Contributions to Statistics

Joachim Kunert
Christine H. Müller
Anthony C. Atkinson *Editors*

mODa 11 - Advances in Model-Oriented Design and Analysis

Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis held in Hamminkeln, Germany, June 12-17, 2016

Springer

Contributions to Statistics

More information about this series at http://www.springer.com/series/2912

Joachim Kunert • Christine H. Müller •
Anthony C. Atkinson
Editors

mODa 11 - Advances in Model-Oriented Design and Analysis

Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis held in Hamminkeln, Germany, June 12-17, 2016

Springer

Editors
Joachim Kunert
Department of Statistics
Technische Universität Dortmund
Dortmund, Germany

Christine H. Müller
Department of Statistics
Technische Universität Dortmund
Dortmund, Germany

Anthony C. Atkinson
Department of Statistics
London School of Economics
London, United Kingdom

ISSN 1431-1968
Contributions to Statistics
ISBN 978-3-319-31264-4 ISBN 978-3-319-31266-8 (eBook)
DOI 10.1007/978-3-319-31266-8

Library of Congress Control Number: 2016940826

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This volume contains articles based on presentations at the 11th workshop on model-oriented data analysis and optimum design (mODa) in Hamminkeln-Dingden, Germany, during June 2016. The 11th workshop was organized by the Department of Statistics of the TU Dortmund and supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

The mODa series of workshops focuses on nonstandard design of experiments and related analysis of data. The main objectives are:

- To promote new advanced research areas as well as collaboration between academia and industry.
- Whenever possible, to provide financial support for research in the area of experimental design and related topics.
- To give junior researchers the opportunity of establishing personal contacts and working together with leading researchers.
- To bring together scientists from different statistical schools – particular emphasis is given to the inclusion of scientists from Central and Eastern Europe.

The mODa series of workshops started at the Wartburg near Eisenach in the former GDR in 1987 and has continued as a tri-annual series of conferences. The locations and dates of the former conferences are as follows:

- mODa 1: Eisenach, former GDR, 1987,
- mODa 2: St. Kyrik, Bulgaria, 1990,
- mODa 3: Peterhof, Russia, 1992,
- mODa 4: Spetses, Greece, 1995,
- mODa 5: Luminy, France, 1998,
- mODa 6: Puchberg/Schneeberg, Austria, 2001,
- mODa 7: Heeze, The Netherlands, 2004,
- mODa 8: Almagro, Spain, 2007,
- mODa 9: Bertinoro, Italy, 2010,
- mODa 10: Łagów Lubuski, Poland, 2013.

The articles in this volume provide an overview of current topics in research on experimental design. The topics covered by the papers are:

- designs for treatment combinations (Atkinson; Druilhet; Grömping and Bailey),
- randomisation (Bailey; Ghiglietti; Shao and Rosenberger),
- computer experiments (Curtis and Maruri-Aguilar; Ginsbourger, Baccou, Chevalier and Perales),
- designs for nonlinear regression and generalized linear models (Amo-Salas, Jiménez-Alcázar and López-Fidalgo; Burclová and Pázman; Cheng, Majumdar and Yang; Mielke; Radloff and Schwabe),
- designs for dependent data (Deldossi, Osmetti and Tommasi; Gauthier and Pronzato; Prus and Schwabe),
- designs for functional data (Aletti, May and Tommasi; Zang and Großmann),
- adaptive and sequential designs (Borrotti and Pievatolo; Hainy, Drovandi and McGree; Knapp; Lane, Wang and Flournoy),
- designs for special fields of application (Bischoff; Fedorov and Xue; Graßhoff, Holling and Schwabe; Pepelyshev, Staroselskiy and Zhigljavsky),
- foundations of experimental design (Müller and Wynn; Zhigljavsky, Golyandina and Gillard).

In this time of Big Data, it is often not emphasized in public discourse that experimental design remains extremely important. The mODa series of workshops wishes to raise public awareness of the continuing importance of experimental design. In particular, the papers from various fields of application show that experimental design is not a mathematical plaything, but is of direct use in the sciences.

Since the first workshop in Eisenach, optimal design for various situations has been at the heart of the research covered by mODa. Sequential design is another long-standing topic in the mODa series. It is clear that computer experiments, designs for dependent data, and functional data become increasingly feasible. For causal inference in particular, old-fashioned methods like randomization, blinding, and orthogonality of factors remain indispensable. In addition to the importance of the research covered here, we think that the articles in this volume show the beauty of mathematical statistics, which should not be forgotten.

For the editors, it was a pleasure reading these research results. We would like to thank the authors for submitting such nice work and for providing revisions in time, wherever a revision was necessary. Last, but not least, we want to thank the referees who provided thoughtful and constructive reviews in time, helping to make this volume a fine addition to any statistician's bookshelves.

Dortmund, Germany

Christine Müller
Joachim Kunert
Anthony Atkinson

Contents

List of Contributors

Giacomo Aletti Università degli Studi di Milano, Milano, Italy

Mariano Amo-Salas Faculty of Medicine, University of Castilla-La Mancha, Camino de Moledores, Ciudad Real, Spain

Anthony C. Atkinson Department of Statistics, London School of Economics, London, UK

Jean Baccou Institut de Radioprotection et de Sûreté Nucléaire, PSN-RES, SEMIA, Centre de Cadarache, France

Laboratoire de Micromécanique et d'Intégrité des Structures, IRSN-CNRS-UMII, Saint-Paul-lès-Durance, France

R.A. Bailey School of Mathematics and Statistics, University of St Andrews, St Andrews, UK

Wolfgang Bischoff Mathematisch-Geographische Fakultaet, Katholische Universitaet Eichstaett-Ingolstadt, Eichstaett, Germany

Matteo Borrotti CNR-IMATI, Milan, Italy

Katarína Burclová Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovak Republic

Qianshun Cheng Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, USA

Clément Chevalier Institut de Statistique, Université de Neuchâtel, Neuchâtel, Switzerland

Institute of Mathematics, University of Zurich, Zurich, Switzerland

Peter Curtis Queen Mary, University of London, London, UK

Laura Deldossi Dipartimento di Scienze statistiche, Università Cattolica del Sacro Cuore, Milan, Italy

Christopher C. Drovandi School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

Pierre Druilhet Laboratoire de Mathématiques, UMR 6620 – CNRS, Université Blaise Pascal, Clermont-Ferrand, France

Valerii V. Fedorov ICONplc, North Wales, PA, USA

Nancy Flournoy University of Missouri, Columbia, MO, USA

Bertrand Gauthier ESAT-STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium

