, taken to be 1 if an action potential occurred in the corresponding time interval and to be 0 otherwise. Each of the three methods of fit, maximum likelihood, cross-spectral analysis, cross-correlation analysis were applied in the cases discussed below.

In the maximum likelihood approach the likelihood function (2.2) is to be maximized as a function of θ and the a_u introduced via (1.2). On the basis of the cross-spectral and cross-correlation analyses to be presented later it was decided to assume $a_u=0$ for $u>25$ [making (2.2) a function of 26 unknown parameters]. To reduce the amount of computing required, maximum likelihood estimates were determined separately for contiguous stretches of 1900 X_t values: this has the further advantage of allowing the direct assessment of the variability of the estimates computed. The derivatives of the function (2.2) were evaluated analytically as required by the particular maximization procedure employed (i.e. the FORTRAN subroutine VA09A of the Harwell library). This procedure further requires choosing initial values for the parameters: several sets of initial values were tried for the first stretch of data and the program converged to the same extreme values in every case. For each subsequent stretch of data, the initial values were taken to be the average of the estimates of the already processed stretches. [Computational and statistical properties of maximum likelihood estimates are presented in Chambers (1977).] Overall estimates of the a_u were constructed by averaging together the maximum likelihood values for the various stretches and then hanning the values corresponding to the a_u's. "Hanning" refers to a running smoothing employing the weights 1/4, 1/2, 1/4 (Tukey, 1977) that suppresses the ripling introduced by the sampling process.

The cross-spectral analysis of time series data (X_t, Y_t), $t=0,1,\ldots, T-1$ is now classical. An effective way to proceed is (i) to Fourier transform the corresponding stretches of data (employing a fast Fourier transform algorithm), (ii) to form the auto- and cross-periodograms from the Fourier transforms, (iii) to smooth these periodograms to obtain estimates of the auto- and cross-spectra $f_{XX}^T(\lambda)$, $f_{YY}^T(\lambda)$, $f_{YX}^T(\lambda)$, (iv) to back Fourier transform $f_{YX}^T(\lambda)/f_{XX}^T(\lambda)$ to obtain an estimate of the impulse response a_u. For the particular data sets at hand, because of series/length, the periodograms were computed for separate stretches of 1024 observations and then averaged together across all the data. Hanning was employed at the final stage here as well. Estimates of the sampling variability of these estimates are available (Brillinger, 1975).

The cross-correlation approach is by far the most direct and rapid. One simply computes

$$\sum_{j=1}^{N} X(t_j - u)/N \tag{6.1}$$

for $u=0,1,2,\dots$ where $t_1, t_2, \dots, t_N$ are the observed t values for which $Y_t=1$.

The cross-spectral and cross-correlation approaches do not produce direct estimates of the threshold level θ. They are further based on the assumption that the upper summation limit in expression (2.3) has no effect. To check the reasonableness of the procedure in the face of this assumption, simulations were carried out. Summation functions a_u were selected. For stretches of X_t generated for the *Aplysia* experiments the values U_t of (2.3) were formed and Y_t set to 1 if a threshold was exceeded. Various threshold values, θ were tried. Simulations were also carried out with a variety of refractory periods. The thresholds (and refractory periods) were selected to give results comparable with those obtained in the experiments. In all of these simulations the cross-spectral and cross-correlation approaches both produced estimates that were close to the "known" a_u.

The threshold character of the neuron may be investigated via the firing probability function,

$$\text{Prob}\,\{Y_t=1|U_t=U\}\ . \tag{6.2}$$

Were it a threshold device with constant threshold θ, this (conditional) probability would be given by the step function $H(U-\theta)$. Were the threshold a normal variate with mean θ and variance 1, the probability would be given by $\Phi(U-\theta)$, with $\Phi(\cdot)$ the normal cumulative, as in expression (2.1). The probability (6.2) may be estimated as follows; compute fitted values, $\hat{U}_t$, of U_t via expression (2.3); compute the proportion of t with $Y_t=1$ and $U-\alpha<\hat{U}_t<U+\alpha$, α small. Graphs of this proportion versus U are presented later in the paper and are in accord with a threshold model.

7. Results

The procedures described in this paper were applied to a broad variety of data sets obtained from identified *Aplysia* neurons. Results will be presented for cells L5 and R2 whose disimilar and even contrasting behaviors are typical of results obtained in the remaining cells (e.g. L10).

L5 is found in the left rostral quadrant of the left hemiganglion, and the results of imposing continuously varying current upon it are reported in Bryant and Segundo (1976). In the experiment whose analysis will be presented here L5 was driven by a Gaussian white noise current having ± 25 nA range and a 12.5 Hz bandpass. A 43 min long stretch of record containing some 2017 action potentials was digitized at 32 Hz for analysis. The running rate of the cell was estimated by counting the number of action potentials in 200 contiguous time intervals of about 13 s and found to fluctuate but narrowly about a constant level; this provided a check for stationarity. The autointensity function of the spike train was estimated (Bryant et al., 1973) and had the character of that of a delayed Poisson process, 0 at the smallest lags (corresponding to a refractory period of about 0.4 s and essentially constant (at the overall rate) thereafter.

Figure 1a is a graph of the function a_u estimated by the method of maximum likelihood. The shape obtained is biphasic, with a strong narrow positive-going peak just before firing and a broad peak of opposite polarity commencing at about 0.75 s. As indicated earlier this function represents the current to potential transformation as affected by the spatial transmission from the soma to the trigger zone and by the temporal decay. Under the conditions of the experiment and the nature of the model (and the relatively long spacing between spikes) it represents the response of the transmembrane potential at the trigger zone to a pulse of positivegoing current injected in the soma.

Figure 1b presents the result of estimating a_u by cross-spectral analysis. It has the same shape as that found by the method of maximum likelihood with the exception that it dies to 0 slightly earlier. This attenuation is undoubtedly caused by the replacement of U_t of (1.2) by V_t of (3.1). The cross-spectral analysis weights in 0 values for various of the a_u.

Figure 1c is the statistic (6.1) of the cross-correlation approach, because the input signal was approximately white. The computation time required to produce Fig. 1c was negligeable compared to that of the other two approaches. The advantage of driving the neuron with Gaussian white noise is apparent. Further the examples computed by cross-correlation in Bryant and Segundo (1976) are plausibly seen to have a broader interpretation as providing summation functions prior to a threshold crossing.

The second set of results to be presented involve experiments with the cell R2. Once again Gaussian "white" noise was applied as a stimulus; in addition however white noise with a uniform marginal distribution was applied too – to the identical cell that received Gaussian noise. Amongst other things it is of interest to examine how robust the fitting procedures of Sects. 3 and 4 are to departures from normality. Figure 2a provides the summation function, a_u, as estimated by the method of maximum likelihood for the case of uniform input. In this case the function is unidirectional weighting immediate values most heavily with the weighting extending perhaps 0.4 s.

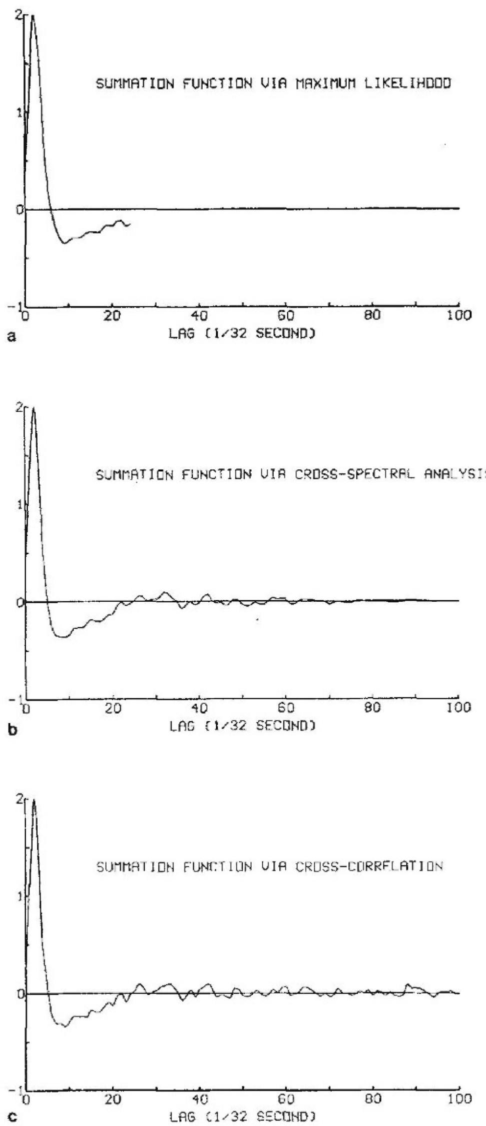


Fig. 1a–c. Estimated summation function for cell L5 driven by Gaussian noise. **a** Formed by averaging maximum likelihood estimates for 43 consecutive stretches of 1900 sampled values and then hanning. **b** Determined by cross-spectral analysis; the spectral estimates have 1200° of freedom and bandwidth 0.32 Hz. **c** Determined by cross-correlation

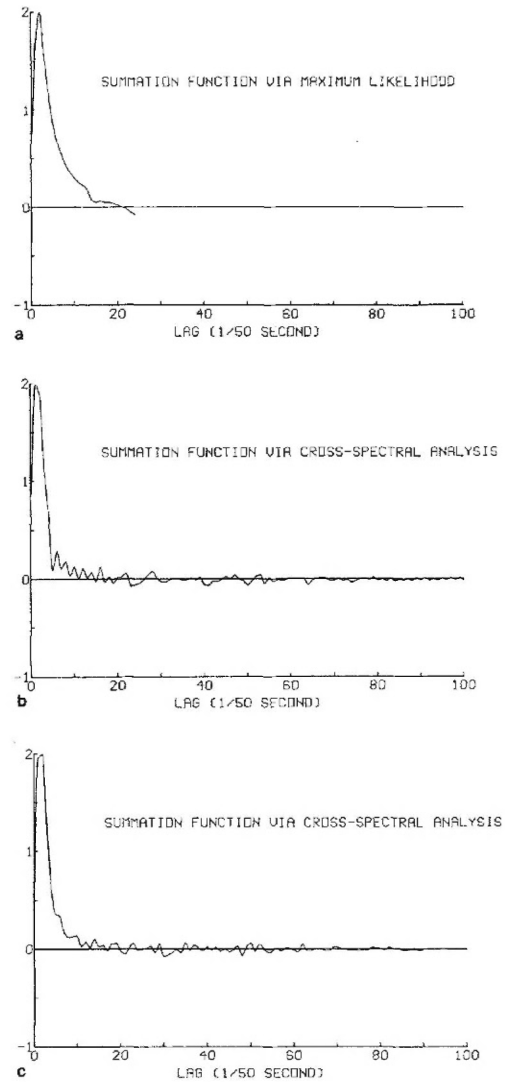


Fig. 2a–c. Estimated summation function for cell is R2. **a** and **b** Driven by uniform noise. Estimate formed by averaging maximum likelihood estimates for 25 consecutive stretches of 1900 sampled values and then hanning (**a**). Estimate determined by cross-spectral analysis; the spectral estimates have 1200° of freedom and bandwidth 0.50 Hz (**b**). **c** Driven by Gaussian noise. Estimate determined by cross-spectral analysis; the spectral estimates have 1200° of freedom and bandwidth 0.50 Hz

Whether one applies uniform or Gaussian input makes no difference as to the logic of the maximum likelihood approach. For nonlinear systems, however, statistics based on cross-spectral or cross-correlation analysis cannot be expected to be independent of the input signal characteristics. Figures 2b and c provide the function a_u as estimated in the same cell by cross-spectral analysis for the cases of uniform and Gaussian input respectively. The curves are very similar to each other and to the maximum likelihood estimate of Fig.

2a. The one substantial difference with the latter is that it tails off more slowly. This undoubtedly relates, as indicated earlier, to the cross-spectral estimate weighting in noise values beyond the last time of firing.

Figure 3 presents estimates of the probability of the neurons firing as a function of the estimated concurrent potential. The potential was estimated by

$$\hat{U}_t = \sum_{u=0}^{B_t-1} \hat{a}_u X_{t-u} \tag{7.1}$$

with the $\hat{a}_u$ values resulting from the maximum likelihood fit. The vertical line is placed at the value, $\hat{\theta}$, of the estimated threshold mean. Figure 3a is for the case of cell L5. The graph has the form of the lower half of a sigmoid function in agreement with expression (2.1). Substantial sampling fluctuations are apparent in the estimate for the larger values of the potential: the experiment was such that few values of $\hat{U}$ where observed in that region. Further no values of $\hat{U}$ were observed at all once one got a bit beyond $\hat{\theta}$: this cell (in contrast to R2 as described below) was never pushed into a region of firing with probability one, apparently.

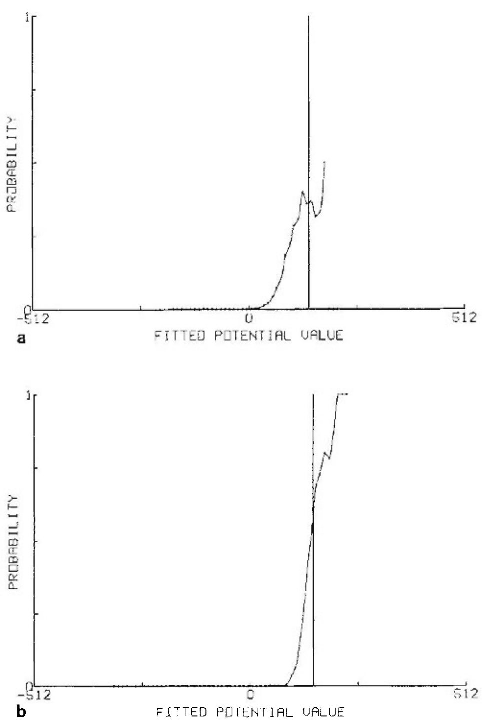


Fig. 3a and b. Estimated firing probability as a function of fitted membrane potential. The binwidth along the horizontal axis is 10.24. In **a** and **b**, the data and estimated summation function are those of Fig. 1a and 2a, respectively

Figure 3b presents the estimated firing probability for the case of cell R2 driven by uniform input. In this case a full sigmoid shape is apparent. When the potential gets large enough the cell is seen to fire with probability one. This last figure provides substantial evidence towards the reasonableness of the threshold model.

In deriving the estimates the threshold, θ_t, has been assumed to fluctuate about a constant level θ. On occasion other workers have assumed exponential decay for the threshold following an action potential. In an examination of this possibility scatter diagrams for various experiments were prepared of the values $(t_j - t_{j-1}, \hat{U}_{t_j})$ where the t_j were the observed spike times. No exponential decay was apparent. The diagrams were consistent with the relationship $\theta_t = \theta + \varepsilon_t$, the series ε_t being zero mean.

Figure 4 presents estimates of the coherence function of the input current series with the output spike train as computed in the spectral analyses of the L5-Gaussian, R2-uniform, R2-Gaussian data sets. The coherence measures the degree of linear time invariant association of the input and output as a function of frequency. Given the essential nonlinearity of the present situation, the coherences are surprisingly large. In each case the input and output appear most strongly related at the low frequencies with the relationship extending up to 3–4 Hz.

Figure 4b and c are for the cases of uniform and Gaussian input respectively. Once again the figures are surprisingly similar suggesting that the Gaussian assumption invoked in the analytic derivations of Sect. 3 and 4 is not crucial. The principal distinction between the two stimuli was found to occur in the estimates of firing probability such as those of Fig. 3. With the uniform stimuli more large values of $\hat{U}$ occurred with the effect that the estimate was more stable at the ends.

8. Conclusions

The threshold model has been fit to certain *Aplysia* neuronal data. Figure 3a and b derived from the fit provide evidence that the neurons studied are indeed firing with a probability that depends on the relationship of an internal variate, denoted by U_t in this paper and called a potential, to a threshold level. The summation function describing the transformation of presynaptic current at one site to potential at the trigger zone has been found to be estimable and as evidenced in Fig. 1 and 2 to depend upon the particular cell studied.

From a physiological viewpoint the summation function represents the potential fluctuations that would occur at the trigger zone had a positive-going current impulse been injected into the soma. At a synapse the initial change provoked by the arrival of a presynaptic spike is a local conductance increase associated with a rapid synaptic current whose electronic repercusion is the more durable postsynaptic potential, as demonstrated in lobster cardiac ganglion cells where synapses occur on both soma and dendrites (Hagiwara et al., 1959). It seems reasonable, therefore, to propose that the summation functions of Fig. 1 and 2 approximates the excitatory postsynaptic potentials that would be encountered were the synapses located upon the *Aplysia* cell soma.

The form of the summation function suggests two comments. The first is that its duration (e.g. around 0.6 s) is far smaller than that (e.g. over 1.0 s) of PSP's of comparable amplitudes observed in the same neurons. Our interpretation is that the different properties of the membranes responsible for the "soma-to-trigger zone" and the "neurophil-to-trigger zone" transformations lead to such a contrast. The second is that the summation function (in L5, for example Fig. 1) can be biphasic, exhibiting an early rapid swing of one polarity and a late slower one of the opposite polarity. This double effect can be responsible for part of the biphasic rate changes associated with PSP's (Bryant et al., 1973).

The soma-axon coupling in living neurons (Junge, 1976) has been represented by an RC soma model connected to an infinitely long cable-like axon. In *Aplysia* neurons the cell body dominates the potential response to constant injected currents simply because the somatic conductance is far greater than that of the initial axon. The situation may change somewhat when bandlimited white noise is delivered but, even if it were to change little, it is not prudent to assimilate uncritically this "soma-to-trigger zone" conversion to that occuring naturally from "neuropil-to-trigger zone" in the same neuron. The physiological interest of the present observations arises mainly from conclusions that are applicable to this large class of other neurons where the synaptic currents are generated principally in a dominant soma.

From a statistical viewpoint, it is worth noting that the functions produced by cross-spectral (or cross-correlation) analysis have the same essential time course as those produced by the specific modelling and fitting of a threshold model. The former functions may be computed much more easily and rapidly than those using the method of maximum likelihood. The results of delivering a Gaussian white noise input, used previously in several other neurophysiological systems (e.g. Marmarelis and Marmarelis, 1978), have been found here to have a broader applicability than might have been expected – a threshold model involving summation back to the previous firing only may be fit.

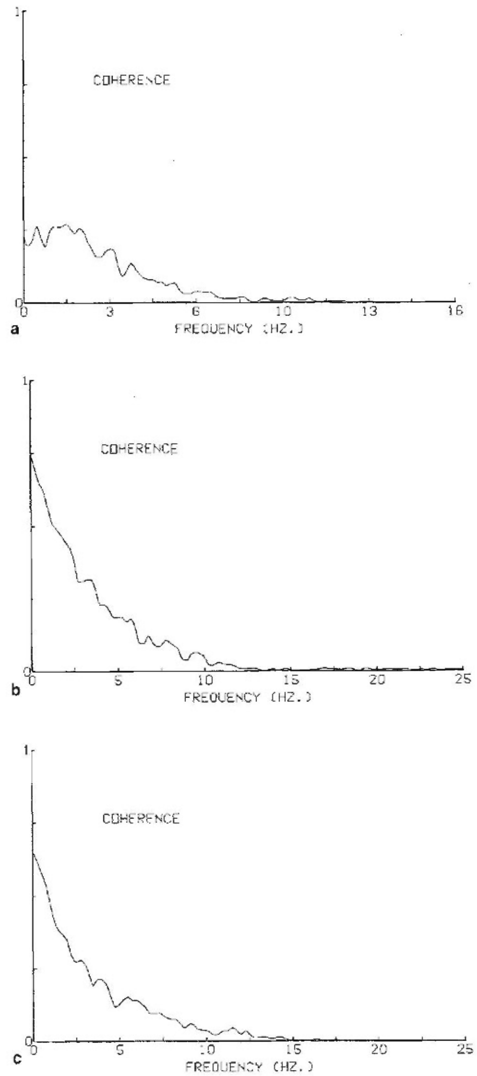


Fig. 4a–c. Estimated coherence functions **a**, **b**, and **c** were determined in the course of constructing Fig. 1b, Fig. 2b, and Fig. 2c, respectively

Further the use of input with a uniform distribution has been found to be more efficient, than a Gaussian, for the estimation of certain parameters of interest (e.g. the firing probability).

References

Boer, E. de, Kuyper, P.: Triggered correlation. IEEE Trans.Biomed. Eng. **15**, 169–179 (1968)

Brillinger, D.R.: Time series: data analysis and theory. New York: Holt, Rinehart, and Winston 1975

Brillinger, D.R.: The identification of a particular nonlinear time series system. Biometrika **64**, 509–515 (1977)

Bryant, H.L., Segundo, J.P.: Spike initiation by transmembrane current: a white noise analysis. J. Physiol. **260**, 279–314 (1976)

Bryant, H.L., Segundo, J.P.: An analysis of spike triggering in neurons using Gaussian and uniform white noise stimulation. To appear (1979)

Bryant, H.L., Ruiz Marcos, A., Segundo, J.P.: Correlations of neuronal spike discharges produced by monosynaptic connections and by common inputs. J. Neurophysiol. **36**, 205–225 (1973)

Chambers, J.M.: Computational methods for data analysis. New York: Wiley 1977

Hagiwara, S., Watanabe, A., Saito, N.: Potential changes in syncytial neurons of lobster cardiac ganglion. J. Neurophysiol. **22**, 554–572 (1959)

Holden, A.V.: Models for the stochastic activity of neurones. Berlin: Springer 1976

Junge, D.: Nerve and muscle excitation. Sunderland: Sinauer Associates 1976

Knox, C.K.: Cross-correlation functions for a neuronal model. Biophys. J. **14**, 567–582 (1974)

Korenberg, M.J.: Cross-correlation analysis of neural cascades. Proc. Ann. Rocky Mount. Bioeng. Symp. **10**, 47–71 (1973)

Marmarelis, P.Z., Marmarelis, V.Z.: Analysis of physiological systems. New York: Plenum 1978

Rao, C.R.: Linear statistical inference and its applications. New York: Wiley 1965

Rubio, J.E., Holden, A.V.: The response of a model neurone to a white noise input. Biol. Cybernetics **19**, 191–195 (1975)

Tukey, J.W.: Exploratory data analysis. Reading: Addison-Wesley 1977

Received: August 16, 1979

Prof. D.R. Brillinger
Statistics Department
University of California
Berkeley, CA 94720, USA

Nerve Cell Spike Train Data Analysis: A Progression of Technique

DAVID R. BRILLINGER*

Collections of occurrence times of events taking place irregularly in time provide a fairly common, but not broadly discussed, data type. This article is concerned with the particular circumstance of firing times in nerve cells that interact and form networks. The article reviews a progression of statistical analysis techniques: description, association as measured by moments and correlation, regression, and finally likelihood. The data is point process, but may be seen as that of regression and of multivariate analysis in standard parlance. A simple description of data collected simultaneously for one or more cells is provided.

KEY WORDS: Binary data; Nerve cell; Network; Point process; Probit analysis; Semiparametric model.

". . . the purpose of inductive reasoning, based on empirical observations, is to improve our understanding of the systems from which these observations are drawn."
Sir R. A. Fisher (1956)

The above statement sets down the spirit of applied statistics. The related goal of this article is a better understanding of the nerve cell system and the construction of better quantitative models of the neuronal firing phenomenon. On the substantive side, the author's collaborator J. P. Segundo has remarked that "the biological goal is understanding in strictly biological terms." This may be viewed as an ultimate goal. The models will change, but the biology will remain.

R. A. Fisher was central to the development of statistics, in particular to the progression of data analysis techniques from description and simple measures of association to the tools of association and regression analysis and finally to likelihood analysis. This article aims to illustrate the same progression for a data type of some contemporary interest—point process data—and to continue on to nonparametric and semiparametric likelihood analysis.

The article is concerned with a particular biological system—small networks of neurons communicating with each other and responding to stimuli. The system studied is of basic interest on both scientific and theoretical grounds. Scientific interest follows from a concern as to how the nervous system works; theoretical interest results in part from the system's strong nonlinearity.

Data from two different living preparations are studied. First discussed are some data for the cat collected by A. E. P. Villa at Lausanne, Switzerland. In Villa's experiments, cats were subjected to sound stimuli and data for eight nerve cells recorded simultaneously (Villa 1988, 1990). Also studied are simultaneous data for networks of two and three identified nerve cells (in particular cells L2, L3, L5, and L10) of *Aplysia californica* (the sea hare) collected by J. P. Segundo at the University of California, Los Angeles (Bryant, Ruiz Marcos, and Segundo 1973; Bryant and Segundo 1976). *Aplysia* is commonly studied by neurophysiologists because the nerve cells are large and accessible and a number are repeatedly identifiable.

As is the pleasant feature of most time series analyses, a broad variety of figures are presented. These figures are central to the analysis.

Important aspects of nerve cell firing not addressed in this article include spatial effects and intracellular data collection and analysis.

1. WHAT IS A NERVE CELL?

Neurons (or nerve cells) are basic building blocks of an animal's central communication system. They are input-output systems of a particular structure having important functions. It is pertinent to discuss both structure and function, because in biology often the two seem directly related. Functions include accumulating, processing, and transmitting information. A nerve cell receives messages through its *dendrites*, root-like strings susceptible to chemical stimulus. The messages propagate to the cell body, or *soma*. Out of the soma grows the axon, with many branches ending at *synapses*, the junctions of neural networks. Figure 1, taken from Cajal (1895), shows a collection of neighboring neurons. The arrows indicate the flow of information. The cell bodies are the five blobs, four of which are labeled A, B, C, and D. The axons run vertically downward from the bodies—except for E, which is an axon entering from a distance. The dendrites include the three treelike structures at the top susceptible to influence from E.

The dendrites absorb input from other neurons through chemical processes that change ionic conductances and thereby induce current flows. The input is thence converted to a *membrane potential* throughout the soma. At the *axon hillock* (or trigger zone), the membrane potential occasionally reaches a threshold and the neuron fires, that is, generates an *action potential* (or spike). This action potential propagates along the axon to synapses, at which point a chemical transmitter is released to affect other neurons. The action potentials are of near-identical size and shape; see the spikes in Figure 2, which shows measured voltage fluctuations within cell R2 of *Aplysia* (Bryant and Segundo 1976). It may be argued that, because of reduced sensitivity to noise, the firing

* David R. Brillinger is Professor of Statistics, University of California, Berkeley, CA 94720. This article formed the R. A. Fisher Lecture given in Atlanta, Georgia, in August 1991. The research was partially supported by the National Science Foundation Grant DMS-8900613. The computations were performed at the Statistical Computing Facility, University of California, Berkeley, under the directorship of Leo Breiman. The figures were prepared employing S; see Becker, Chambers, and Wilks (1988). The author thanks the many individuals who provided help and advice with the computing and the presentation of the material. The author particularly thanks J. P. Segundo, Department of Anatomy & Cell Biology, UCLA, who for almost 20 years has helped him with the intricacies of the pertinent neurophysiology.

Journal of the American Statistical Association
June 1992, Vol. 87, No. 418, 1991 R. A. Fisher Memorial Lecture

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_33, © Springer Science+Business Media, LLC 2012

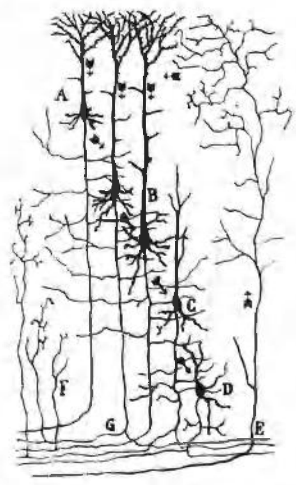


Figure 1. Drawing of Cajal (1895) Illustrating a Network of Five Cells of the Cerebral Cortex. The arrows (↟) suggest that the input arrives along the fiber E and progresses from it both directly and indirectly to the cells A, B, C, and D of different types.

times are the crucial variates in communication among neurons. Some discussion of the reduction to point processes is given in Segundo, Altshuler, Stiber, and Garfinkel (1991).

Synaptic connections may be *excitatory* or *inhibitory;* that is, depending on the type of connection, the firing of one neuron may make a second neuron either more likely or less likely to fire. Neurons also may fire spontaneously with no outside stimulus. Further is the phenomenon of *refractoriness,* wherein after a neuron has fired, the chance of it firing again is reduced (perhaps to zero) for a period.

Questions of interest include:

1. Can an analytic model incorporating the basic features of neuron behavior be developed and fit?
2. Given the firing times of a network of neurons, can one infer their causal connections?

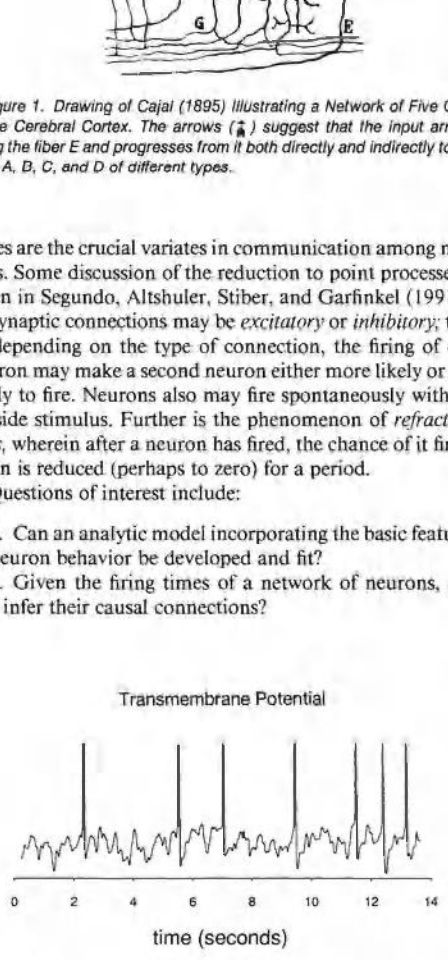


Figure 2. Fluctuating Intracellular Voltage of the Cell R2 of Aplysia Showing the Occurrence of Point Process Data. The amplitudes of the spikes are approximately 100 millivolts. Figure adapted from Bryant and Segundo (1976).

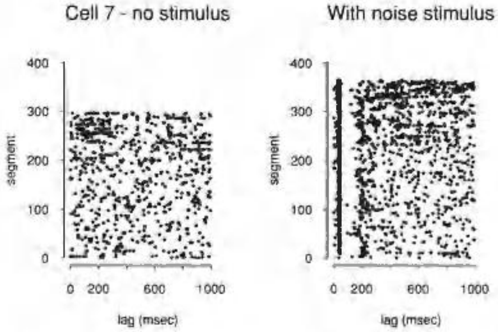


Figure 3. Rastor Plots Providing Nerve Cell 7 Firing Times. Each row of asterisks represents the events in a time interval of 1,000 msec. In the left panel, there was no experimental stimulus. In the right panel, a noise stimulus was applied at the beginning of each time interval, so each column represents the events occurring at the indicated lag after the noise stimulus.

General references for pertinent neurophysiological background information include Koch and Segev (1989), Segundo (1968, 1984, 1986), Segundo et al. (1991), and Stein (1972).

2. WHAT ARE POINT PROCESS DATA?

A stretch of point process data is a set of ordered numbers

$$\tau_1 \leq \tau_2 \leq \cdots \leq \tau_K,$$

to be thought of as the times of events that occurred in some time interval, say $(0, T)$. Usual examples are the times of telephone calls and the times of particle emission by some radioactive material. A naive descriptive statistic derived from such data is the observed rate, given here by K/T. This statistic has dimensions of counts per unit of time and is useful in elementary comparisons of point process behavior. For the data studied in this article, the rates range from about 1 spike per second to about 20 spikes per second. Figure 2 shows 7 spikes in about 14 seconds.

Descriptive statistics conducive to insight are provided by the plots in Figure 3. These plots are based on data collected in experiments studying the auditory system of the cat. Microelectrodes were inserted in a cat's brain at a location related to hearing. The plots refer to firing times for a single particular nerve cell (cell 7) which the probe happened upon. In the case of the lefthand plot, there is no applied stimulus. To describe the plot, suppose that the observation period is broken up into L segments, each 1,000 milliseconds long. Let τ_{kl} refer to the time elapsed since the start of the lth segment, of the kth spike of that segment. The points plotted are now $\{(\tau_{kl}, k), k = 1, \ldots, K_l\}$ for $l = 1, \ldots, L$. No dramatic structure is apparent in the lefthand panel. The second panel of Figure 3 refers to the same experiment but with a noise stimulus introduced into the ears of the cat every 1,000 milliseconds. The points are plotted as before, with τ referring to the time elapsed since each stimulus presentation. This picture shows that this neuron typically fires a short time after the application of the stimulus. Then there is a time period during which the neuron is unlikely to fire

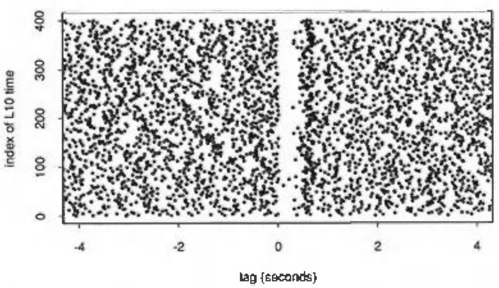


Figure 4. Times of Neuron L3's Firings Relative to L10's Firings. Each row corresponds to a single firing of L10.

and perhaps then a rebound period when the cell is more likely to fire. Plots such as those in Figure 3 are known as *rastor plots.*

A second set of experimental data of some interest comes from experiments with *Aplysia,* the sea hare. Suppose that firing times are available for two related neurons—in the analysis to be presented, neurons L3 and L10 of *Aplysia.* Let $\{\sigma_j\}$ represent the firing times of L10 and $\{\tau_k\}$ represent the firing times of L3. In the case of these neurons, it "has been demonstrated almost beyond reasonable doubt" that L10 drives L3 (see Bryant, Ruiz, Marcos, and Segundo 1973 p. 205). Figure 4 plots the points $\{(\tau_k - \sigma_j, j), k = 1, 2, \ldots, K_j\}$ for $j = 1, 2, 3, \ldots$. This plot is consistent with the idea that firing of L10 tends to inhibit firing of L3. There is an indication of a brief acceleration or rebound at a lag of about .5 second. The bulk of the points appear randomly distributed.

To progress with the analysis, it is convenient to introduce some probability structure. A *stochastic point process* is a random process whose realizations are collections of points $\{\tau_k\}$, ordered by $\tau_k \leq \tau_{k+1}$, on the interval $(-\infty, \infty)$. Such a process can be described by giving the joint distributions of all the $N(I_1), \ldots, N(I_J)$, where I_j is a Borel set and $N(I_j)$ is the number of points falling in I_j for $j = 1, \ldots, J$ and $J = 1, 2, \ldots$. The process is said to be *stationary* when the joint distributions are unaffected by simple time translation, $I \rightarrow I + t$. An alternate way to describe a point process is via the joint distributions of the intervals $Y_k = \tau_{k+1} - \tau_k$ between successive points. In the stationary case, the *rate* of the process is given by $E\{N(I)\}/|I|$, where $|I|$ is the length of the interval.

It is worth remarking that there are many similarities between the concepts and techniques of time series analysis and those of point process analysis; see Brillinger (1978), as well as the classic reference for the analysis of point process data, Cox and Lewis (1966).