Andrea Ghiglietti Università degli Studi di Milano, Milan, Italy

Jonathan Gillard Cardiff University, Cardiff, UK

David Ginsbourger Centre du Parc, Idiap Research Institute, Martigny, Switzerland

Department of Mathematics and Statistics, IMSV, University of Bern, Bern, Switzerland

Nina Golyandina St.Petersburg State University, St.Petersburg, Russia

Ulrike Graßhoff School of Business and Economics, Humboldt University, Berlin, Germany

Ulrike Grömping Department II, Beuth University of Applied Sciences Berlin, Berlin, Germany

Heiko Großmann Otto-von-Guericke-University, Magdeburg, Germany

Markus Hainy Department of Applied Statistics, Johannes Kepler University, Linz, Austria

Heinz Holling Institute of Psychology, University of Münster, Münster, Germany

Alfonso Jiménez-Alcázar Environmental Sciences Institute, University of Castilla-La Mancha, Toledo, Spain

Guido Knapp Department of Statistics, TU Dortmund University, Dortmund, Germany

Adam Lane Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA

Jesús López-Fidalgo Higher Technical School of Industrial Engineering, University of Castilla-La Mancha, Ciudad Real, Spain

Dibyen Majumdar Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, USA

Hugo Maruri-Aguilar Queen Mary, University of London, London, UK

Caterina May University of Eastern Piedmont, Novara, Italy

James M. McGree School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

Tobias Mielke ICON Clinical Research, Cologne, Germany

Werner G. Müller Department of Applied Statistics, Johannes Kepler University, Linz, Austria

Silvia Angela Osmetti Dipartimento di Scienze statistiche, Università Cattolica del Sacro Cuore, Milan, Italy

Andrej Pázman Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovak Republic

Andrey Pepelyshev Cardiff University, Cardiff, UK

Frédéric Perales Institut de Radioprotection et de Sûreté Nucléaire, PSN-RES, SEMIA, Centre de Cadarache, France

Laboratoire de Micromécanique et d'Intégrité des Structures, IRSN-CNRS-UMII, Saint-Paul-lès-Durance, France

Antonio Pievatolo CNR-IMATI, Milan, Italy

Luc Pronzato Laboratoire I3S – UMR 7271, CNRS, Université de Nice-Sophia Antipolis/CNRS, Nice, France

Maryna Prus Institute for Mathematical Stochastics, Otto-von-Guericke University, Magdeburg, Germany

Martin Radloff Institute for Mathematical Stochastics, Otto-von-Guericke-University, Magdeburg, Germany

William F. Rosenberger Department of Statistics, George Mason University, Fairfax, VA, USA

Rainer Schwabe Institute for Mathematical Stochastics, Otto-von-Guericke University, Magdeburg, Germany

Hui Shao Department of Statistics, George Mason University, Fairfax, VA, USA

Yuri Staroselskiy Crimtan, 1 Castle Lane, London, UK

Chiara Tommasi Dipartimento DEMM, Università degli Studi di Milano, Milan, Italy

HaiYing Wang University of New Hampshire, Durham, NH, USA

Henry P. Wynn Department of Statistics, London School of Economics, London, UK

Xiaoqiang Xue Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Min Yang Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, USA

Bairu Zhang Queen Mary University of London, London, UK

Anatoly Zhigljavsky Cardiff University, Cardiff, UK

Lobachevskii Nizhnii Novgorod State University, Nizhny Novgorod, Russia

On Applying Optimal Design of Experiments when Functional Observations Occur

Giacomo Aletti, Caterina May, and Chiara Tommasi

Abstract In this work we study the theory of optimal design of experiments when functional observations occur. We provide the best estimate for the functional coefficient in a linear model with functional response and multivariate predictor, exploiting fully the information provided by both functions and derivatives. We define different optimality criteria for the estimate of a functional coefficient. Then, we provide a strong theoretical foundation to prove that the computation of these optimal designs, in the case of linear models, is the same as in the classical theory, but a different interpretation needs to be given.

1 Introduction

In many statistical contexts data have a functional nature, since they are realizations from some continuous process. For this reason functional data analysis is an interest of many researchers. Reference monographs on problems and methods for functional data analysis are, for instance, the books of [6, 12] and [7].

Even in the experimental context functional observations can occur in several situations. In the literature many authors have already dealt with optimal design for experiments with functional data (see, for instance, [1, 3, 9, 10, 13, 14, 16]). Sometimes the link between the infinite-dimensional space and the finite-dimensional projection is not fully justified and may unknowingly cause errors. In this work we offer a theoretical foundation to obtain the best estimates of the functional coefficients and the optimal designs in the proper infinite-dimensional space, and its finite-dimensional projection which is used in practice.

When dealing with functional data, derivatives may provide important additional information. In this paper we focus on a linear model with functional response and multivariate (or univariate) predictor. In order to estimate the functional coefficient,

G. Aletti (✉) • C. Tommasi
Università degli Studi di Milano, Milano, Italy
e-mail: giacomo.aletti@unimi.it; chiara.tommasi@unimi.it

C. May
University of Eastern Piedmont, Novara, Italy
e-mail: caterina.may@uniupo.it

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_1

we exploit fully the information provided by both functions and derivatives, obtaining a strong version of the Gauss-Markov theorem in the Sobolev space H^1.

Since our goal is precise estimation of the functional coefficients, we define some optimality criteria to reach this aim. We prove that the computation of the optimal designs can be obtained as in the classical case, but the meaning of the of A- and D- criteria cannot be traced back any more to the confidence ellipsoid. Hence we give the right interpretation of the optimal designs in the functional context.

2 Model Description

We consider a linear regression model where the response y is a random function which depends on a vector (or scalar) known variable $\mathbf{x}$ through a functional coefficient, which needs to be estimated. Whenever n experiments can be performed the model can be written in the following form, for $t \in \tau$,

$$y_i(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) + \varepsilon_i(t) \qquad i = 1, \dots, n, \tag{1}$$

where $y_i(t)$ denote the response curve for the i-th value of the regressor $\mathbf{x}_i$; $\mathbf{f}(\mathbf{x}_i)$ is a p-dimensional vector of known functions; $\boldsymbol{\beta}(t)$ is an unknown p-dimensional functional vector; $\varepsilon_{ij}(t)$ is a zero-mean error process. This model is a functional response model described, for instance, in [7].