3. ASSOCIATION—SECOND ORDER MOMENTS

In the case of a bivariate stochastic point process (M, N) with components $M = \{\sigma_j\}$ and $N = \{\tau_k\}$, one can define the *cross-intensity function*

$$\lim_{h \downarrow 0} \Pr\{N \text{ point in } (u + t, u + t + h] \,|\, M \text{ point at } t\}/h.$$

This will be a function of lag u alone in the stationary case This parameter may be estimated by

$$\frac{\#\{u + \sigma_j < \tau_k \leq u + \sigma_j + h\}}{\#\{\sigma_j\}h} \tag{3.1}$$

for small $h > 0$. Figure 5 gives the estimate for the data of Figure 4. It is essentially the histogram of the $\{\tau_k - \sigma_j\}$ and comes from counting the points in vertical strips of Figure 4. In fact, because of simpler sampling properties (including more stable variance, more symmetric distribution, and more near normal distribution), it is often more convenient to plot the square root of the estimate (Brillinger 1976); this was done here. The horizontal dashed lines provide ±2 standard error limits set about 0. The diagram shows a period of initial inhibition after L10's firing, followed by a rebound at about .3 second. In some sense, Figure 5 does not add new information to that of Figure 4; but it does provide a specific way to interpret and assess the phenomena that occur. This cross-intensity function provides a precise measure of second-order association in the stationary case.

If two processes are associated, one can anticipate that functions of their realizations will be correlated. A particular function to study, because of its simplifying characteristics, is the empirical Fourier transform. Consider the Fourier transforms of two stretches of point process data, specifically $\sum_{0<\sigma_j\leq T} e^{-i\lambda\sigma_j}$, $\sum_{0<\tau_k\leq T} e^{-i\lambda\tau_k}$ for $0 \leq \lambda < \infty$. The quantity

$$R_{MN}(\lambda) = \lim_{T\to\infty} \operatorname{corr}\{\sum e^{-i\lambda\sigma_j}, \sum e^{-i\lambda\tau_k}\}$$

is called the *coherency* at frequency λ. Its modulus-squared, $|R_{MN}|^2$, is called the *coherence.* The coherence lies between 0 and 1 and measures the extent of linear time invariant association between two processes (Brillinger, Bryant, and Segundo 1976).

Figure 6 provides an estimate of the coherence for the L10-L3 data above. The estimate is seen to be highest for frequencies $\lambda/2\pi$ less than 1 cycle per second. The dashed line in the figure gives the (approximate) 95% upper point

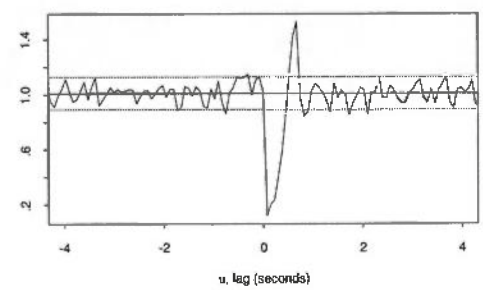


Figure 5. The Square Root of the Cross-intensity Statistic (3.1). The dashed lines give upper and lower two standard error limits placed about 0 level.

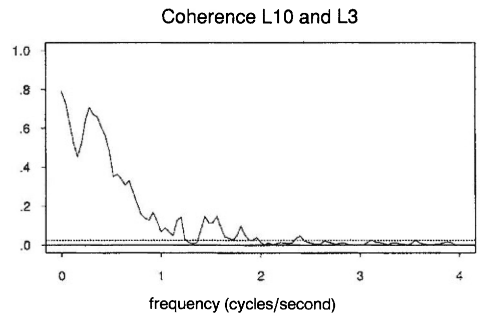


Figure 6. An Estimate of the Coherence of Neurons L10 and L3 Obtained in the Fashion Described in Brillinger, Bryant, and Segundo (1976). Dashed line gives 95% null point.

of the null distribution of the estimate. Except in the case of simple translation of all the events by a common amount, mappings between realizations of point processes are inherently nonlinear. In view of this, the high magnitude here of the coherence estimate at the low frequencies is surprising.

4. REGRESSION—A LINEAR MODEL

Consider next a model

$$\lim_{h \downarrow 0} \Pr\{N \text{ point in } (t, t + h] \,|\, M\}/h = \mu + \sum_j a(t - \sigma_j). \tag{4.1}$$

This model is linear and time-invariant. The function $a(\cdot)$ is meant to represent the various chemical, electrical, spatial, and temporal delay processes involved in the influence of neuron M's firing on neuron N's firing. For example, if the τ's were given by $\tau_j = \sigma_j + Y_j$, with the Y's independent and of density function $a(\cdot)$, then the result (4.1) would hold with $\mu = 0$, see Brillinger (1974). The model (4.1) may be fit by cross-spectral analysis (Brillinger 1974). The resulting estimate of $a(\cdot)$ for the Aplysia data addressed in Section 3 is given in Figure 7. The estimate is seen to mimic that in

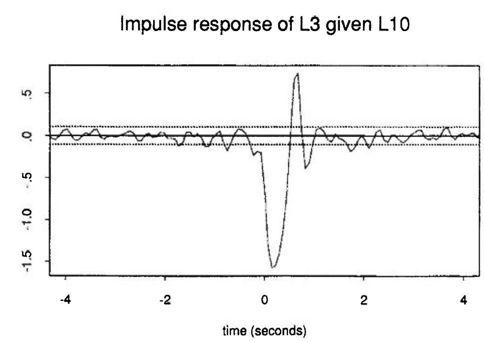


Figure 7. An Estimate of the Function a(·) of (4.1) Obtained in the Fashion Described in Brillinger, Bryant, and Segundo (1976).

Figure 5. The distinction is that, as is the case in ordinary regression analysis, one is nearer to an object unaffected by elementary reexpressions. This analysis for this particular data set is not dramatically enlightening, but interesting examples may be found in Brillinger, Bryant, and Segundo (1976). The following section presents a more satisfying analysis of the present data in any case.

5. LIKELIHOOD—CONCEPTUAL MODELING

A model with a long history in neurophysiology involves a neuron firing when the membrane potential at its trigger zone exceeds a *threshold.* The threshold is a time-varying quantity that is reset to a high level on the neuron's firing and then is subject to slow (although not always monotonic) decay. The effect of the reset is to prevent firing from recurring immediately, and thus to incorporate the phenomenon of refractoriness. The model may be described in formal terms as follows: Let $M = \{\sigma_j\}$ refer to the times at which a first (or input) neuron fires. Given the function $a(\cdot)$, consider the following time-varying state variable

$$U(t) = \sum_{\sigma_j \le t} a(t - \sigma_j). \tag{5.1}$$

The quantity $U(t)$ is meant to represent the membrane potential at time t at the trigger zone of the neuron whose firing is of interest. Here, $a(\cdot)$ is a *summation function,* meant to represent the various processes involved in the influence of M's firing on N's firing. The character of the function affects whether the firing of the neuron M increases (excites) or decreases (inhibits) the chance of the neuron N firing. The threshold decay is represented by the function $b(\cdot)$.

Figure 8 provides a layout of the situation. The bottom two panels give hypothetical $a(\cdot)$ and $b(\cdot)$ for the case of an inhibitory synapse. (Shortly, empirical estimates of $a(\cdot)$ and $b(\cdot)$ will be provided.) The vertical asterisks of the top plot are the firing times of the input neuron, M. The hook-shaped curves are the translates of the function $b(\cdot)$, with a new translate introduced with each firing of the principal neuron, N. If γ_t denotes the time elapsed since N's last firing, then the threshold curve may be represented by $\theta(t) = b(\gamma_t)$. The lower continuous curve of the figure is $U(t)$. One is concerned with the membrane potential, $U(t)$, crossing $\theta(t)$.

Consideration turns to developing a stochastic version of this model and of a corresponding likelihood function to employ in analyzing available data. Suppose first that the point processes are simplified to discrete time ($t = 0, \pm 1, \pm 2, \cdots$) and to 0–1 valued series. That is, a sampling interval of small length is selected such that only 0 or 1 points occur within each interval, and one defines $M_t = 1$ if there is a point in the unit interval starting at t and $M_t = 0$ if there is no point, for $t = 0, \pm 1, \pm 2, \ldots$. Corresponding discrete versions of N and $a(\cdot)$ are similarly defined. Now

$$U(t) = \sum a(t - \sigma_j) \approx \sum a_{t-u} M_u, \tag{5.2}$$

and it is convenient to represent the effect of the threshold by

$$\theta_t = \sum_{v=1}^{\gamma_t} b_v N_{t-v}, \tag{5.3}$$

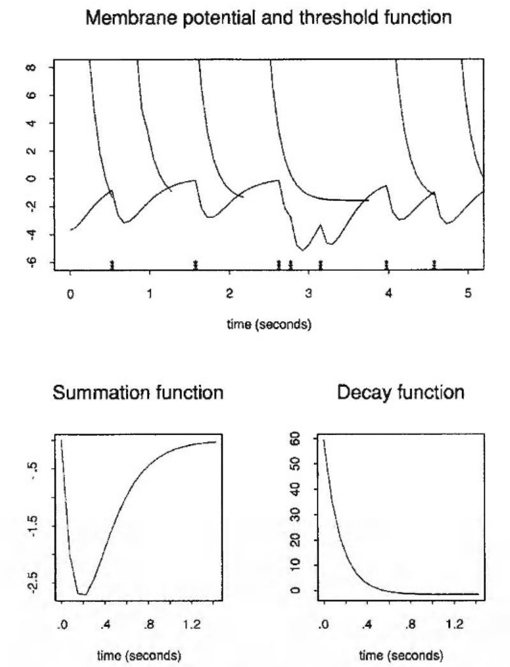


Figure 8. The Threshold Model. The lower curve of the top panel gives U(t) of (5.1) with a(·) given by the lower left function. The hook-shaped functions of the top panel are translates of the function of the lower right panel initiated each time the curve U(t) is crossed. The spikes of the top panel are the times of M firing.

with γ_t again the time elapsed since the last N firing. (That the expression (5.3) is linear in the parameters aids in their estimation.)

Suppose that there is noise, with c.d.f. $P(\cdot)$, superposed on the threshold. This makes the model stochastic. The conditional probability of the neuron firing, given the past, is taken to be

$$P_t = \Pr\{N_t = 1 \mid \text{the past}\} = P(\psi_t), \qquad (5.4)$$

where

$$\psi_t = \sum a_u M_{t-u} - \theta_t.$$

The log-likelihood is

$$\sum [N_t \log P_t + (1 - N_t)\log(1 - P_t)]. \qquad (5.5)$$

Estimates of the a's and b's may now be determined by the maximization of (5.5), employing iteratively reweighted least squares algorithms such as those described in McCullagh and Nelder (1989).

Figure 9 presents the results of these computations, taking $P(\cdot)$ to be $\Phi(\cdot)$, the standard normal cumulative (as in probit analysis) and the sampling interval to be .075 seconds. The estimated summation function $\hat{a}_u$ is seen to swing negative directly. This corresponds to M's (or L10's) firing inhibiting the N's (or L3's) firing. This effect of L10 appears to last for approximately one second. No apparent rebound effect is present. The estimate of the decay function $\hat{b}_u$ is ∞ for the first five coefficients, reflecting the fact that no output spikes occurred closer than .49 second for this particular data set. The standard errors are estimated via the usual formulas of probit analysis. For convenience of display, in the case of $\hat{a}_u$ the errors are graphed about the horizontal axis.

The preceding analysis involved the assumption that the perturbing noise values had a standard normal distribution. Suppose, however, that the noise comes from an unknown distribution and that it is desired to estimate that distribution. It is convenient to write that distribution as

$$P(\psi) = \Phi(g(\psi)), \qquad (5.6)$$

with the consequence that $g(\cdot)$ will be linear if the noise is in fact normal. (The function $g(\cdot)$ is not assumed monotonic here.)

The estimation procedure employed in this case is locally weighted maximum likelihood. The computations are carried out recursively. To begin, set $\hat{g}(\psi) = \psi$ and $\hat{g}'(\psi) = 1$.

Step 1. Given N_t, $\hat{g}(\cdot)$, $\hat{g}'(\cdot)$ obtain estimates of the remaining parameters of the model, and in particular $\hat{\psi}_t$, by ordinary maximum likelihood.

Step 2. Given N_t, $\hat{\psi}_t$ obtain $\hat{g}(\cdot)$, $\hat{g}'(\cdot)$ to maximize the locally weighted log-likelihood

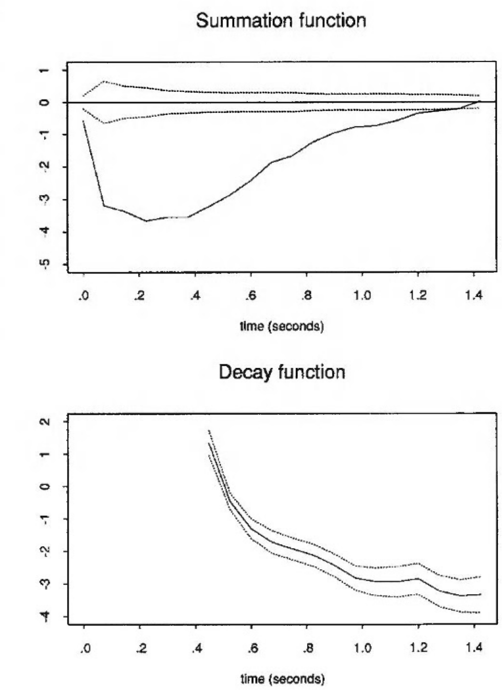


Figure 9. Estimates of the Functions a_u and b_u of (5.2) and (5.3). The dashed lines provide two standard error limits.

$$\sum w(\psi - \hat{\psi}_t)[N_t \log P_t + (1 - N_t)\log(1 - P_t)], \quad (5.7)$$

with $w(\cdot)$ a weight function concentrated near 0 and with $g(\psi) = \alpha + \beta\psi$ assumed (locally) linear. (This assumption of linearity means that except for the additional weight term, the computations are usual probit ones.) The weight function focuses the local estimation towards the center of the function's support. The estimate of $g(\psi)$ is now taken to be $\hat{\alpha}_\psi + \hat{\beta}_\psi\psi$; the estimate of the derivative, $\hat{\beta}_\psi$.

Step 3. Return to Step 1 until convergence is achieved.

The function estimation procedure of Step 2 here may be found, at various stages of development, in Gilchrist (1967), Cleveland and Kleiner (1975), Brillinger (1977), Cleveland (1979), Hastie and Tibshirani (1984), Tibshirani and Hastie (1987), and Staniswalis (1989). An early version of GAIM (Almudevar and Tibshirani 1990) gave the author confidence that this procedure was feasible for the present situation. The weight function of (5.7) was taken to be the tricube, as in Cleveland and Devlin (1988).

Figures 10 and 11 present the results of these computations. The dashed lines give estimated ±2 standard error limits. In the case of $\hat{g}'(\cdot)$, they are placed about the level 1.0. The derivative estimate $\hat{g}'(\cdot)$ is seen to not deviate much from 1.0 in the region of apparent probability mass. The computations are seen to support an assumption of linearity of $g(\cdot)$ and hence of normality. This is further reflected in

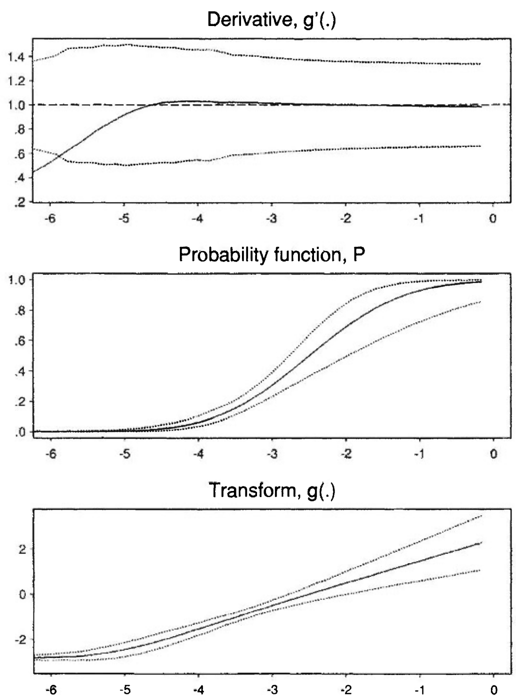


Figure 10. Estimates of the Functions $g(\cdot)$ and $P(\cdot)$ of (5.6) and of the Derivative of $g(\cdot)$

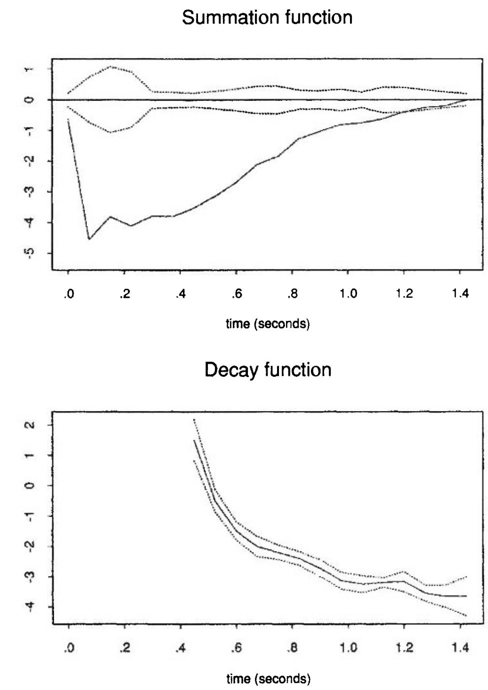


Figure 11. Estimates of a_u and b_u for the Case of Unknown $P(\cdot)$

the similarity of Figures 9 and 11 giving the respective estimates of a_u and b_u. The approximate standard errors were determined via the jackknife (Mosteller and Tukey 1977). In this case, replicates were based on 99% of the data, and 20 replicates were formed.

Consideration next turns to an alternate type of experiment involving *Aplysia* with a different stimulus and a correspondingly altered state variable. In the experiment, noise current is fed directly into the neuron L5 and evoked spike times are recorded. Some input and corresponding output are provided in Figure 12. Numerous neurophysiological experiments have suggested that neuronal firing depends on more than a single-state variable, such as the membrane potential's crossing a threshold. For example, the speed of the crossing, perhaps quantified via the derivatives of the functions involved, also appears to be pertinent (Segundo 1968). The preceding threshold model suggests consideration of the state variable

$$U(t) \approx \sum_{u<t} a_{t-u}X_u, \quad (5.8)$$

with X the input noise and ψ_t the corresponding linear predictor

$$\psi_t = U_t - d - e\gamma_t - f\gamma_t^2 - g\gamma_t^3, \quad (5.9)$$

where θ_t is here restricted to have cubic form. (In these computations it was convenient to take the threshold decay func-

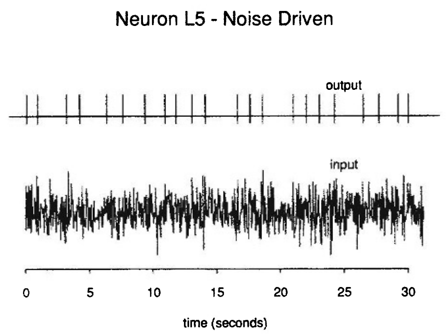


Figure 12. Input and Output of a Neuron. The neuron L5 of Aplysia is stimulated directly by the Gaussian noise of the lower panel and fires as in the upper panel.

tion to be cubic in order to avoid excessive computations.) Consider also a second state variable

$$\nu_t = \sum c_u X_{t-u}. \tag{5.10}$$

Suppose further that

$$\Pr\{N_t = 1 \mid \text{the past}\} = \Phi(\psi_t)\Phi(\nu_t) \tag{5.11}$$

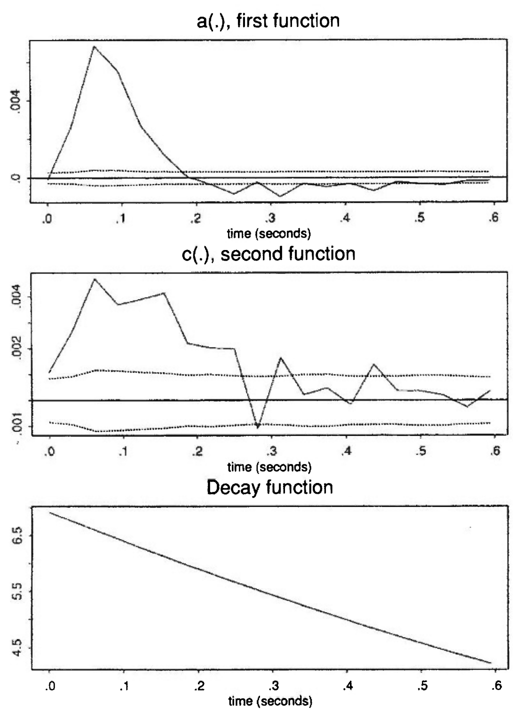


Figure 13. Estimates of a_u and c_u of (5.8) and (5.10) and of the Cubic Decay Function of (5.9).

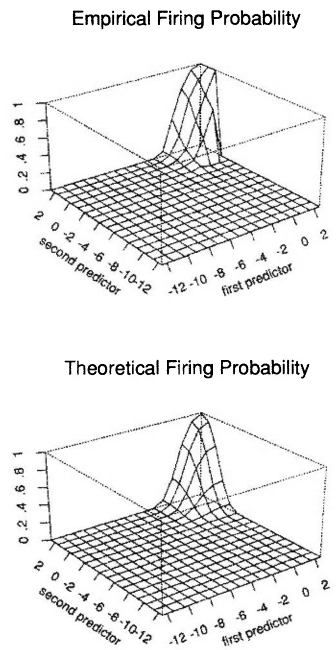


Figure 14. Firing Probability. The bottom panel gives the right side of (5.11). The top panel provides the observed proportion of times the neuron fires as a function of the first and second linear predictor values.

as a naive extension of (5.4). It is assumed that approximating the actions of the two state variables as independent will not lead to wildly deviating estimates. Figure 13 gives the results of fitting this model. The fitting here is carried out iteratively, first assuming the coeficients of ψ_t given and estimating those of ν_t, then assuming the coeficients of ν_t given and estimating those of ψ_t. In both cases, the estimation procedures are probit. The second panel gives the estimate of c_u with two standard error limits set about 0. There is evidence for the presence of a second state variable, although in the case of the present computations it does not have the appearance of the derivative of the first. The estimate of a_u given in the first panel shows how the noise current is exciting the neuron.

The problem of assessing goodness of fit has not yet been commented on. Figure 14 provides an informal procedure for the model (5.11). The top panel is a plot of (5.11), the bottom panel gives the empirical firing probability as a function of the first and second predictors. To obtain this, one bins the values of the predictors and computes the corresponding proportion of firing occurrences. The agreement does seem reasonable. One could proceed to formal goodness-of-fit tests based on the quantities just graphed, such as chi-squared statistic, but this seems premature because the temporal dependency leaves the sampling properties in doubt.

Brillinger and Segundo (1979) fit the threshold model to some *Aplysia* data by maximum likelihood. Brillinger

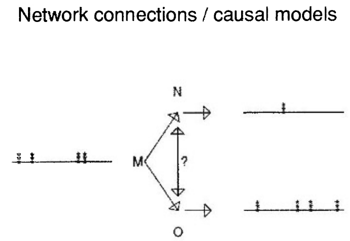


Figure 15. A Network of Three Neurons. Neuron M influences neurons N and O, but one wonders if there is a direct connection from N to O or vice versa.

(1988b) provides a number of references to the threshold modeling of nerve cells' actions and presents further empirical examples.

6. NETWORKS—3 CELL

Suppose one has three neurons, M, N, O, which may be influencing each other. In the experiment analyzed below (see Brillinger, Bryant, and Segundo 1976), it was understood that neuron M was driving both neurons N and O, but it was not known if there were direct connections from N to O or vice versa. The scheme of the situation has been illustrated in Figure 15. One tool for addressing questions of connectivity is partial coherence. The partial coherency at frequency λ of point processes M and N, given the point process O, is defined as

$$R_{NO|M} = \frac{R_{NO} - R_{NM}R_{MO}}{\sqrt{(1 - |R_{NM}|^2)(1 - |R_{OM}|^2)}}. \qquad (6.1)$$

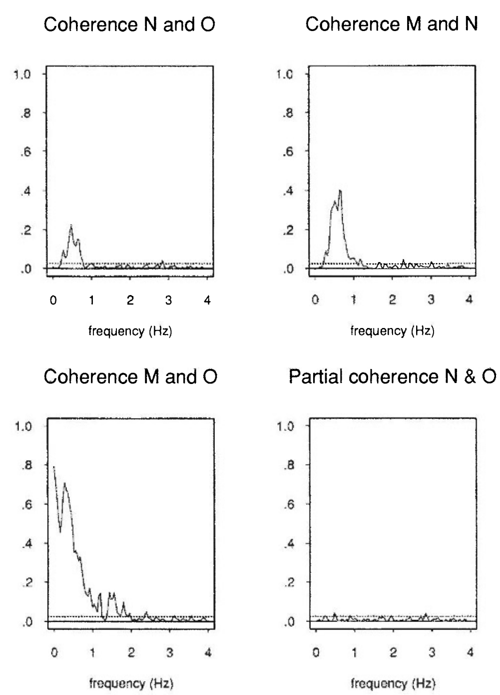


Figure 16. Coherences of Three Neurons. The first three panels provide estimates of the indicated coherences. The final panel is an estimate of the partial coherence of N and O "removing" the effects of the input M. The dashed line gives the upper 95% point of the null distribution.

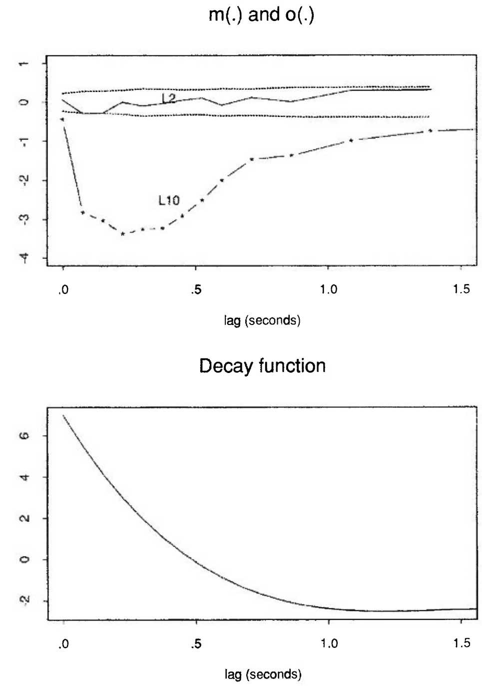


Figure 17. Estimates of m(·) and o(·) of (6.2) and of the Cubic Decay Function of (6.3).

Here, R_{NO} denotes a coherency of two stationary point processes as before. Dependence on λ has been surpressed to simplify the display (6.1). The partial coherency may be interpreted via

$$R_{NO|M} = \lim_{T\to\infty} \operatorname{corr}\{d_N^T - \alpha d_M^T, d_O^T - \beta d_M^T\},$$

with

$$d_M^T(\lambda) = \sum_j e^{-i\lambda\sigma_j},$$

for example, as before. Here α and β are the regression coefficients of d_N^T on d_M^T and of d_O^T on d_M^T. The intent of their inclusion is to remove the (linear) effects of the Fourier transform of M from those of N and O.

Figure 16 provides the results of such computations for data on a network of cells O = L2, N = L3, and M = L10 of *Aplysia.* The particular experiments are discussed in Brillinger et al. (1976). The effect of the analysis is quite dramatic.

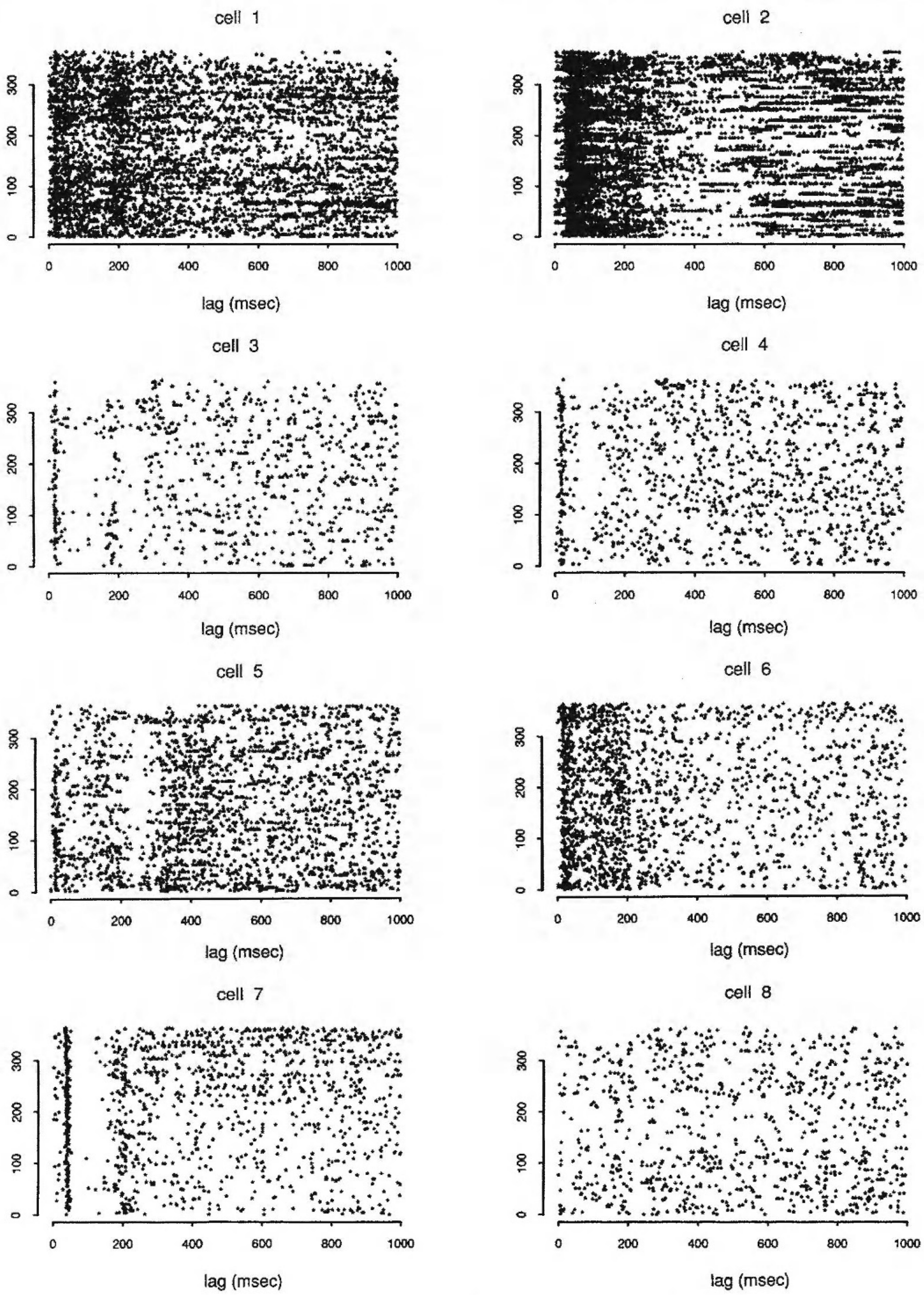


Figure 18. Rastor Plots of the Firings of Eight Cells Following Application of a Noise Stimulus Every 1,000 msec.

From the fourth panel, one can infer that the apparent association of cells N and O, as shown in the first panel, is due to their common association with cell M.

This problem can also be addressed from a likelihood approach by employing a threshold model. Suppose the firing times of cell M are denoted by $\{\sigma_j\}$ and those of cell O by $\{\rho_l\}$. Consider the membrane potential of cell N at time t to be given by

$$U(t) = \sum_j m(t - \sigma_j) + \sum_l o(t - \rho_l), \tag{6.2}$$

and suppose

$$\Pr\{N_t = 1 \mid \text{the past}\} = \Phi(U_t - d - e\gamma_t - f\gamma_t^2 - g\gamma_t^3), \tag{6.3}$$

γ_t being the elapsed time since N last fired. Here, $m(\cdot)$ and $o(\cdot)$ are summation functions associated with the effects of neurons M and O. One wonders if the function $o(\cdot) \equiv 0$.

Figure 17 on page 267 gives the maximum likelihood estimates of m_u, o_u, and the decay function. The two standard error limits for the cell O = L2, set at about 0, suggest an insignificant effect. This is consistent with the results of the coherence analysis. One could do a similar analysis relating O to M and N and achieve the same result.

Various references relating to network analysis are given in Brillinger (1988a), as are further examples. Tick (1963) is an early reference to partial coherence analysis. Gersch (1972) discusses empirical partial coherence analysis as a tool to study causality in electrophysiological signal analysis. More examples are provided in Rosenberg, Amjad, Breeze, Brillinger, and Haliday (1989).

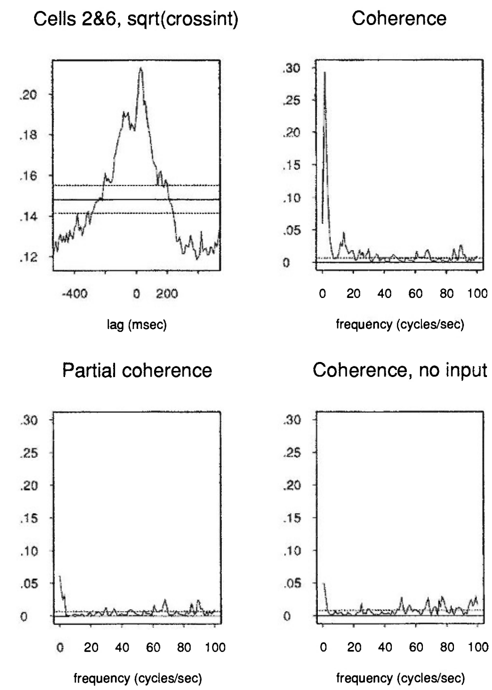


Figure 20. Statistics to Investigate the Association of Cells 2 and 6.

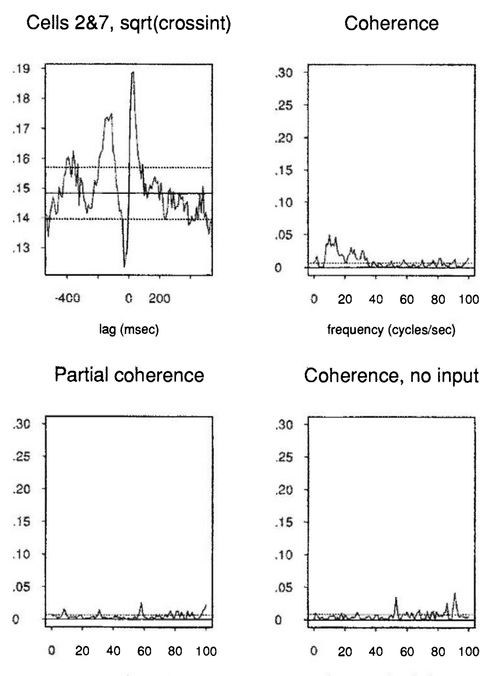


Figure 19. Statistics to Investigate the Association of Cells 2 and 7.