In a real world setting, the functions $y_i(t)$ are not directly observed. By a smoothing procedure from the original data, the investigator can reconstruct both the functions and their derivatives, obtaining $y_i^{(f)}(t)$ and $y_i^{(d)}(t)$, respectively. Hence we can assume that the model for the reconstructed functional data is

$$\begin{cases} y_i^{(f)}(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t) + \varepsilon_i^{(f)}(t) \\ y_i^{(d)}(t) = \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t) + \varepsilon_i^{(d)}(t) \end{cases} \qquad i = 1, \dots, n, \tag{2}$$

where all the n couples $\{\varepsilon_i^{(f)}(t), \varepsilon_i^{(d)}(t)\}$ are zero-mean identically distributed processes, each process being independent of all the other processes, with $E(\|\varepsilon_{ij}^{(f)}(t)\|_{L^2}^2 + \|\varepsilon_{ij}^{(d)}(t)\|_{L^2}^2) < \infty$.

Note that the investigator might reconstruct each function $y_i^{(f)}(t)$ and its derivative $y_i^{(d)}(t)$ separately. In this case, the terms of the second equation of (2) are not the derivative of the terms of the first equation. The particular case when $y_i^{(d)}(t)$ is obtained deriving $y_i^{(f)}(t)$ is the most simple situation in model (2) and can be seen as model (1).

Let us consider an estimator $\hat{\boldsymbol{\beta}}(t)$ of $\boldsymbol{\beta}(t)$, formed by p random functions in the Sobolev space $H^1 = H^1(\tau)$. Recall that a function $g(t)$ is in H^1 if $g(t)$ and its derivative function $g'(t)$ belongs to L^2. Moreover, H^1 is a Hilbert space with inner product

$$\begin{aligned}\langle g_1(t), g_2(t)\rangle_{H^1} &= \langle g_1(t), g_2(t)\rangle_{L^2} + \langle g_1'(t), g_2'(t)\rangle_{L^2} \\ &= \int g_1(t)g_2(t)dt + \int g_1'(t)g_2'(t)dt, \qquad g_1(t), g_2(t) \in H^1.\end{aligned}$$

Definition 1 We define the H^1-*generalized covariance matrix* $\Sigma_{\hat{\boldsymbol{\beta}}}$ of $\hat{\boldsymbol{\beta}}(t)$ as the $p \times p$ matrix whose (l_1, l_2)-th element is

$$E\langle \hat{\beta}_{l_1}(t) - \beta_{l_1}(t), \hat{\beta}_{l_2}(t) - \beta_{l_2}(t)\rangle_{H^1}. \tag{3}$$

Definition 2 In analogy with classical settings, we define the H^1-functional best linear unbiased estimator (H^1-BLUE) as the estimator with minimal (in the sense of Loewner Partial Order) H^1-generalized covariance matrix (3), in the class of the linear unbiased estimators of $\boldsymbol{\beta}(t)$.

Given a couple $\{y^{(f)}(t), y^{(d)}(t)\} \in L^2 \times L^2$, a linear continuous operator on H^1 may be defined as follows

$$\phi(h) = \langle y^{(f)}, h\rangle_{L^2} + \langle y^{(d)}, h'\rangle_{L^2}, \qquad \forall h \in H^1.$$

From the Riesz representation theorem, there exists a unique $\tilde{y} \in H^1$ such that

$$\langle \tilde{y}, h\rangle_{H^1} = \langle y^{(f)}, h\rangle_{L^2} + \langle y^{(d)}, h'\rangle_{L^2}, \qquad \forall h \in H^1. \tag{4}$$

Definition 3 We call $\tilde{y} \in H^1$ in (4) the *Riesz representative* of the couple $(y^{(f)}(t), y^{(d)}(t)) \in L^2 \times L^2$.

This definition will be useful to provide a nice expression for the functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$. Actually the Riesz representative synthesizes, in some sense, the information of both $y^{(f)}(t)$ and $y^{(d)}(t)$ in H^1.

The *functional OLS estimator* for the model (2) is

$$\begin{aligned}\hat{\boldsymbol{\beta}}(t) &= \arg\min_{\boldsymbol{\beta}(t)} \Big(\sum_{i=1}^{n} \|y_i^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T\boldsymbol{\beta}(t)\|_{L^2}^2 + \sum_{i=1}^{n} \|y_i^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T\boldsymbol{\beta}'(t)\|_{L^2}^2 \Big) \\ &= \arg\min_{\boldsymbol{\beta}(t)} \sum_{i=1}^{n} \Big(\|y_i^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T\boldsymbol{\beta}(t)\|_{L^2}^2 + \|y_i^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T\boldsymbol{\beta}'(t)\|_{L^2}^2 \Big).\end{aligned}$$

The quantity

$$\|y_i^{(f)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{L^2}^2 + \|y_i^{(d)}(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}'(t)\|_{L^2}^2$$

resembles

$$\|y_i(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{H^1}^2,$$

because $y_i^{(f)}(t)$ and $y_i^{(d)}(t)$ reconstruct $y_i(t)$ and its derivative function, respectively. The functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ minimizes, in this sense, the sum of the H^1-norm of the unobservable residuals $y_i(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)$.

3 Infinite and Finite-Dimensional Results

This section contains the fundamental theoretical results for estimation of functional linear models given in Sect. 2; they can be proved as particular cases of the theorems contained in [2].

Theorem 1 *Given the model in (2),*

(a) the functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ can be computed by

$$\hat{\boldsymbol{\beta}}(t) = (F^T F)^{-1} F^T \tilde{\mathbf{y}}(t), \tag{5}$$

where $\tilde{\mathbf{y}}(t) = \{\tilde{y}_1(t), \ldots, \tilde{y}_n(t)\}$ is the vector whose components are the Riesz representatives of the replications, and $F = [\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n)]^T$ is the $n \times p$ design matrix.

(b) The estimator $\hat{\boldsymbol{\beta}}(t)$ is unbiased and its generalized covariance matrix is

$$\Sigma_{\hat{\boldsymbol{\beta}}} = \sigma^2 (F^T F)^{-1},$$

where $\sigma^2 = E(\|y_i(t) - \mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}(t)\|_{H^1}^2)$.

The functional OLS estimator obtained in Theorem 1 by means of the Riesz representatives is also the best linear unbiased estimator in the Sobolev space, as stated in the next theorem, which is a functional version of the well known Gauss-Markov theorem.

Theorem 2 *The functional OLS estimator $\hat{\boldsymbol{\beta}}(t)$ for the model (2) is a H^1-functional BLUE, when the Riesz representatives of the eigenfunctions of the error terms are independent.*

In a real world context, we work with a finite dimensional subspace $\mathscr{S}$ of H^1. Let $S = \{w_1(t), \ldots, w_N(t)\}$ be a base of $\mathscr{S}$. Without loss of generality, we may assume that $\langle w_h(t), w_k(t) \rangle_{H^1} = \delta_h^k$, where δ_h^k is the Kronecker delta symbol, since a Gram-Schmidt orthonormalization procedure may always be applied. More precisely,

given any base $\tilde{S} = \{\tilde{w}_1(t), \ldots, \tilde{w}_N(t)\}$ in H_1, the corresponding orthonormal base is given by:

for $k = 1$, define $w_1(t) = \frac{\tilde{w}_1(t)}{\|\tilde{w}_1(t)\|_{H^1}}$,

for $k \geq 2$, let $\hat{w}_k(t) = \tilde{w}_k(t) - \sum_{h=1}^{n-1} \langle \tilde{w}_k(t), w_h(t) \rangle_{H^1} w_h(t)$, and $w_k(t) = \frac{\hat{w}_k(t)}{\|\hat{w}_k(t)\|_{H^1}}$.

With this orthonormalized base, the projection $\tilde{y}(t)_{\mathscr{S}}$ on $\mathscr{S}$ of the Riesz representative $\tilde{y}(t)$ of the couple $\{y^{(f)}(t), y^{(d)}(t)\}$ is given by

$$\begin{aligned} \tilde{y}(t)_{\mathscr{S}} &= \sum_{k=1}^{N} \langle \tilde{y}(t), w_k(t) \rangle_{H^1} \cdot w_k(t) \\ &= \sum_{k=1}^{N} \Big(\langle y^{(f)}(t), w_k(t) \rangle_{L^2} + \langle y^{(d)}(t), w_k'(t) \rangle_{L^2} \Big) w_k(t), \end{aligned} \tag{6}$$

where the last equality comes from the definition (4) of the Riesz representative. Now, if $\mathbf{m}_l = (m_{l,1}, \ldots, m_{l,n})^T$ is the l-th row of $(F^T F)^{-1} F^T$, then

$$\begin{aligned} \langle \hat{\boldsymbol{\beta}}_l(t), w_k(t) \rangle_{H^1} &= \sum_{i=1}^{n} \langle m_{l,i} y_i(t), w_k(t) \rangle_{H^1} \\ &= \sum_{i=1}^{n} m_{l,i} \langle y_i(t), w_k(t) \rangle_{H^1}, \qquad \text{for any } k = 1, \ldots, N, \\ \hat{\boldsymbol{\beta}}_l(t)_{\mathscr{S}} &= \mathbf{m}_l^T \mathbf{y}(t)_{\mathscr{S}}, \end{aligned}$$

hence $\hat{\boldsymbol{\beta}}(t)_{\mathscr{S}} = (F^T F)^{-1} F^T \mathbf{y}(t)_{\mathscr{S}}$.

Let us note that, even if the Riesz representative (4) is implicitly defined, its projection on $\mathscr{S}$ can be easily computed by (6). From a practical point of view, the statistician can work with the data $\{y_{ij}^{(f)}(t), y_{ij}^{(d)}(t)\}$ projected on a finite linear subspace $\mathscr{S}$ and the corresponding OLS estimator $\hat{\boldsymbol{\beta}}(t)_{\mathscr{S}}$ is the projection on $\mathscr{S}$ of the obtained H^1-OLS estimator $\hat{\boldsymbol{\beta}}(t)$. As a consequence of Theorem 2, $\hat{\boldsymbol{\beta}}(t)_{\mathscr{S}}$ is also H^1-BLUE in $\mathscr{S}$, since it is unbiased and the projection is linear. For the projection, it is crucial to take a base of $\mathscr{S}$ which is orthonormal in H^1.

It is straightforward to prove that the estimator (5) becomes

$$\hat{\boldsymbol{\beta}}(t) = (F^T F)^{-1} F^T \mathbf{y}^{(f)}(t),$$

in two cases: when we do not take into consideration $y^{(d)}$, or when $y^{(d)} = y'^{(f)}$. In both cases, from the results obtained, $\hat{\boldsymbol{\beta}}$ is an L^2-BLUE. To our knowledge, this is the most common situation considered in the literature.

4 Optimal Designs

Assume we work in an experimental setup. Therefore, $\mathbf{x}_i$, with $i = 1, \dots, n$, are not observed auxiliary variables; they can be freely chosen by an experimenter on the design space $\mathscr{X}$. The set of experimental conditions $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ is called an exact design. A more general definition is that of a continuous design, as a probability measure ξ with support on $\mathscr{X}$ (see, for instance, [8]). The choice of ξ may be made with the aim of obtaining accurate estimates of the model functional parameters.

From Theorem 2, $\hat{\boldsymbol{\beta}}(t)$ given in (5) is the H^1-BLUE for the model (2). This optimal estimator can be further improved by a "clever" choice of the design. By analogy with the criteria proposed in the finite-dimensional theory (see for instance, [4, 11, 15]) we define a *functional optimal design* as a design which minimizes an appropriate convex function of the generalized covariance matrix $\Sigma_{\hat{\boldsymbol{\beta}}}$ given in Definition 1. In particular, we define the following optimality criteria.

Definition 4 A *functional D-optimum* design is a design ξ_D^* which minimizes $\det(\Sigma_{\hat{\boldsymbol{\beta}}})$; a *functional A-optimum* design is a design ξ_A^* which minimizes $trace(\Sigma_{\hat{\boldsymbol{\beta}}})$; a *functional E-optimum* design is a design ξ_E^* which minimizes the maximum eigenvalue of $\Sigma_{\hat{\boldsymbol{\beta}}}$.

Observe that Definition 4 may be applied also in the case of functional non-linear models. When we deal in particular with models (1) or (2), part (b) of Theorem 1 shows that

$$\Sigma_{\hat{\boldsymbol{\beta}}} \propto (F^T F)^{-1},$$

and, from the definition of continuous design,

$$F^T F \propto \int_{\mathscr{X}} \mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^T d\xi(\mathbf{x}).$$

Hence we have proved that, in the case of models (1) and (2), a functional optimal design can be computed as in the classical theory.

4.1 Interpretation

We describe here the meaning of the optimality criteria given by Definition 4 in the functional context. Observe that these interpretations are strongly connected to the definition of generalized covariance matrix given in Definition 1.