7. NETWORKS—8 CELL

In the next analyses presented (albeit preliminarily, as this is work in progress), data were collected in an attempt to understand the auditory pathways of the cat. Microelectrodes were inserted with location tuned to an apparent response to sound and to anatomical knowledge, and responding neurons were located.

The animal was stimulated by white noise bursts of 200 msec duration and at the rate of one per second, through speakers inserted in the ears. The stimulus was applied 364 times. For the eight cells located, Figure 18 on preceding page provides rastor displays of firing times for lags up to 1,000 msec following stimulus application. Various behaviors are exhibited, ranging from the strong association of cells 1, 2, and 6 to the weak association of cell 8. One sees excitation, inhibition, and rebounding.

This work has defined various measures of association of point processes. Figures 19–21 provide them for a selected three of the 28 possible cell pairs. In Figure 19, concerning cells 2 and 7, the cross-intensity and coherence show association. Not much is present, however, when the stimulus is "removed" by partial coherence analysis. This inference is

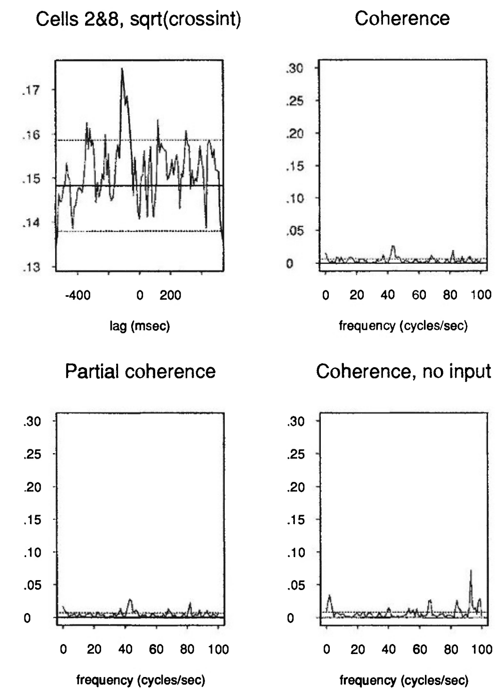


Figure 21. Statistics to Investigate the Association of Cells 2 and 8.

confirmed by the directly measured coherence between the two cells in the case of no applied experimental noise stimulus. Figure 20 provides the same information for cells 2 and 6. Again, the cross-intensity and coherence estimates show association. In this case, however, the partial coherence does suggest that the cells are related beyond the dependence introduced by the common noise stimulus. This inference is again confirmed by the coherence for the case of no experimental stimulus. Figure 21, based on cells 2 and 8, suggests little, if any, connection for these cells. This is consistent with the apparent weak dependence of cell 8 on the stimulus, as shown in Figure 18.

8. DISCUSSION AND SUMMARY

The article has sought to follow the historical statistical progression of description, association, regression, and likelihood analysis. It then continues to the contemporary topics of semiparametric maximum likelihood and causal structure recognition. The data is of a particular type—point process—and is taken from the field of neurophysiology. The paper has illustrated that a calculus is available for point process data analysis and that the calculus allows the computation of standard errors to provide uncertainty measures.

It has been seen that linear techniques—specifically coherence analysis—can elucidate highly nonlinear situations. It has also been seen that stochastic models incorporating basic features of neuron firing and network connections can be set down.

Work ahead includes inferring causal connections for the 8-cell cat network (taking note of the issues and techniques mentioned in Wold 1956, for example), maximum likelihood analysis of the cat data, modeling at the ionic level and, as is topical in contemporary statistical work, improving estimates by borrowing strength (e.g., via random effects models).

[*Received September 1991. Revised October 1991.*]

REFERENCES

Almudevar, T., and Tibshirani, R. (1990), *GAIM, Personal Computer Software for Generalized Additive Interactive Modelling,* Toronto: S. N. Tibshirani Enterprises.

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988), *The New S Language,* Pacific Grove, CA: Wadsworth.

Brillinger, D. R. (1974), "Cross-Spectral Analysis of Processes With Stationary Increments Including the $G/G/\infty$ Queue," *The Annals of Probability* 2, 815–827.

——— (1976), "Estimation of the Second-Order Intensities of a Bivariate Stationary Point Process," *Journal of the Royal Statistical Society,* Ser. B, 38, 60–66.

——— (1977), "Consistent Nonparametric Regression," Discussion of Stone, *The Annals of Statistics* 5, 622–623.

——— (1978), "Comparative Aspects of the Study of Ordinary Time Series and of Point Processes," in *Developments in Statistics 1,* ed. P. R. Krishnaiah, New York: Academic Press, pp. 33–133.

——— (1988a), "Maximum Likelihood Analysis of Spike Trains of Interacting Nerve Cells," *Biological Cybernetics,* 59, 189–200.

——— (1988b), "The Maximum Likelihood Approach to the Identification of Neuronal Firing Systems," *Annals of Biomedical Engineering,* 16, 3–16.

Brillinger, D. R., and Segundo, J. P. (1979), "Empirical Examination of the Threshold Model of Neuron Firing," *Biological Cybernetics,* 35, 213–220.

Brillinger, D. R., Bryant, H. L., and Segundo, J. P. (1976), "Identification of Synaptic Interactions," *Biological Cybernetics,* 22, 213–228.

Bryant, H. L., Ruiz Marcos, A., and Segundo, J. P. (1973), "Correlations of Neuronal Spike Discharges Produced by Monosynaptic Connexions and by Common Inputs," *Journal of Neurophysiology,* 36, 205–225.

Bryant, H. L., and Segundo, J. P. (1976), "Spike Initiation by Transmembrane Current: A White-Noise Analysis," *Journal of Physiology,* 260, 279–314.

Cajal, S. Ramon y (1895), *Les Nouvelles Idées sur la Structure du Système Nerveux Chez L'Homme et Chez les Vertebrés,* Paris: Reinwald.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association,* 74, 829–836.

Cleveland, W. S., and Devlin, S. J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by local Fitting," *Journal of the American Statistical Association,* 83, 596–610.

Cleveland, W. S., and Kleiner, B. (1975), "A Graphical Technique for Enhancing Scatterplots With Moving Statistics," *Technometrics,* 17, 447–454.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events,* London: Methuen.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference,* London: Oliver and Boyd.

Gersch, W. (1972), "Causality or Driving in Electrophysiological Signal Analysis," *Mathematical Biosciences,* 14, 177–196.

Gilchrist, W. G. (1967), "Methods of Estimation Involving Discounting," *Journal of the Royal Statistical Society,* Ser. B, 29, 355–369.

Hastie, T., and Tibshirani, R. (1984), *Generalized Additive Models,* Technical Report 2, Laboratory for Computational Statistics, Stanford University.

Koch, C., and Segev, I. (1989), *Methods in Neuronal Modelling,* Cambridge, MA: MIT Press.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression,* Reading, MA: Addison-Wesley.

Rosenberg, J. R., Amjad, A. M., Breeze, P., Brillinger, D. R., and Haliday, D. M. (1989), "The Fourier Approach to the Identification of Functional Coupling Between Neuronal Spike Trains," *Progress in Biophysics and Molecular Biology,* 53, 1–31.

Segundo, J. P. (1968), "Functional Possibilities for Communication and Coding of Neuronal Properties and Interactions," *Acta Neurologica Latinoamericana,* 14, 340–344.

——— (1984), *La Neurofisiologia: Alguno Supuestos y Bases. Recovecos e Implicaciones,* Mexico City: SECEP.

——— (1986), "What Can Neurons Do to Serve as Integrating Devices?" *Journal of Theoretical Neurobiology,* 5, 1–59.

Segundo, J. P., Altshuler, E., Stiber, M., and Garfinkel, A. (1991), "Periodic Inhibition of Living Pacemaker Neurons: I. Locked, Intermittent, Messy, and Hopping Behaviors," *International Journal of Bifurcation and Chaos,* 1, 549–581.

Staniswalis, J. G. (1989), "The Kernel Estimate of a Regression Function in Likelihood-Based Models," *Journal of the American Statistical Association,* 84, 276–283.

Stein, R. B. (1972), "The Stochastic Properties of Spike Trains Recorded From Nerve Cells," in *Stochastic Point Processes,* ed. P. A. W. Lewis, New York: John Wiley, pp. 700–731.

Tibshirani, R., and Hastie, T. (1987), "Local Likelihood Estimation," *Journal of the American Statistical Association,* 82, 559–567.

Tick, L. J. (1963), "Conditional Spectra, Linear Systems, and Coherency," in *Time Series Analysis,* ed. M. Rosenblatt, New York: John Wiley, pp. 197–203.

Villa, A. E. P. (1988), *Influence de l'Ecorce Cérebrale sur l'Activite Spontanée et Evoquée du Thalamus Auditif du Chat,* unpublished thesis, University of Lausanne, Faculty of Sciences.

——— (1990), "Physiological Differentiation Within the Auditory Part of the Thalamic Reticular Nucleus of the Cat," *Brain Research Reviews,* 15, 25–40.

Wold, H. (1956), "Causal Inference From Observational Data: A Review of Ends and Means," *Journal of the Royal Statistical Society,* Ser. A, 119, 28–61.

A GENERALIZED LINEAR MODEL WITH "GAUSSIAN" REGRESSOR VARIABLES

David R. Brillinger
University of California, Berkeley

ABSTRACT

A model in which the conditional expected value of a response variate is an unknown nonlinear function of an unknown linear combination of regressor variates is considered. It is shown that in the case that the regressors are stochastic and jointly Gaussian, or are deterministic and quasi-Gaussian, the ordinary least squares estimates provide useful estimates of the coefficients of the linear combination up to an arbitrary multiplier. The cases of both conditional and unconditional inference are investigated.

KEY WORDS: *Gaussian regressors, Generalized linear model, Multiple regression, Quasi-Gaussian regressors.*

1. INTRODUCTION

Multiple regression is one of the most powerful of statistical techniques. The procedure has been given numerous justifications and interpretations. The traditional approach to it rests on a linear model

$$y_j = \alpha + \beta x_j + \varepsilon_j \ , \qquad (1.1)$$

with the y_j, x_j, $j=1,\ldots,n$ observed, with α, β unknown parameters of interest, with the ε_j zero mean error variates, with the x_j p column-vectors, and with β a p row-vector.

Prepared with the support of National Science Foundation Grant PFR-7901642. I would like to thank Paul Feigin for a number of helpful comments.

P. Guttorp and D. Brillinger (eds.), *Selected Works of David Brillinger*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1344-8_34, © Springer Science+Business Media, LLC 2012

Letting

$$\bar{x} = \sum_{j=1}^{n} x_j/n \ , \tag{1.2}$$

the ordinary least squares estimate, $\hat{\beta}$, satisfies

$$\hat{\beta} \sum_j (x_j - \bar{x})(x_j - \bar{x})^{\tau} = \sum_j y_j(x_j - \bar{x})^{\tau} \tag{1.3}$$

with τ denoting the operation of matrix transposition. In some circumstances the entries of β have causal interpretations, though these must be exercised cautiously (see Box, 1966 and Mosteller and Tukey, 1977, Chapter 13). It seems that substantive scientists have gotten more service out of ordinary least squares estimates than the narrow assumptions of the traditional approach might lead one to suspect possible. In many of these situations it is not the actual value of the coefficients that is of interest, rather it is their relative values, which are somehow measuring the relative importance of the regressor variates of interest. In this paper it is demonstrated that, in the case where the regressors are jointly "Gaussian," the ordinary least squares estimates have a working interpretation for a broader class of models then one might have imagined. The solution, $\hat{\beta}$, of (1.3) is shown to provide an estimate of β in the model

$$y_j = g(\alpha + \beta x_j) + \varepsilon_j \ , \tag{1.4}$$

up to an unknown constant of proportionality. The practical implication is that if the regressors are chosen to be Gaussian, or happen to be approximately so, then despite the possible presence of an unknown nonlinearity, $\hat{\beta}$ still reflects the relative importance of the regressor variates.

After computing $\hat{\beta}$, one may go on to prepare a scatter plot of the points $(\hat{\beta}x_j, y_j)$, $j=1,\ldots,n$ and look for a functional form for $g(\cdot)$. Alternatively, one might compute a nonparametric estimate of $g(u)$ by smoothing the y_j values with $\hat{\beta}x_j$ near u.

It is the usual statistical practice to examine the sampling properties of the least squares estimate conditional on the x values that come to hand. Both the unconditional and conditional distributions are investigated in the paper. Interesting questions arise in the present context, because the fact that the xs are Gaussian is an integral part of the study. It will be seen that it is not convenient to construct confidence regions conditional on a realization of a Gaussian sequence; however, useful regions may be constructed if $x_1, x_2, \ldots$ is a deterministic quasi-Gaussian sequence of a particular sort.

The paper further investigates the extent to which the results require an assumption of normality and describes an application of the results to an identification problem in neurophysiology and an estimation problem in economics.

2. AN ELEMENTARY LEMMA

The whole basis of the procedure is the following simple result given in Brillinger (1977).

Lemma 1. *Let* (U,V) *be bivariate normal with* U *nondegenerate. Let* $g(\cdot)$ *be a measurable function with* $E\{|g(U)|\}$ *and* $E\{|g(U)U|\} < \infty$. *Then*

$$\operatorname{cov}\{g(U),V\} = \operatorname{cov}\{U,V\} \operatorname{cov}\{g(U),U\}/\operatorname{var} U \, . \qquad (2.1)$$

Proof. One has $E\{V|U\} = \mu + \Theta U$ with $\Theta = \operatorname{cov}\{U,V\}/\operatorname{var} U$. Now

$$\operatorname{cov}\{g(U),V\} = \operatorname{cov}\{g(U),E\{V|U\}\} = \Theta \operatorname{cov}\{g(U),U\} \, ,$$

giving the result.

That the regression of V on U is linear is key to the result. It is perhaps worth noting that for $g(\cdot)$, an almost differentiable function (defined in Stein, 1981) satisfying $E\{|g'(U)|\} < \infty$, one may write

$$\operatorname{cov}\{g(U),U\}/\operatorname{var} U = E\{g'(U)\} \qquad (2.2)$$

This last is an identity that Stein (1981) makes use of in his construction of improved estimates of the mean of a multivariate Gaussian.

Now consider the model (1.4) with x_j Gaussian of covariance matrix Σ and ε_j independent of x_j. Then, from (2.1),

$$\text{cov}\{y_j, x_j\} = \beta\Sigma \,\text{cov}\{g(U), U\}/\text{var } U \qquad (2.3)$$

with $U = \alpha + \beta x_j$. (Here $\text{cov}\{y,x\} = E\{(y - \mu_y)(x - \mu_x)^\tau\}$.) The linear regression coefficient of y on x is proportional to β of expression (1.4). Provided $\text{cov}\{g(U),U\} \neq 0$, the constant of proportionality will not be 0. If consistent estimates of $\text{cov}\{y,x\}$ and Σ are constructed, then a consistent estimate of β (up to an arbitrary multiplier) may be constructed. The details of the estimate are presented in the next section for the unconditional case.

3. UNCONDITIONAL INFERENCE

The estimate of interest is the ordinary least squares estimate defined by (1.3). Its properties will be investigated when the variates are related by $y_j = g(\alpha + \beta x_j) + \varepsilon_j$ and when the x_j are Gaussian.

Assumption I. $x_1, x_2, \ldots$ *are independent normals with mean* μ_x *and nonsingular covariance matrix* Σ. $\varepsilon_1, \varepsilon_2, \ldots$ *are independent of the* xs *and have finite variance* σ^2. $E\{x_j^\tau x_j |g(\alpha + \beta x_j)|^2\} < \infty$ *for* $j = 1,2,\ldots$.

From expressions (1.3) and (2.3) one can see that, almost surely, the ordinary least squares estimate $\hat{\beta}$ tends to $\text{cov}\{y,x\}\Sigma^{-1} = k\beta$, where

$$k = \text{cov}\{g(\alpha + \beta x), \alpha + \beta x\}/\text{var}\{\alpha + \beta x\} \ . \qquad (3.1)$$

That is, $\hat{\beta}$ is a strongly consistent estimate of β, up to a constant, k, of proportionality. For $\hat{\beta}$ to be useful, one needs $k \neq 0$.

Turning to the question of the asymptotic distribution of $\hat{\beta}$, set

$$h(x) = g(\alpha + \beta x) - \gamma - \delta x \tag{3.2}$$

where $\delta = k\beta$ and $\gamma = E\{g(\alpha + \beta x) - \delta x\}$. Then one has

Theorem 1. Suppose Assumption I is satisfied. Let $y_j = g(\alpha + \beta x_j) + \varepsilon_j$, $j = 1,2,\ldots$. *Let* $\hat{\beta}$ *be given by (1.3) and* k *by (3.1). Then* $\sqrt{n}(\hat{\beta} - k\beta)$ *is asymptotically normal with mean* **0** *and covariance matrix*

$$\sigma^2\Sigma^{-1} + \Sigma^{-1}E\{h(x)^2(x - \mu_x)(x - \mu_x)^\tau\}\Sigma^{-1} . \tag{3.3}$$

This theorem may be demonstrated using a result of Freedman (1981). The proof is presented in the Appendix. In the case that $g(\cdot)$ is a linear function, the second term in (3.3) will be absent and one has the usual expression for the asymptotic covariance matrix of a least squares estimate.