4.1.1 Functional D-Optimum Designs

Let $\hat{\boldsymbol{\beta}}(t)$ be an unbiased estimator for a functional parameter $\boldsymbol{\beta}(t)$ having H^1-generalized covariance matrix $\Sigma_{\hat{\boldsymbol{\beta}}}$ according to Definition 1. Then, for $\boldsymbol{\theta}$ in R^p, the equation

$$\boldsymbol{\theta}^T \, \Sigma_{\hat{\boldsymbol{\beta}}} \, \boldsymbol{\theta} \leq \text{constant} \tag{7}$$

defines an ellipsoid of R^p such that the linear combinations

$$\boldsymbol{\theta}^T \, \hat{\boldsymbol{\beta}}(t) = \sum_{i=1}^{p} \theta_i \, \hat{\beta}_i(t), \tag{8}$$

with $\boldsymbol{\theta}$ in the ellipsoid (7), have H^1-generalized variance bounded by the same arbitrary constant:

$$\Sigma_{\boldsymbol{\theta}^T \hat{\boldsymbol{\beta}}} = E(\| \sum_{i=1}^{p} \theta_i \, \hat{\beta}_i(t) - \sum_{i=1}^{p} \theta_i \, \beta_i(t) \|^2_{H^1}) \leq \text{constant}.$$

A functional D-optimum design maximizes the volume of ellipsoid (7) and hence the estimate of $\boldsymbol{\beta}(t)$ is more accurate since the "volume" of linear combinations $\sum_{i=1}^{p} \theta_i \, \hat{\beta}_i(t)$ with bounded variance is greater.

4.1.2 Functional A-Optimum Designs

A functional A-optimum design minimizes the trace of $\Sigma_{\hat{\boldsymbol{\beta}}}$; it can be proved that this is equivalent to minimizing

$$\int_{\|\boldsymbol{\theta}\| \leq 1} \boldsymbol{\theta}^T \, \Sigma_{\hat{\boldsymbol{\beta}}} \, \boldsymbol{\theta} \; d\boldsymbol{\theta}.$$

Observe that $\boldsymbol{\theta}^T \, \Sigma_{\hat{\boldsymbol{\beta}}} \, \boldsymbol{\theta}$ is the H^1-generalized variance of the linear combinations (8). In other words, a functional A-optimum design minimizes the mean H^1-generalized variance of the linear combinations $\sum_{i=1}^{p} \theta_i \, \hat{\beta}_i(t)$ with coefficients on the unit ball $\|\boldsymbol{\theta}\| \leq 1$. We are able to prove that this can be also achieved with coefficients on the unit sphere $\|\boldsymbol{\theta}\| = 1$.

4.1.3 Functional E-Optimum Designs

Finally, the E-optimality criterion has the following interpretation: a functional E-optimum design minimizes the maximum H^1-generalized variance of the linear combinations $\sum_{i=1}^{p} \theta_i \, \hat{\beta}_i(t)$ with the constraint $\|\boldsymbol{\theta}\| \leq 1$ or $\|\boldsymbol{\theta}\| = 1$.

5 Future Developments

The advantages of applying the theory discussed in this paper are shown in [2] in a real example, where a linear model with functional response and vectorial predictor is used for an ergonomic problem, as proposed in [13]. To forecast the motion response of drivers within a car (functional response), different locations are chosen (experimental conditions). The original, non-optimal design adopted provides a D-efficiency equal to 0.3396; this D-efficiency is raised to 0.9779 through a numerical algorithm for optimal designs.

Regression models with functional variables can cover different situations: we can have functional responses, or functional predictors, or both. In this work we have considered optimal designs for the case of functional response and non-functional predictor. A future goal is to develop the theory of optimal designs also for the scenarios with functional experimental conditions (see also [5]).

References

1. Aletti, G., May, C., Tommasi, C.: Optimal designs for linear models with functional responses. In: Bongiorno, E.G., Salinelli, E., Goia, A., Vieu, P. (eds.) Contributions in Infinite-Dimensional Statistics and Related Topics, pp. 19–24. Società Editrice Esculapio (2014)
2. Aletti, G., May, C., Tommasi, C.: Best estimation of functional linear models. Arxiv preprint 1412.7332. http://arxiv.org/abs/1412.7332 (2015)
3. Chiou, J.-M., Müller, H.-G., Wang, J.-L.: Functional response models. Stat. Sin. **14**, 675–693 (2004)
4. Fedorov, V. V.: Theory of Optimal Experiments. Probability and Mathematical Statistics, vol. 12. Academic, New York/London (1972). Translated from the Russian and edited by W.J. Studden, E.M. Klimko
5. Fedorov, V.V., Hackl, P.: Model-oriented Design of Experiments. Volume 125 of Lecture Notes in Statistics. Springer, New York (1997)
6. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis. Springer Series in Statistics. Springer, New York (2006)
7. Horváth, L., Kokoszka, P.: Inference for Functional Data with Applications. Springer Series in Statistics. Springer, New York (2012)
8. Kiefer, J.: General equivalence theory for optimum designs (approximate theory). Ann. Stat. **2**, 849–879 (1974)
9. Marley, C.J.: Screening experiments using supersaturated designs with application to industry. PhD thesis, University of Southampton (2011)
10. Marley, C.J., Woods, D.C.: A comparison of design and model selection methods for supersaturated experiments. Comput. Stat. Data Anal. **54**, 3158–3167 (2010)
11. Pukelsheim, F.: Optimal Design of Experiments. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York (1993)
12. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer Series in Statistics, 2nd edn. Springer, New York (2005)
13. Shen, Q., Faraway, J.: An F test for linear models with functional responses. Stat. Sin. **14**, 1239–1257 (2004)
14. Shen, Q., Xu, H.: Diagnostics for linear models with functional responses. Technometrics **49**, 26–33 (2007)

15. Silvey, S.D.: Optimal Design: An Introduction to the Theory for Parameter Estimation. Monographs on Applied Probability and Statistics. Chapman & Hall, London/New York (1980)
16. Woods, D.C., Marley, C.J., Lewis, S.M.: Designed experiments for semi-parametric models and functional data with a case-study in tribology. In: World Statistics Congress, Hong Kong (2013)

Optimal Designs for Implicit Models

Mariano Amo-Salas, Alfonso Jiménez-Alcázar, and Jesús López-Fidalgo

Abstract In this paper the tools provided by the theory of the optimal design of experiments are applied to a model where the function is given in implicit form. This work is motivated by a dosimetry problem, where the dose, the controllable variable, is expressed as a function of the observed value from the experiment. The best doses will be computed in order to obtain precise estimators of the parameters of the model. For that, the inverse function theorem will be used to obtain the Fisher information matrix. Properly the *D*-optimal design must be obtained directly on the dose using the inverse function theorem. Alternatively a fictitious *D*-optimal design on the observed values can be obtained in the usual way. Then this design can be transformed through the model into a design on the doses. Both designs will be computed and compared for a real example. Moreover, different optimal sequences and their *D*-effiencies will be computed as well. Finally, *c*-optimal designs for the parameters of the model will be provided.