For the estimate $\hat{\beta}$ to be of practical use, one needs some estimate of its covariance matrix. Several general methods are available for obtaining the latter: the delta method, the jackknife, and the bootstrap. $\hat{\beta}$ is a function of U-statistics, hence the use of the jackknife estimate of the covariance matrix is justified by the results of Arvesen (1969). With a further assumption of $E\{|g(\alpha + \beta x)|^4\} < \infty$, the use of the bootstrap estimate is justified by the results of Freedman (1981). The delta method estimate will now be constructed.

Write expression (1.3) as

$$[\hat{\mu}\ \hat{\beta}]\ \sum_j \begin{bmatrix} 1 \\ x_j \end{bmatrix} [1\ x_j^\tau]\ \frac{1}{n} = \sum_j y_j [1\ x_j^\tau]\ \frac{1}{n} \tag{3.4}$$

or $[\hat{\mu}\ \hat{\beta}]A = B$. Here A and B are means of (matrix-valued) sample values. As A and B are means, the variances and covariances of all their entries may be estimated directly, by the usual expressions. Now if A_0, B_0 denote the expected values of A and B respectively, then one has the perturbation expansion

$$[\hat{\mu}\ \hat{\beta}] = BA^{-1} = B_0A_0^{-1} + (B - B_0)A_0^{-1} - B_0A_0^{-1}(A - A_0)A_0^{-1} + \dots . \tag{3.5}$$

This gives $\hat{\beta}$ as an (approximate) linear function of A and B, whose covariance matrix may now be estimated using the estimates of the variances and covariances of the entries of A and B and replacing A_0, B_0 by A, B respectively.

Having an approximation to the large sample distribution of $\hat{\beta}$ and an estimate of its covariance matrix, one can go on to construct approximate confidence intervals, test hypotheses, and the like.

A concern with these results, however, is that they are unconditional — averaging over all realizations of the xs. Yet in practice, $x_1,\dots,x_n$ will usually be ancillary and one would like to carry out inference conditional on its value at hand. The next section considers this issue.

4. CONDITIONAL INFERENCE

Let $X = \{x_1, x_2,\dots\}$ denote the sequence of regressor variables. This section is concerned with inferences conditional on X. To begin, consider the case where $x_1, x_2,\dots$ are independent realizations of a p-variate normal with mean μ_x and covariance matrix Σ. Directly from expression (1.3) one has

$$E\{\hat{\beta}|X\} = \sum_j g(\alpha+\beta x_j)(x_j - \bar{x})^\tau \ [\sum_j (x_j - \bar{x})(x_j - \bar{x})^\tau]^{-1} = \mu_n(X) \tag{4.1}$$

$$\mathrm{Var}\{\hat{\beta}|X\} = \sigma^2[\sum_j (x_j - \bar{x})(x_j - \bar{x})^\tau]^{-1} = \sigma^2 S_n(X) . \tag{4.2}$$

The variance is the usual least squares expression. In the case that $g(\cdot)$ is linear, the conditional expected value is β; however it will generally be different from β or $k\beta$. A question of interest is how close may it be expected to be to $k\beta$?

The asymptotic conditional distribution of $\hat{\beta}$ is normal. Specifically one has:

Theorem 2. *Suppose Assumption I is satisfied. Then almost surely*

$$\text{Prob}\{\sqrt{n}(\hat{\beta} - \mu_n(X))S_n(X)^{-1/2}/\sigma \leq b|X\} \to \Phi(b_1)\ldots\Phi(b_p) \quad (4.3)$$

as $n \to \infty$, *where* $b = (b_1,\ldots,b_p)$ *and* $\Phi(\cdot)$ *is the standard normal cumulative.*

This result provides information concerning the deviations of $\hat{\beta}$ from $\mu_n(X)$ for a given X. However, one is interested in the deviations of $\hat{\beta}$ from $k\beta$. The next lemma indicates that while $\mu_n(X) - k\beta = o_{a.s.}(1)$, it is not generally $o_{a.s.}(1/\sqrt{n})$ and so (4.3) is not of great use in conditional inference questions concerning β. Theorem 2 is proved in the Appendix of the paper.

Lemma 2. *Let* $x_1, x_2 \ldots$ *be independent normals with mean* μ_x *and nonsingular covariance matrix* Σ. *Suppose* $E\{x_j^\tau x_j|g(\alpha+\beta x_j)|^2\} < \infty$, $j = 1,2,\ldots$. *Then*

$$E\{\hat{\beta}|X\} = \mu_n(X) = k\beta + \frac{1}{\sqrt{n}}W + o_{a.s.}(\frac{1}{\sqrt{n}}) , \quad (4.4)$$

where k *is given by (3.1) and* W *is normal with mean* 0 *and covariance matrix*

$$\Sigma^{-1}E\{h(x)^2(x - \mu_x)(x - \mu_x)^\tau\}\Sigma^{-1} , \quad (4.5)$$

$h(x)$ *being given by (3.2).*

The deviations of $\mu_n(X)$ from $k\beta$ are seen to be of order $1/\sqrt{n}$, generally. One implication of this is that the result (4.3) cannot be used to construct approximate confidence regions for $k\beta$. Some other approach is needed. The lemma is proved in the Appendix.

As the lemma and discussion make clear, for a typical realization of the Gaussian process X, $\mu_n(X)$ does not tend to $k\beta$ rapidly enough to be useful. Consider expression (4.1). The term

$$\sum_j g(\alpha + \beta x_j)x_j/n$$

may be considered an approximation to the integral, or expected value, $E\{g(\alpha + \beta x)x\}$. This suggests that by choosing a sequence $x_1, x_2, \ldots$ corresponding to a clever numerical integration rule, one might be able to have $E\{\hat{\beta}|X\}$ closer to $k\beta$ than $O_{a.s.}(1/n)$. This does turn out to be possible.

Halton (1960) has demonstrated the existence of a sequence of points $u_1, u_2, \ldots$ in the unit cube $[0,1]^p$ with the property that

$$D_n = \sup_{I \varepsilon J} \left| \frac{\#(u_1,\ldots,u_n \ \varepsilon \ I)}{n} - \mu(I) \right| = O(n^{-1}(\log n)^p) \qquad (4.6)$$

where J is the family of all subintervals of $[0,1]^p$ and where $\mu(I)$ is the Lebesgue measure of I. (A computer algorithm for generating the sequence is given in Halton and Smith, 1964.) The usefulness of this sequence is that for a function, f, with variation, $V(f)$, in the sense of Hardy and Krause (see Neiderreiter, 1978, p. 967), one has

$$\left| \frac{1}{n} \sum_{j=1}^{n} f(u_j) - \int f(u)du \right| \leq V(f)D_n = O(n^{-1}(\log n)^p) \qquad (4.7)$$

for bounded $V(f)$. The sequence $u_1, u_2, \ldots$ may be said to be quasi-uniform. Writing $u_j = (u_{j1},\ldots,u_{jp})$ and $x_j = (x_{j1},\ldots,x_{jp})$ with $x_{jk} = \Phi^{-1}(u_{jk})$, the sequence $x_1, x_2, \ldots$ may be said to be quasi-Gaussian. Letting $h(x_j) = f(u_j)$, one has from (4.7),

$$\left|\frac{1}{n} \sum_{j=1}^{n} h(x_j) - \int h(x)\phi(x_1)\ldots\phi(x_p)dx \right| = O(n^{-1}(\log n)^p) \qquad (4.8)$$

for $h(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_p))$ of bounded variation with $\phi(\cdot)$ denoting the standard normal density. One might say that for $p > 1$, quasi-Monte Carlo techniques exist that outperform naive Monte Carlo.

Returning to the question of the estimation of $\hat{\beta}$ of the model (1.4), suppose now that the values of the regressors may be chosen by the experimenter. Suppose he takes $x_1, x_2, \ldots$ to be the above quasi-Gaussian sequence. Consider $\hat{\beta}$ satisfying

$$\hat{\beta} \sum_j x_j x_j^\tau = \sum_j y_j x_j^\tau \, . \tag{4.9}$$

(There is no need to correct for the mean with this sequence.) One has

$$\begin{aligned} E\{\sum_j y_j x_j^\tau/n &= \sum_j g(\alpha + \beta x_j)x_j^\tau/n \\ &= \int g(\alpha + \beta x)x^\tau \phi(x_1)\ldots\phi(x_p)dx + O(n^{-1}(\log n)^p) \\ &= k\beta + O(n^{-1}(\log n)^p) \end{aligned}$$

from (4.8), provided g is of bounded variation as required. Similarly,

$$\sum_j x_j x_j^\tau/n = I + O(n^{-1}(\log n)^p) \, .$$

In summary, for the deterministic quasi-Gaussian sequence indicated above, one has

$$E\,\hat{\beta} = k\beta + O(n^{-1}(\log n)^p) \, . \tag{4.10}$$

The variance-covariance matrix of β is $\sigma^2[\sum_j x_j x_j^\tau]^{-1}$, and hence the conclusion of Theorem 2 becomes

$$\text{Prob}\{(\hat{\beta} - k\beta)[\sum_j x_j x_j^\tau]^{1/2}/\sigma \leq b\} \to \Phi(b_1)\ldots\Phi(b_p) \, . \tag{4.11}$$

Once an estimate of σ is at hand, approximate confidence regions for $k\beta$ may be constructed using (4.11).

With an estimate of $g(\cdot)$, σ^2 may be estimated from the residuals of the fit. Various nonparametric estimates of a regression function are available. A bibliographic review of these is given in Collomb (1981). In the present context one might form

$$\hat{g}(u) = \sum_{j=1}^n y_j W_n(u - \hat{\beta}x_j) / \sum_{j=1}^n W_n(u - \hat{\beta}x_j) \tag{4.12}$$

for example, with W_n a sequence of weight functions becoming concentrated at 0 as n increases. For large n, $\hat{g}(u)$ may be expected to be near $g(\alpha + u/k)$. The error variance may be estimated by

$$\hat{\sigma}^2 = \sum_{j=1}^{n} [y_j - \hat{g}(\hat{\beta}x_j)]^2/n \ . \tag{4.13}$$

A procedure for constructing approximate confidence regions for $k\beta$ has been set down.

5. DISCUSSION

Section 3 discussed inference in the unconditionsl case when $x_1, x_2, \ldots$ was any realization of a sequence of independent p-variate normals. Section 4 developed inference for the case that $x_1, x_2, \ldots$ was a very particular deterministic sequence (that was quasi-Gaussian). It would appear that the latter conditional inference procedure is the preferred one — as is the case in the usual (linear) regression situation — since $x_1, \ldots, x_n$ is generally an ancillary statistic. Lehmann (1981) comments on some aspects of ancillaries and conditional inference.

If the form of the function $g(\cdot)$ is known, then one will be able to determine other estimates of β, for example, the maximum likelihood. These other estimates may be expected to be more efficient. There have been at least two studies in which the ordinary least squares estimate has been compared with the maximum likelihood estimate. In both cases it has been found to perform well, even when the xs were not Gaussian.

Greene (1981) considered the model

$$y_j = \max\{0, \alpha + \beta x_j + \varepsilon_j\} \tag{5.1}$$

with the εs independent normals of mean 0 and variance σ^2. He derived both the ordinary least squares and the maximum likelihood estimate of β for a set of data from a study of female labor supply. Here y was the number of hours worked in a survey year. The xs are listed in Table 1. (Eight of them are dummy variables.) The estimate, $\hat{\beta}$, has been standardized to $\hat{\beta}\hat{\beta}^\tau = 1$. The proportion of nontruncated observations was .460.

There is close agreement between the results of least squares and maximum likelihood. This occurs despite some of the xs having far from normal distributions. Greene (1981) is able to construct an estimate for of (3.1), since $g(\cdot)$ is known, and so obtain an estimate of β itself.

Table 1

Variable	Maximum Likelihood	Least Squares
x_1 = small child	-.4140	-.3831
x_2 = health	-.5072	-.4472
x_3 = other income	.0005	.0008
x_4 = wage	.5156	.6053
x_5 = south	.2953	.2989
x_6 = farm	-.2266	-.2218
x_7 = urban	.0554	.0523
x_8 = age	.0097	.0094
x_9 = education	.0113	.0125
x_{10} = rel. wage	.1438	.1346
x_{11} = 2nd marriage	.0127	.0143
x_{12} = mean divorce prob.	.2416	.2381
x_{13} = high divorce prob.	.2906	.2652

Brillinger and Segundo (1979) present an example of a successful application of the estimation procedure discussed in this paper, in a more complicated situation. A neuron was stimulated by a fluctuating current, causing it to fire every so often. The stimulating current was taken to be stationary Gaussian. In the classic model of neuron firing, the input current $X(t)$ is filtered to form the membrane potential

$$U(t) = \int_0^{B(t)} a(u)X(t-u)du$$

with $B(t)$ the time elapsed at time t since the neuron last fired. The neuron then fires next when $U(t)$ crosses an approximately constant threshold. It is of interest to estimate the function $a(\cdot)$ of (5.1) and to confirm the presence of a threshold.

A time series analog of the procedure considered in this paper was applied to experimental data consisting of a record of the current taken as input and the times at which the neuron fired. Strictly speaking, the model is not appropriate here because of correlation introduced by $B(t)$ being present in (5.1). A maximum likelihood procedure was developed to deal with this difficulty. It was found that the results of the procedure of this paper were quite consistent with the maximum likelihood results. In the principal experiment, the input current was taken to be Gaussian. In a second experiment, the input current was taken to have a uniform distribution. Figure 1 gives the time series analog of the regression estimate of $a(\cdot)$ when $X(t)$ is Gaussian. Figure 2 gives it for $X(t)$ uniform. The two estimates are surprisingly close, suggesting that the procedure may be robust.

In a part of the study analagous to the estimation of the function $g(\cdot)$, the nonlinearity was estimated and found to have a threshold character. Sampling fluctuations of the estimates were estimated by splitting the data up into a number of segments and estimating the parameters separately for each segment, rather than attempting to use any of the procedures of Section 3.

6. A PARTIAL CONVERSE

The development of the results of this paper made essential use of an assumption of normality for the xs. A question of some interest is whether there is any other distribution leading to similar results. The following theorem indicates that normality is required for regressor variates of one important type.

Theorem 3. Let the p-*variate* x *be of the form* $a + b\varepsilon$ *with* a *a* p-*vector,* b $p \times p$, *and nonsingular, and the entries of* ε *independent, identically distributed of mean* 0, *and finite nonzero variance. Let* Σ *denote the covariance matrix of* x. *Suppose* $p > 1$ *and*

$$\operatorname{cov}\{g(\beta x), x\} = k\beta\Sigma \tag{6.1}$$

for some $\beta \neq 0$, *some* $k \neq 0$, *and all* $g(u)$ *of the form* $\exp\{itu\}$, t *real-valued. Then,* x *is normally distributed.*

This theorem is proved in the Appendix. This result is far from a converse; however, it does suggest strongly that normal regressors will prove the most useful.

7. CONCLUDING REMARKS

So far, the work of this paper has been predicated on the assumption (1.4) of a model with an additive error. When the xs were Gaussian and independent of the error, this model led to the relationship $\operatorname{cov}\{y,x\} = k\beta \operatorname{var} x$, on which the estimation procedure proposed was based. In fact, this relationship follows from the weaker assumption that

$$E\{y|x\} = g(\alpha + \beta x) . \tag{7.1}$$

The estimation procedure is now seen to be of use in a broader class of situations. Consider, for example, the binomial response (or regression) model. Here $y = 1$ or 0 with

$$\operatorname{Prob}\{y = 1|x\} = g(\alpha + \beta x) \tag{7.2}$$

with $g(\cdot)$ normal for the probit model and logistic for the logit model. From what has gone before in the paper, one sees that if $g(\cdot)$ is unknown and x is Gaussian, then β may be estimated, up to a constant of proportionality, by ordinary least squares. As a second example, consider the Cox (1972) model of proportional hazards. This involves a random variate y (a survival time), and associated covariates x, with

$$\text{Prob}\{y \leq t|x\} = 1 - [1 - F_0(t)]^{\exp\{\beta x\}} ,$$

$F_0(\cdot)$ being an unknown c·d·f. (This class of models is sometimes referred to as the class of Lehmann alternatives, introduced in Lehmann, 1953.) It is clear that, when the expected value exists, $E\{y|x\} = g(\beta x)$, for some $g(\cdot)$. If x is taken to be Gaussian and associated ys recorded, then the procedure of this paper allows the estimation of $k\beta$. As a final example, one has linear regression with a censored dependent variate; for example,

$$\begin{aligned} y &= \alpha + \beta x + \varepsilon \quad \text{if the right-hand side is nonnegative} \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Such models are discussed in Green (1981), Nelson (1981), and references given therein. It is clear that $E\{y|x\} = g(\alpha + \beta x)$ and that ordinary least squares estimates are of use in the Gaussian case once again.

APPENDIX

Proof of Theorem 1. By writing expression (1.3) in the form (3.4), without the 1/n's one has $[\hat{\mu}\ \hat{\beta}]$ of the form of the statistic $\beta(n)$ considered on page 1219 of Freedman (1981). His result gives the asymptotic normality of $\hat{\beta}$. His expression for the asymptotic covariance matrix may be manipulated to give (3.3).

Proof of Theorem 2. Expand the equations (1.3) to the form (3.4) once again, that is, to the form of the usual normal equations of multiple regression. If $\hat{\theta} = [\hat{\mu}\ \hat{\beta}]$ and $\theta_X = E\{\hat{\theta}|X\}$, this gives

$$(\hat{\theta} - \theta_X) \sum_j \begin{bmatrix} 1 \\ x_j \end{bmatrix} [1\ x_j^T] = \sum_j \varepsilon_j [1\ x_j^T] .$$

This corresponds to standard multiple regression with the regression coefficient 0. That $\hat{\theta} - \theta_X$ is asymptotically normal under the stated conditions of the xs and εs is shown in Miller (1974).