1 Introduction

This paper is focused on the case of nonlinear models where the explanatory variable is expressed as a function of the dependent variable or response and this function is not invertible. That is, we consider the model

$$y = \eta(x, \theta) + \varepsilon, \quad \varepsilon \sim N(0, \sigma), \tag{1}$$

M. Amo-Salas
Faculty of Medicine, University of Castilla-La Mancha, Camino de Moledores s/n 13071-Ciudad Real, Spain
e-mail: Mariano.Amo@uclm.es

A. Jiménez-Alcázar
University of Castilla-La Mancha, Environmental Sciences Institute Avda. Carlos III s/n. 45071, Toledo, Spain
e-mail: Alfonso.JAlcazar@uclm.es

J. López-Fidalgo (✉)
Higher Technical School of Industrial Engineering, University of Castilla-La Mancha, Avda. Camilo José Cela 3, 13071-Ciudad Real, Spain
e-mail: Jesus.LopezFidalgo@uclm.es

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_2

where y is the dependent variable, x is the explanatory variable, θ is the vector of parameters of the model and $\mu(y,\theta) = \eta^{-1}(x,\theta)$ has a known expression, but a mathematical expression of $\eta(x,\theta)$ is not available. The challenge of this situation is to find optimal experimental designs for the explanatory variable when the expression of the function $\eta(x,\theta)$ is unknown. This situation is presented in a dosimetry study which will be used as case study in this work. Firstly, the description of the case study and a general introduction to the theory of Optimal Experimental Design is given. In Sect. 2 the inverse function theorem is applied to compute the information matrix. Finally, in Sect. 3 D-optimal designs are computed and compared for the case study proposed. Moreover, arithmetic and geometric optimal sequences, c-optimal designs and their D-effiencies are computed.

1.1 Case Study Background

The use of digital radiographs has been a turning point in dosimetry. In particular, radiochromic films are very popular nowadays because of their near tissue equivalence, weak energy dependence and high spatial resolution. In this area, calibration is frequently used to determine the right dose. The film is irradiated at known doses for building a calibration table, which will be used to fit a parametric model, where the dose plays the role of the dependent variable. The nature of this model is phenomenological since the darkness of the movie is only known qualitatively. An adjustment is necessary to filter noise and interpolate the unknown doses.

Ramos-García and Pérez-Azorín [9] used the following procedure. The radiochromic films were scanned twice. The first scanning was made when a pack of films arrived and the second 24 h after being irradiated. With the two recorded images the optical density, *netOD*, was calculated as the base 10 logarithm of the ratio between the means of the pixel values before (PV_0) and after (PV) the irradiation. They used patterns formed by 12 squares of $4 \times 4\,\text{cm}^2$ irradiated at different doses. This size is assumed enough to ensure the lateral electronic equilibrium for the beam under consideration. A resolution of 72 pp, without color correction and with 48-bit pixel depth was used for the measurements. The pixel values were read at the center of every square. Then, the mean and standard error were calculated. The authors assumed independent and normally distributed errors with constant variance as well as we do in this paper.

To adjust the results to the calibration table the following model was used:

$$netOD = \eta(D,\theta) + \varepsilon,$$

where D is the dose and the error ε will be assumed normally distributed with mean zero and constant variance, σ^2. The expression of the function $\eta(D,\theta)$ is unknown but the mathematical expression of the inverse is known

$$\eta^{-1}(D,\theta) = \mu(netOD,\theta) = \alpha\, netOD + \beta\, netOD^{\gamma},\ D \in [0,B], \tag{2}$$

where $\theta = (\alpha, \beta, \gamma)^T$ are unknown parameters to be estimated using maximum likelihood (MLE).

1.2 Optimal Experimental Design: General Background

Let a general nonlinear regression model be given by Equation (1). An *exact experimental design of size n* consists of a planned collection of points x_i, $i = 1, \ldots, n$, in a given compact *design space*, $\mathscr{X}$. Some of these points may be repeated and a probability measure can be defined assigning to each different point the proportion of times it appears in the design. This leads to the idea of extending the definition of experimental design to any probability measure (*approximate design*). It can be seen that, from the optimal experimental design viewpoint, we can restrict the search to finite designs of the type

$$\xi = \left\{ \begin{matrix} x_1 & x_2 & \ldots & x_k \\ p_1 & p_2 & \ldots & p_k \end{matrix} \right\},$$

where x_i, $i = 1, \ldots, k$ are the support points and $\xi(x_i) = p_i$ is the proportion of experiments made at point x_i. Thus, $p_i \geq 0$ and $\sum_{i=1}^{k} p_i = 1$.

For the exponential family of distributions the Fisher Information Matrix (FIM) of a design ξ is given by

$$M(\xi, \theta) = \sum_{x \epsilon \chi} I(x, \theta)\xi(x), \tag{3}$$

where $I(x, \theta) = \frac{\partial \eta(x,\theta)}{\partial \theta} \frac{\partial \eta(x,\theta)}{\partial \theta^T}$ is the FIM at a particular point x. If the model is nonlinear, in the sense that function $\eta(x, \theta)$ is nonlinear in the parameters, the FIM depends on the parameters and nominal values for them have to be provided.

It can be proved that the inverse of this matrix is asymptotically proportional to the covariance matrix of the parameter estimators. An optimal design criterion aims to minimize the covariance matrix in some sense and therefore the inverse of the information matrix, $\Phi[M(\xi, \theta)]$. For simplicity $\Phi(\xi)$ will be used instead of $\Phi[M(\xi, \theta)]$. In this paper two popular criteria will be used, *D*-optimality and *c*-optimality. The *D*-optimality criterion minimizes the volume of the confidence ellipsoid of the parameters and is given by $\Phi_D(\xi) = \det M^{-1/m}(\xi, \theta)$, where m is the number of parameters in the model. The *c*-optimality criterion is used to estimate a linear combination of the parameters, say $c^T\theta$, and is defined by $\Phi_c(\xi) = c^T M^{-}(\xi, \theta)c$. The superscript "$-$" stands for the generalized inverse class of the matrix. Although the generalized inverse is unique only for nonsingular matrices the value of $c^T M(\xi, \theta)^{-}c$ is constant for any representative of the generalized inverse class if and only if $c^T\theta$ is estimable with the design. These criterion functions are convex and non-increasing. A design that minimizes one of these functions Φ over

all the designs defined on $\mathscr{X}$ is called a Φ-optimal design, or more specifically, a D- or c-optimal design. It is worth to mention here that c-optimality raises important difficulties in nonlinear models when the optimal matrix is singular. In particular the actual covariance matrix may be different from the one predicted ([8], chap. 5).