Proof of Lemma 2. Taking $\sigma = 0$ in Theorem 1, it follows that $\sqrt{n}(\mu_n(X) - k\beta)$ is asymptotically normal with mean 0 and covariance matrix (4.5). That $\mu_n(X)$ then has the representation (4.4) follows from a theorem of Skorokhod (1956) (see also Wichura, 1970).

Proof of Theorem 3. One can assume $a = E\{x\} = 0$. Then, from (6.1),

$$E\{g(\beta x)x^\tau\} = kE\{\beta x x^\tau\} . \quad (A.1)$$

Set $L_p = \beta x$ and $L_j = \gamma_j x$ with the γ_j chosen so that $L_1, \ldots, L_p$ are mutually uncorrelated. Multiplying (A.1) by γ_j^τ, one has

$$E\{g(\beta x)L_j\} = kE\{L_p L_j\} = 0$$

and so $E\{g(L_p)L_j\} = 0$. From Lemma 1.1.1 of Kagan et al (1973), this last gives $E\{L_j | L_p\} = 0$. That x is necessarily normal now follows from Theorem 5.5.3 of Kagan et al or Theorem 2 of Cacoullos (1967).

REFERENCES

ARVESEN, J.N. (1969), "Jackknifing U-Statistics," *Annals of Mathematical Statistics*, 40, 2076-2100.

BOX, G.E.P. (1966), "Use and Abuse of Regression," *Technometrics*, 8, 625-629.

BRILLINGER, D.R. (1977), "The Identification of a Particular Nonlinear Time Series System," *Biometrika*, 64, 509-515.

______________, and SEGUNDO, J.P. (1979), "Empirical Examination of the Threshold Model of Neuron Firing," *Biological Cybernetics*, 35, 213-220.

CACOULLOS, T. (1967), "Characterizations of Normality by Constant Regression of Linear Statistics on Another Linear Statistic," *Annals of Mathematical Statitistics*, 38, 1894-1898.

COLLOMB, G. (1981), "Estimation Non-Parametrique de la Regression: Revue Bibliographique," *International Statistical Review*, 49, 75-94.

COX, D.R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, Ser. B, 34, 187-220.

FREEDMAN, D.A. (1981), "Bootstrapping Regression Models," *Annals of Statistics*, 9, 1218-1228.

GREENE, W.H. (1981), "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," *Econometrica*, 49, 505-513.

HALTON, J.H. (1960), "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals," *Numerische Mathematik*, 2, 84-90.

__________ and SMITH, G.B. (1964), "Algorithm 247: Radical-Inverse Quasi-Random Point Sequence |G5|," *Communications Association of Computing Machinery*, 7, 701-702.

KAGAN, A.M., LINNIK, Yu.V., and RAO, C.R. (1973), *Characterization Problems in Mathematical Statistics*, New York: Wiley.

LEHMANN, E.L. (1953), "The Power of Rank Tests," *Annals of Mathematical Statistics*, 24, 23-43.

__________ (1981), "An Interpretation of Completeness and Basu's Theorem," *Journal of the American Statistical Association*, 76, 335-340.

MILLER, R.G. (1974), "An Unbalanced Jackknife," *Annals of Mathematical Statistics*, 2, 880-891.

MOSTELLER, F.M., and TUKEY, J.W. (1977), *Data Analysis and Regression*, Reading, Ma.: Addison-Wesley.

NELSON, F.D. (1981), "A Test for Misspecification in the Censored Normal Model," *Econometrica*, 49, 1317-1330.

NIEDERREITER, H. (1978), "Quasi-Monte Carlo Methods and Pseudo-Random Numbers," *Bulletin American Mathematical Society*, 84, 957-1041.

SKOROKHOD, A.V. (1956), "Limit Theorems for Stochastic Processes," *Theory of Probability and Its Applications*, 1, 261-290.

STEIN, C.M. (1981), "Estimation of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 9, 1135-1151.

WICHURA, M. (1970), "On the Construction of Almost Uniformly Convergent Random Variables With Given Weakly Convergent Image Laws," *Annals of Mathematical Statistics*, 41, 284-291.

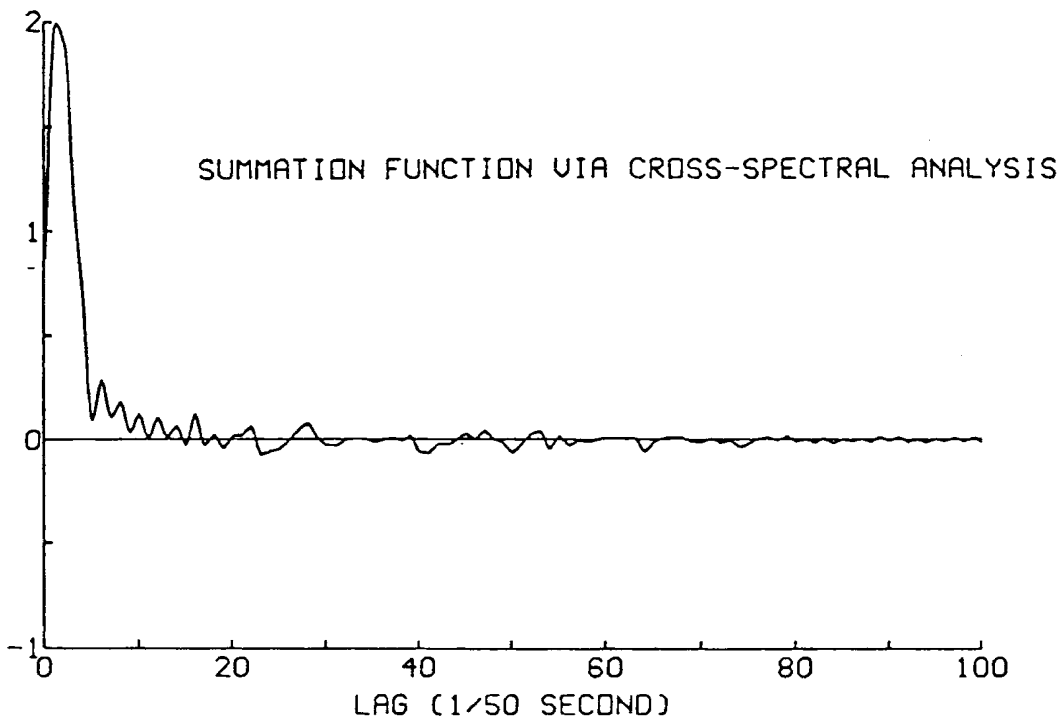


Figure 1. Estimate of the summation function a(•) obtained with Gaussian input.

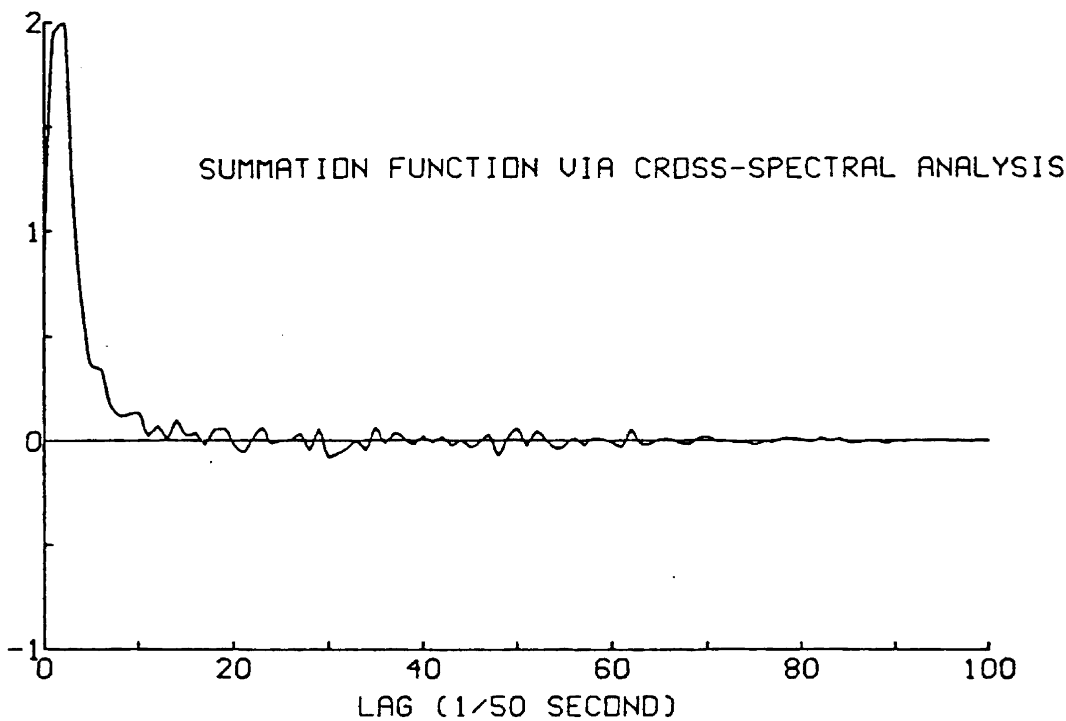


Figure 2. Estimate of the summation function a(•) obtained with uniform input.

Biometrika (1977), **64**, 3, *pp.* 509–15
Printed in Great Britain

The identification of a particular nonlinear time series system

BY DAVID R. BRILLINGER
Department of Statistics, University of California, Berkeley

SUMMARY

A nonlinear time series system is considered. The system has the property that the output series corresponding to a given input series is the sum of a noise series and the result of applying in turn the operations of linear filtering, instantaneous functional composition and linear filtering to the input series. Given a stretch of Gaussian input series and corresponding output series, estimates are constructed of the transfer functions of the linear filters, up to constant multipliers. The investigation discloses that for such a system, the best linear predictor of the output given Gaussian input, has a broader interpretation than might be suspected. The result is derived from a simple expression for the covariance function of a normal variate with a function of a jointly normal variate.

Some key words: Cumulant; Gaussian series; Multiple regression; Nonlinear system; Normal variate; Stationary time series; System identification.

1. INTRODUCTION

Let (U, V) be a bivariate normal with U nondegenerate. Then one can write $V = \beta U + \epsilon$ with ϵ independent of U and $\beta = \operatorname{cov}(U, V)/\operatorname{var}(U)$. Let $G(u)$ $(-\infty < u < \infty)$ be a real-valued function. Then one has

$$\operatorname{cov}\{G(U), V\} = \operatorname{cov}\{G(U), \beta U + \epsilon\} = \operatorname{cov}\{G(U), U\} \operatorname{cov}(U, V)/\operatorname{var}(U), \tag{1·1}$$

provided that $E\{|G(U)|\}$ and $E\{|G(U)\,U|\} < \infty$. Suppose next that

$$Y = \mu + G\left(\sum_{j=1}^{J} \alpha_j X_j\right) + \epsilon, \tag{1·2}$$

with $(X_1, \ldots, X_J)$ multivariate normal, with ϵ a zero mean, finite second-order moment variate independent of $(X_1, \ldots, X_J)$ and with $\mu, \alpha_1, \ldots, \alpha_J$ constants. It follows from (1·1) that

$$\operatorname{cov}(Y, X_k) = \sum_{j=1}^{J} \alpha_j \operatorname{cov}(X_j, X_k) \operatorname{cov}\{G(U), U\}/\operatorname{var}(U) \quad (k = 1, \ldots, J) \tag{1·3}$$

with $U = \Sigma \alpha_j X_j$. Now, in linear regression theory the regression coefficients of the variate Y on $(X_1, \ldots, X_J)$ are determined by solving the system of equations

$$\operatorname{cov}(Y, X_k) = \sum_{j=1}^{J} \alpha_j \operatorname{cov}(X_j, X_k). \tag{1·4}$$

It follows from (1·3) that when the nonlinear model (1·2) applies, the linear regression coefficients of Y on $(X_1, \ldots, X_J)$ are proportional to the α's of (1·2). Provided $\operatorname{cov}\{G(U), U\} \neq 0$, the constant of proportionality will not be zero. One implication of this result is that when the independent variates of a regression analysis are jointly normal, estimates of the linear regression coefficients are relevant to the linear parameters of a broader class of models than might have been suspected.

When data $(Y_i, X_{i1}, \ldots, X_{iJ})$ $(i = 1, \ldots, n)$ are available and estimates $\hat{\alpha}_1, \ldots, \hat{\alpha}_J$ are determined by solving the usual normal equations, the form of the function $G(u)$ may be examined by plotting the points $(\Sigma \hat{\alpha}_j X_{ij}, Y_i)$ $(i = 1, \ldots, n)$. The unknown constant of proportionality may here be considered part of the function $G(u)$.

Consider next the model

$$Y = \mu + \sum_{j=1}^{J} \alpha_j G(X_j) + \epsilon, \tag{1·5}$$

with the previous assumptions and in addition the assumption that the X_j have identical distributions. It follows from (1·1) that

$$\operatorname{cov}(Y, X_k) = \sum_{j=1}^{J} \alpha_j \operatorname{cov}(X_j, X_k) \operatorname{cov}\{G(X), X\}/\operatorname{var}(X), \tag{1·6}$$

where X denotes a variate with the distribution of the X_j. Once again the linear regression coefficients are seen to be proportional to those of a broader model, when the independent variates are normal.

Expressions (1·3) and (1·6) follow directly from (1·1). This expression was derived under a normality assumption; however, an examination of the argument shows that the key requirement is that

$$\operatorname{cov}\{G(U), V - \beta U\} = 0 \tag{1·7}$$

for β, the linear regression coefficient, $\operatorname{cov}(U, V)/\operatorname{var}(U)$. The relationship (1·7) may be expected to hold, approximately, for a broad class of variates U, V and functions $G(u)$.

In § 2 the relationship (1·1) is generalized to cumulants and functions of several variables. In § 3 it is applied to certain time invariant nonlinear time series models. In § 4, the asymptotic distributions of certain estimates of the parameters of the time series models are investigated.

2. A particular cumulant

Let $\operatorname{cum}(U_1, \ldots, U_J)$ denote the joint cumulant of order J of the variates $U_1, \ldots, U_J$. This functional has the properties of vanishing if some subset of the U's is statistically independent of the remainder and of being multilinear in its arguments (Brillinger, 1975, § 2·3). These two properties lead to the following lemma.

Lemma 2·1. *Let $(U_1, \ldots, U_J, V_1, \ldots, V_K)$ be multivariate normal with $U_1, \ldots, U_J$ nondegenerate and statistically independent of each other. Set $\sigma_{jk} = \operatorname{cov}(U_j, V_k)$. Let $G(u_1, \ldots, u_J)$ be a measurable function of $(u_1, \ldots, u_J)$ satisfying*

$$E\{|G(U_1, \ldots, U_J) Z_1 \ldots Z_L|\} < \infty \tag{2·1}$$

for $Z_1, \ldots, Z_L$ any subset of $(U_1, \ldots, U_J, V_1, \ldots, V_K)$. Then, for $J, K \geqslant 1$,

$$\operatorname{cum}\{G(U_1, \ldots, U_J), V_1, \ldots, V_K\} = \sum_{j=1}^{J} \operatorname{cum}\{G(U_1, \ldots, U_J), U_j, \ldots, U_j\} \sigma_{j1} \ldots \sigma_{jK}/\{\operatorname{var}(U_j)\}^K. \tag{2·2}$$

Proof. The cumulant of (2·2) exists in view of (2·1). Applying the Gram–Schmidt orthogonalization procedure to the variates $U_1, \ldots, U_J, V_1, \ldots, V_K$ one sees that one can write

$$U_j = \alpha_j \epsilon_j \quad (j = 1, \ldots, J)$$

and

$$V_k = \sum_{l=1}^{J+k} \beta_{kl} \epsilon_l \quad (k = 1, \ldots, K),$$

with the ϵ's independent standard normal variates. Hence the cumulant on the left-hand side of (2·2) is given by

$$\sum_{j=1}^{J} \operatorname{cum}\{G(\alpha_1 \epsilon_1, \ldots, \alpha_J \epsilon_J), \beta_{1j}\epsilon_j, \ldots, \beta_{Kj}\epsilon_j\} = \sum_{j=1}^{J} \operatorname{cum}\{G(U_1, \ldots, U_J), U_j, \ldots, U_j\}\beta_{1j} \ldots \beta_{Kj}/\alpha_j^K.$$

The result (2·1) now follows because $\operatorname{cov}(U_j, V_k) = \alpha_j \beta_{kj}$.

Expression (1·1) corresponds to the result of this lemma with J, $K = 1$. The case of dependent $U_1, \ldots, U_J$ may be handled similarly, for one may write $G(U_1, \ldots, U_J) = H(W_1, \ldots, W_J)$ with $W_1, \ldots, W_J$ independent and obtained from $U_1, \ldots, U_J$ by a Gram–Schmidt orthogonalization procedure.

3. A time series system

A time series system is a collection of a space of input series, a space of output series, and an operation carrying an input series into an output series. Suppose $X(t)$ $(t = 0, \pm 1, \ldots)$ denotes an input series and $Y(t)$ $(t = 0, \pm 1, \ldots)$ the corresponding output series. Then a common time series system has the form

$$Y(t) = \mu + \sum_{u=-\infty}^{\infty} a(t-u)\, X(u) + \epsilon(t) \quad (t = 0, \pm 1, \ldots), \tag{3·1}$$

for some sequence of filter coefficients $a(u)$ $(u = 0, \pm 1, \ldots)$, for some constant μ, and for some zero mean noise series $\epsilon(t)$ $(t = 0, \pm 1, \ldots)$. The problem of system identification is that of determining characteristics of the system from corresponding stretches of input and output series, say, $\{X(t), Y(t)\}$ $(t = 0, \ldots, T-1)$.