The goodness of a design is measured by its efficiency, defined by

$$\text{eff}_{\Phi}(\xi) = \frac{\Phi(\xi^*)}{\Phi(\xi)}.$$

In order to check whether a particular design is optimal or not there is a celebrated equivalence theorem [4] for approximate designs and convex criteria. This theorem consists in verifying that the directional derivative is non–negative in all directions. More details on the theory of optimal experimental designs may be found, e.g., at [3, 7] or [1].

2 Inverse Function Theorem for Computing the FIM

In the general theory, the experiments are designed for the explanatory variable, x, assumed under the control of the experimenter. In the case studied in this work, the function $\eta(x, \theta)$ is unknown but we know $\mu(y, \theta) = \eta^{-1}(x, \theta)$. Therefore the FIM should be defined in terms of y instead of x. The FIM is then given by (3), in particular, for one point the FIM is

$$I(x, \theta) = \frac{\partial \eta(x, \theta)}{\partial \theta} \frac{\partial \eta(x, \theta)}{\partial \theta^T}.$$

We can calculate the FIM in terms of the response variable y through the inverse function theorem. Differentiating the equation

$$x = \mu(y, \theta) = \mu[\eta(x, \theta), \theta],$$

we obtain

$$0 = \left(\frac{\partial \mu(y, \theta)}{\partial y}\right)_{y=\eta(x,\theta)} \frac{\partial \eta(x, \theta)}{\partial \theta} + \left(\frac{\partial \mu(y, \theta)}{\partial \theta}\right)_{y=\eta(x,\theta)}.$$

Then

$$\frac{\partial \eta(x, \theta)}{\partial \theta} = -\left(\frac{\partial \mu(y, \theta)}{\partial y}\right)^{-1}_{y=\eta(x,\theta)} \left(\frac{\partial \mu(y, \theta)}{\partial \theta}\right)_{y=\eta(x,\theta)}. \tag{4}$$

Using this result the FIM can be computed and therefore optimal designs on the explanatory variable may be obtained. This is the same model to be used for design

when the variable y is heteroscedastic instead of homoscedastic with a sensitivity function (inverse of the variance),

$$\left| \left(\frac{\partial \mu(y,\theta)}{\partial y} \right)^{-1}_{y=\eta(x,\theta)} \right|.$$

This makes sense since assuming the response is a trend model plus some error with constant variance implies a trend model for x, which is the inverse of the original trend model plus an error with a non-constant variance coming from the transformation of the model.

3 Optimal Designs for the Case Study

In this section, the model proposed by the case study is considered. In this model, function $\eta(D,\theta)$ is unknown but $\eta^{-1}(D,\theta) = \mu(netOD,\theta)$ is known and defined by Equation (2). Computing the regressors vector with (4),

$$\frac{\partial \eta(D,\theta)}{\partial \theta} = \left[\frac{1}{\alpha_0 + \beta_0 \gamma_0 netOD^{\gamma_0 - 1}} \begin{pmatrix} netOD \\ netOD^{\gamma_0} \\ \beta_0 netOD^{\gamma_0} \log(netOD) \end{pmatrix} \right]_{netOD=\eta(D,\theta)},$$

where $\alpha_0, \beta_0, \gamma_0$ are some nominal values assumed for the parameters to compute the optimal design. Thus, the FIM for a design ξ is

$$M(\xi;\theta_0) = \sum_i \xi(D_i) I(D_i;\theta_0),$$

where $\theta_0^T = (\alpha_0, \beta_0, \gamma_0)$ and

$$I(D;\theta_0) = \frac{\partial \eta(D,\theta)}{\partial \theta} \frac{\partial \eta(D,\theta)}{\partial \theta^T}. \tag{5}$$

Now, the function of the original model, $\eta(D,\theta)$, needs to be plugged into these formulas instead of *netOD*, but it cannot be inverted analytically. Using the results of [9], the design space will be $\mathcal{X}_D = [0,B] = [0,972]$ and the following nominal values for the parameters will be considered: $\alpha_0 = 690, \beta_0 = 1550, \gamma_0 = 2$. For these values the inverse function will be computed numerically when needed.

Assuming the D-optimal design is a three–point design, it should have equal weights at all of them. Since D-optimality is invariant for reparametrizations, the D-optimal design is computed for variable *netOD* using matrix (5). Computing the determinant of the information matrix for an equally weighted design, $\xi_{netOD} = \{netOD_1, netOD_2, netOD_3\}$, and minimizing $\Phi_D(\xi_{netOD})$ for values of $netOD_1$,

$netOD_2$ and $netOD_3$ in the interval $\mathscr{X}_{netOD} = [0, b] = [0, 0.6]$ the following design is obtained: $\xi_{netOD} = \{0.091, 0.348, 0.6\}$. The equivalence theorem states numerically that this design is actually D-optimal.

Transforming the three points through the equation model $\mu(netOD, \theta)$, with the previous nominal values of the parameters, $D = 690netOD + 1550netOD^2$, the optimal design on D is $\xi_D = \{75.6, 427.8, 972\}$.

Now a design for *netOD* will be computed in the usual way for the function $\mu(netOD, \theta)$. This is the optimal design for a wrong MLE from the explicit inverse model. Then this design will be compared with the right one checking the loss of efficiency. We will consider *netOD* as the explanatory variable and after computing the optimal design for *netOD*, we will invert it to compute the design for D. That is, we consider $\mu(netOD, \theta)$ as the function of the original model. Using the previous nominal values the design space is then $\mathscr{X}_{netOD} = [0, 0.6]$, and the D-optimal design is obtained in a similar way as above, $\xi^I_{netOD} = \{0.137, 0.409, 0.6\}$.

The equivalence theorem states numerically that this design is actually D-optimal. At this point a design for the response, D, can be obtained by transforming the design points again using the equation model $\mu(netOD, \theta)$, $\xi^I_D = \{123.4, 541.0, 972\}$.

Apparently the design is quite different, e.g. the first points, 75.6 and 123.4, are quite different. But the efficiency of this design with respect to the optimal one, ξ_D, is rather high,

$$\text{eff}_D(\xi^I_D) = \left(\frac{\Phi_D(\xi_D)}{\Phi_D(\xi^I_D)}\right)^{\frac{1}{3}} = 0.924.$$

Usually experimenters do not like extremal designs with a few different points (three in this case). Before obtaining the data there is always a reasonable doubt about the right model to be used and therefore more different points provide more safety (frequently just subjective). In practice, it is common to find designs distributing the points equidistantly (arithmetic sequence) or with geometric decreasing or increasing distances. This is usually made in a reasonable way taking into account the experience and intuition of the experimenter, but sometimes they can be far from optimal among the different possibilities. López Fidalgo and Wong [6] optimized different types of sequences according to D-optimality, including arithmetic, geometric, harmonic and an arithmetic inverse of the trend model. In the example considered in this paper we knew by personal communication that there was particular interest in arithmetic sequences. Optimal sequences of ten points are considered for both variables, *netOD* and D.