Suppose that the series $X(.)$ and $\epsilon(.)$ are stationary and independent. Let

$$c_{XY}(u) = \operatorname{cov}\{X(t+u), Y(t)\}$$

denote the crosscovariance function of the two series and let the autocovariance functions $c_{XX}(u)$ and $c_{YY}(u)$ be defined similarly. Then the system (3·1) leads to the relationship

$$c_{XY}(u) = \Sigma_v\, a(v)\, c_{XX}(v+u) \tag{3·2}$$

for suitable $a(.)$. Let

$$f_{XY}(\lambda) = (2\pi)^{-1} \Sigma_u\, c_{XY}(u)\, e^{-i\lambda u} \quad (-\infty < \lambda < \infty), \tag{3·3}$$

denote the cross-spectrum of the series $X(.)$ with the series $Y(.)$ and make corresponding definitions of the power spectra $f_{XX}(\lambda), f_{YY}(\lambda)$. Let

$$A(\lambda) = \Sigma_u\, a(u)\, e^{-i\lambda u} \tag{3·4}$$

denote the transfer function of the filter. Then the relationship (3·2) leads to

$$f_{XY}(\lambda) = A(-\lambda) f_{XX}(\lambda),$$

or, if $f_{XX}(\lambda) \neq 0$, to $A(\lambda) = f_{YX}(\lambda)\{f_{XX}(\lambda)\}^{-1}$. The parameter $f_{YX}(\lambda)\{f_{XX}(\lambda)\}^{-1}$ is called the complex regression coefficient of the series $Y(.)$ on the series $X(.)$ at frequency λ. It provides the transfer function of the best linear filter for predicting the series $Y(.)$ from the series $X(.)$ (Brillinger, 1975, § 8·3).

The remainder of this paper is concerned with the identification of the following system, generalizing (3·1), for $t = 0, \pm 1, \ldots$,

$$\begin{gathered} U(t) = \Sigma_u\, a(t-u)\, X(u), \quad V(t) = G\{U(t)\}, \\ Y(t) = \mu + \Sigma_u\, b(t-u)\, V(u) + \epsilon(t), \end{gathered} \tag{3·5}$$

where $G(.)$ is a function from reals to reals, and where $\epsilon(.)$ is a stationary noise series independent of the stationary series $X(.)$. The system (3·1) corresponds to $G(u) = u$, the identity transformation, and to $b(0) = 1$, $b(u) = 0$ for $u \neq 0$, the identity filter. Suppose now that the series $X(.)$ is stationary Gaussian; then Lemma 2·1 gives the relationship

$$c_{XY}(u) = L_1 \Sigma_w \Sigma_v a(w) b(v) c_{XX}(w+v+u), \tag{3·6}$$

where $L_1 = \text{cov}\{U(0), V(0)\}/\text{var}\{U(0)\}$, provided, for example, that all of

$$\Sigma_u|c_{XX}(u)|, \quad \Sigma_u|a(u)|, \quad \Sigma_u|b(u)|, \quad E[|G\{U(t)\}|], \quad E[|U(t+u) G\{U(t)\}|] \tag{3·7}$$

are bounded. Formation of the Fourier transform of the relationship (3·6) gives

$$f_{XY}(\lambda) = L_1 A(-\lambda) B(-\lambda) f_{XX}(\lambda). \tag{3·8}$$

That is, for the system (3·5), the complex regression coefficient $f_{YX}(\lambda)\{f_{XX}(\lambda)\}^{-1}$ is proportional to the product $A(\lambda)B(\lambda)$ of the transfer functions of the two linear filters of the system. It is interesting that in this situation, the best linear predictor corresponds to the composition of the two linear components of the system. It may be remarked that for the relationship (3·8) to be meaningful it should be the case that $L_1 \neq 0$. An expression equivalent to (3·6) was set down by Korenberg (1973) for the case of $X(.)$ a continuous time white noise process, $G(u)$ a polynomial in u and $\epsilon(.)$ identically zero.

When one of the filters $a(.)$ and $b(.)$ is the identity, expression (3·8) indicates that the transfer function of the other is proportional to the complex regression coefficient $f_{YX}(\lambda)\{f_{XX}(\lambda)\}^{-1}$. The transfer function may be estimated, up to a constant, once estimates of the second-order spectra have been constructed. An indeterminate constant is to be expected since no restrictions have been placed on the function $G(u)$. The character of the function $G(u)$ may be examined by using the estimate of the unknown filter to determine approximately the series $U(.)$, or $V(.)$, and by plotting the values $\{U(t), V(t)\}$ $(t = 0, 1, \ldots, T-1)$.

The relationship (3·8) is not sufficient to construct individual estimates of $A(\lambda)$ and $B(\lambda)$. Further relationships involving these parameters exist, however. First, some further parameters must be defined. Given a trivariate stationary series $\{X(t), Y(t), Z(t)\}$ $(t = 0, \pm 1, \ldots)$, let

$$\begin{aligned} c_{XYZ}(u,v) &= E([X(t+u) - E\{X(t+u)\}][Y(t+v) - E\{Y(t+v)\}][Z(t) - E\{Z(t)\}]) \\ &= \text{cum}\{X(t+u), Y(t+v), Z(t)\}, \end{aligned}$$

for $t, u, v = 0, \pm 1, \ldots$, denote the third-order cumulant function of the series and

$$f_{XYZ}(\lambda, \nu) = (2\pi)^{-2} \Sigma_u \Sigma_v c_{XYZ}(u,v) \exp\{-i(\lambda u + v\nu)\} \quad (-\infty < \lambda, \nu < \infty),$$

the corresponding third-order cumulant spectrum. Now Lemma 2·1 leads to the relationship

$$c_{XXY}(u,v) = L_2 \Sigma_w \Sigma_x \Sigma_y a(x) a(y) c_{XX}(x+u+w) c_{XX}(y+v+w) b(w), \tag{3·9}$$

where $L_2 = \text{cum}\{U(0), U(0), V(0)\}/[\text{var}\{U(0)\}]^2$. An expression equivalent to this was given by Korenberg (1973) when $X(.)$ is a continuous time white noise process, $G(u)$ a polynomial in u and $\epsilon(.)$ identically zero. The relationship will be valid provided in addition to the conditions of (3·7) one has $E[|U(t+u) U(t+v) G\{U(t)\}|] < \infty$. Formation of the Fourier transform of the relationship (3·9) gives

$$f_{XXY}(\lambda, \nu) = L_2 A(-\lambda) A(-\nu) B(-\lambda-\nu) f_{XX}(\lambda) f_{XX}(\nu). \tag{3·10}$$

Notice that (3·10) is not symmetric in $A(.)$ and $B(.)$. It may not be hoped to identify the filters $a(.)$ and $b(.)$ completely, even with this added relationship, for it is clear that with the model (3·5) when the filter $a(.)$ is translated τ time units to form the filter $a(.+\tau)$, and the filter $b(.)$ is translated to $b(.-\tau)$, then the output series is unaffected. It does follow from (3·8) and (3·9) though, that, provided $L_1, f_{XY}(\lambda+\nu), f_{XX}(\lambda), f_{XX}(\nu) \neq 0$, then

$$\frac{f_{XXY}(\lambda,\nu) f_{XX}(\lambda+\nu)}{f_{XY}(\lambda+\nu) f_{XX}(\lambda) f_{XX}(\nu)} = \frac{L_2 A(-\lambda) A(-\nu)}{L_1 A(-\lambda-\nu)}, \tag{3·11}$$

and by setting $\nu = -\lambda$

$$\{L_2 |A(\lambda)|^2\}/\{L_1 A(0)\} = \{f_{XXY}(\lambda, -\lambda) f_{XX}(0)\}/\{f_{XY}(0) f_{XX}(\lambda)^2\}.$$

Therefore, up to a constant multiplier, the gain of the filter, $|A(\lambda)|$, is given by

$$|f_{XXY}(\lambda, -\lambda)|^{\frac{1}{2}}/f_{XX}(\lambda). \tag{3·12}$$

It should be remarked that for this result to be meaningful, it is necessary that $L_1, L_2 \neq 0$.

For the problem of identifying $A(\lambda)$ beyond its modulus, let $\phi(\lambda) = \arg\{A(\lambda)\}$ and let $\psi(\lambda,\nu) = \arg\{f_{XXY}(\lambda,\nu)/f_{XY}(\lambda+\nu)\}$. Then from (3·11), $\phi(\lambda+\nu)-\phi(\lambda)-\phi(\nu) = \psi(\lambda,\nu) \pmod{\pi}$. It may be checked by simple substitution that the formula

$$\phi(\lambda) = \left\{2\int_0^\lambda \phi(\alpha)\,d\alpha + \int_0^\lambda \psi(\alpha, \lambda-\alpha)\,d\alpha\right\}\Big/\lambda \tag{3·13}$$

provides a recursive means of obtaining $\phi(\lambda)$ given $\phi(\alpha)$ $(0 \leqslant \alpha < \lambda)$. The formula (3·13) provides a determination of the phase with $\phi'(0) = 0$. In the next section estimates of the filters $A(.)$ and $B(.)$ will be constructed using the relationships (3·12) and (3·14). In some circumstances, the context of the problem may suggest that the filter $a(.)$ is realizable, that is $a(u) = 0$ for $u < 0$, and one may be willing to assume further that the filter is of minimum phase type. Then the phase $\phi(\lambda)$ may be determined from the amplitude $|A(.)|$; see Solodovnikov (1960, § 13) or Robinson (1962, Chapter VII).

The particular cases of the system when the filter $a(.)$ or the filter $b(.)$ is the identity were mentioned earlier. These possibilities may be investigated to some extent. If the filter $b(.)$ is the identity, then from (3·8) and (3·10)

$$\frac{f_{XXY}(\lambda,\nu) f_{XX}(\lambda) f_{XX}(\nu)}{f_{XX}(\lambda) f_{XX}(\nu) f_{XY}(\lambda) f_{XY}(\nu)} = \frac{f_{XXY}(\lambda,\nu)}{f_{XY}(\lambda) f_{XY}(\nu)} \tag{3·14}$$

will be constant as a function of λ and ν. On the other hand, if the filter $a(.)$ is the identity then from (3·11)

$$\{f_{XXY}(\lambda,\nu) f_{XX}(\lambda+\nu)\}/\{f_{XY}(\lambda+\nu) f_{XX}(\lambda) f_{XX}(\nu)\} \tag{3·15}$$

will be constant as a function of λ and ν.

In the general case, from (3·11) and (3·12),

$$\{|f_{XXY}(\lambda,\nu)|\,|f_{XXY}(\lambda+\nu, -\lambda-\nu)|^{\frac{1}{2}}\}/\{|f_{XY}(\lambda+\nu)|\,|f_{XXY}(\lambda, -\lambda) f_{XXY}(\nu, -\nu)|^{\frac{1}{2}}\} \tag{3·16}$$

will be constant as a function of λ, ν. It is apparent that estimates of the expressions (3·14), (3·15) and (3·16) may be formed to examine the plausibilities of the respective models. The next section considers the construction of such estimates.

4. Estimation theory

Brillinger (1975, Chapter 8) constructed consistent and asymptotically normal estimates, based on data $\{X(t), Y(t)\}$ $(t = 0, 1, \ldots, T-1)$ for $f_{YX}(\lambda)\{f_{XX}(\lambda)\}^{-1}$, equal to $L_1 A(\lambda) B(\lambda)$ here, and of its inverse Fourier transform, equal to $L_1\, a * b(u)$ here, when the series $\{X(.), Y(.)\}$ is stationary with

$$\Sigma_{u_1} \dots \Sigma_{u_J} (1+|u_1|+\dots+|u_J|)\,[|\mathrm{cum}\{X(t+u_1),\dots,X(t+u_J),X(t)\}| + |\mathrm{cum}\{Y(t+u_1),\dots,Y(t+u_J),Y(t)\}|] < \infty,$$

$$\Sigma_{u_1} \dots \Sigma_{u_J} \Sigma_{v_1} \dots \Sigma_{v_K} (1+|u_1|+\dots+|u_J|+|v_1|+\dots+|v_K|)\,[|\mathrm{cum}\{X(t+u_1),\dots,X(t+u_J), Y(t+v_1),\dots,Y(t+v_K),Y(t)\}|] < \infty, \quad (4\cdot 1)$$

for $J, K = 1, 2, \dots$. In the present situation the series $X(.)$ is Gaussian and so its cumulant functions of order greater than two vanish. In connexion with it therefore it is only necessary to assume $\Sigma(1+|u|)\,|c_{XX}(u)| < \infty$. Suppose further that the noise series $\epsilon(.)$ satisfies

$$\Sigma_{u_1} \dots \Sigma_{u_J} (1+|u_1|+\dots+|u_J|)\,|\mathrm{cum}\{\epsilon(t+u_1),\dots,\epsilon(t+u_J),\epsilon(t)\}| < \infty, \quad (4\cdot 2)$$

for $J = 1, 2, \dots$, and that the function $G(.)$ has the property $E[|G\{U(t)\}|^k] < \infty$ $(k = 1, 2, \dots)$. Then the method employed in the proof of Lemma 1 of Brillinger (1968), invoking Kibble's (1945) extension of Mehler's Theorem, may be used to show that the condition on $G(.)$ and (4·2) imply (4·1).

The preceding discussion is relevant to the case in which one of $a(.)$, $b(.)$ and $G(.)$ is the identity. In the general case, from expression (3·12) it is seen that $|A(\lambda)|$ may be estimated, up to a constant multiplier, by

$$|f^T_{XXY}(\lambda, -\lambda)|^{\frac{1}{2}}/f^T_{XX}(\lambda), \quad (4\cdot 3)$$

where $f^T_{XXY}(\lambda,\nu)$ is an estimate of the third-order cumulant spectrum $f_{XXY}(\lambda,\nu)$. Suppose that $f^T_{XXY}(\lambda,\nu)$ is constructed as by Brillinger & Rosenblatt (1967) by smoothing the third-order periodogram using the weight function $S^T(\alpha,\beta)$. Then Theorems 1 and 4 of Brillinger & Rosenblatt (1967) may be used to show that the estimate (4·3) is consistent and asymptotically normal with variance

$$K_\lambda\, T^{-1}\tfrac{1}{2}\pi \iint S^T(\alpha,\beta)^2\, d\alpha\, d\beta\, f_{XX}(0)/|f_{XXY}(\lambda,-\lambda)|,$$

where $K_\lambda = 1$ if $\lambda \neq 0, \pi$ and $K_0 = 6$, $K_\pi = 2$.

As an estimate of $\phi(\lambda) = \arg\{A(\lambda)\}$, expression (3·13) suggests the consideration of

$$\phi^T(2\pi s/S) = \{2 \sum_{1<j<s} \phi^T(2\pi j/S) + \psi^T(2\pi j/S, 2\pi(s-j)/S)\}\, S\,(2\pi s)^{-1},$$

for S a large integer, $s = 2, 3, \dots$, with $\phi^T(2\pi/S) = 0$ and

$$\psi^T(\alpha, \lambda-\alpha) = \arg\{f^T_{XXY}(\alpha, \lambda-\alpha)/f^T_{XY}(\lambda)\}.$$

As an estimate of $A(\lambda)$, up to a constant multiplier, now take

$$A^T(\lambda) = \exp\{i\phi^T(\lambda)\}\,|f^T_{XXY}(\lambda,-\lambda)|^{\frac{1}{2}}/f^T_{XX}(\lambda).$$

As an estimate of $|B(\lambda)|$, up to a constant multiplier, take $|f^T_{YX}(\lambda)|/|f^T_{XXY}(\lambda,-\lambda)|^{\frac{1}{2}}$ and as an estimate of $\arg\{B(\lambda)\}$ (mod π) take $\arg\{f^T_{YX}(\lambda)\} - \phi^T(\lambda)$.

5. Concluding remarks

The results of the previous section indicate that the proposed identification procedure will be most satisfying when one of the filters $a(.)$ and $b(.)$ is the identity. The results also indicate that the procedures of cross-spectral analysis have a much broader domain of applicability than might be expected wherever the input series is Gaussian.

For the particular case $G(u) = \alpha_1 u + \alpha_2 u^2$ of a quadratic system, there already exist results concerning identification by means of Gaussian input (Tick, 1961; Brillinger, 1970). It may be checked readily that those results agree with the present ones.

This research was carried out at the University of Auckland. It was partially supported by the J. S. Guggenheim Memorial Foundation and a National Science Foundation Grant.

References

BRILLINGER, D. R. (1968). Estimation of the cross-spectrum of a stationary bivariate Gaussian process from its zeros. *J. R. Statist. Soc.* B **30**, 145–59.

BRILLINGER, D. R. (1970). The identification of polynomial systems by means of higher order spectra. *J. Sound Vib.* **12**, 301–13.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

BRILLINGER, D. R. & ROSENBLATT, M. (1967). Asymptotic theory of estimates of k-th order spectra. In *Spectral Analysis of Time Series*, Ed. B. Harris, pp. 158–88. New York: Wiley.

KIBBLE, W. F. (1945). An extension of a theorem of Mehler's on Hermite polynomials. *Proc. Camb. Phil. Soc.* **41**, 12–15.

KORENBERG, M. J. (1973). Cross-correlation analysis of neural cascades. In *Proc. 10th Annual Rocky Mountain Bioeng. Symp.*, pp. 47–51. Pittsburgh: Instrument Society of America.

ROBINSON, E. A. (1962). *Random Wavelets and Cybernetic Systems*. London: Griffin.

SOLODOVNIKOV, V. V. (1960). *Introduction to the Statistical Dynamics of Automatic Control Systems*. New York: Dover.

TICK, L. J. (1961). The estimation of "transfer functions" of quadratic systems. *Technometrics* **3**, 563–7.

[*Received November* 1976. *Revised February* 1977]