Table 1 shows the equally weighted optimal sequence designs, including the fixed equidistant designs with the corresponding D-efficiencies. Subscripts stand for the variable in which the sequence is computed (A = arithmetic, G = geometric and E = equidistant). The last point is always the upper extreme of the design space.

The geometric sequence is quite efficient while the equidistant sequence is by far the worst design.

Table 1 Suboptimal designs according to different patterns and efficiencies (last point, 975, is omitted)

	Design points									D-eff (%)
ξ^A_{netOD}	49.0	107.3	176.6	257.1	348.5	451.1	564.7	689.4	825.1	78.1
ξ^A_D	57.2	158.8	260.5	362.1	463.7	565.4	667.	768.7	870.3	75.5
ξ^G_{netOD}	55.6	73.3	97.2	130.2	176.3	241.4	334.9	471.	671.8	76.9
ξ^G_D	52.9	73.1	101.1	139.7	193.0	266.7	368.5	509.1	703.4	77.8
ξ^E_{netOD}	0.0	52.8	119.5	200.0	294.2	402.2	524	659.5	808.8	71.1
ξ^E_D	0	108	216	324	432	540	648	756	864	64.9

Table 2 c-efficiencies (%) of the D-optimal design for estimating each parameter

γ	56.7
α	42.4
β	65.2

Table 3 Efficiencies (%) of the optimal designs with respect to the nominal value of γ

γ	ξ_D	ξ^α_D	ξ^β_D	ξ^γ_D	ξ^A_{netOD}	ξ^A_D	ξ^G_{netOD}	ξ^G_D
1.8	99.5	95.5	99.6	98.2	98.6	99.2	98.9	98.9
2.2	99.4	95.6	99.5	98.0	99.7	99.3	98.8	98.9

In the example considered here, there is special interest in accurately estimating the parameter γ. Elfving's method [2] is a graphical procedure for calculating c-optimal designs. Although the method can be applied to any number of parameters it is not used directly for more than two parameters. López-Fidalgo and Rodríguez-Díaz [5] proposed a computational procedure for finding c-optimal designs using Elfving's method for more than two dimensions.

The c-optimal designs to estimate each of the parameters of the model are

$$\xi^\alpha_D = \begin{Bmatrix} 46.25 & 439.36 & 972 \\ 0.742 & 0.186 & 0.0717 \end{Bmatrix}, \quad \xi^\beta_D = \begin{Bmatrix} 170.7 & 972 \\ 0.622 & 0.378 \end{Bmatrix},$$

$$\xi^\gamma_D = \begin{Bmatrix} 46.25 & 439.36 & 972 \\ 0.476 & 0.359 & 0.165 \end{Bmatrix}.$$

Table 2 shows the c-efficiencies of the D-optimal design for estimating each parameter. These efficiencies are low, specifically the c_γ-efficiency is lower than 60 %. The $c-$optimal design for β is a singular two–point one.

Table 3 displays the efficiencies of the different designs computed with respect the misspecification of parameter γ. This parameter can be considered as the most important parameter. Its nominal value is $\gamma_0 = 2$. The sensitivity analysis has been performed considering a deviation of ± 10 % from this value. The efficiencies from Table 3 show that the optimal designs are rather robust with respect to the misspecification of this parameter.

4 Concluding Remarks

This work deals with the problem of a model where the function is given in implicit form. In this case the FIM could not be computed in the usual way because the expression of the function of the model is unknown. Using the inverse function theorem, the FIM can be obtained and the D-optimal design may be computed. The D-optimal design was also determined directly on the dependent variable and then it was transformed into a design on the explanatory variable. This design displayed a moderate loss of efficiency when compared with the right one in this particular case.

Dependent errors or other distribution for them can be treated as well and it is one the future research lines.

Since three–point designs may be not acceptable from a practical point of view, ten different points were forced to be in the design restricting them to follow a regular sequence. In particular, arithmetic, geometric and inverse (through the trend model) sequences were considered. All of them were more efficient than the sequence used by the researchers. The geometric sequence achived the highest efficiency.

Finally, c-optimal designs for estimating the parameters of the model were computed. The c-efficiencies of the D-optimal design were lower than 70 % and specifically the c_γ-efficiency was lower than 60 %.

Acknowledgements The authors have been sponsored by Ministerio de Economía y Competitividad and fondos FEDER MTM2013-47879-C2-1-P. They want to thank the two referees for their interesting comments.

References

1. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs with SAS. Oxford Statistical Science Series. Oxford University Press, Oxford (2007)
2. Elfving, G.: Optimum allocation in linear regression theory. Ann. Math. Stat. **23**, 255–262 (1952)
3. Fedorov, V.V., Hackl, P.: Model-Oriented Design of Experiments. Lecture Notes in Statistics. Springer, New York (1997)
4. Kiefer, J.: General equivalence theory for optimum designs (aproximate theory). Ann. Stat. **2**(5), 848–879 (1952)
5. López-Fidalgo, J., Rodríguez-Díaz, J.M.: Elfving's methods for m-dimensional models. Metrika **59**, 235–244 (2004)
6. López Fidalgo, J., Wong, W.K.: Design for the Michaelis-Menten model. J. Theor. Biol. **215**, 1–11 (2002)
7. Pazman, A.: Foundations of Optimum Experimental Design. D. Reidel publishing company, Dordrecht (1986)
8. Pronzato, L., Pazman, A.: Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties. Springer, New York (2013)
9. Ramos-García, L.I., Pérez-Azorín, J.F.: Improving the calibration of radiochromic films. Med. Phys. **40**(7), 071726 (2013)

Optimum Experiments with Sets of Treatment Combinations

Anthony C. Atkinson

Abstract Response surface designs are investigated in which points in the design region corresponds to single observations at each of s distinct settings of the explanatory variables. An extension of the "General Equivalence Theorem" for D-optimum designs is provided for experiments with such sets of treatment combinations. The motivation was an experiment in deep-brain therapy in which each patient receives a set of eight distinct treatment combinations and provides a response to each. The experimental region contains sixteen different sets of eight treatments.

1 Introduction

The scientific motivation is an experiment in deep-brain therapy in which each patient receives a set of eight treatment combinations and provides a response to each. The structure of such experiments is more easily seen in a response surface setting where each choice of an experimental setting provides a response at each of s distinct settings of the explanatory variables. Throughout the focus is on D-optimum designs for homoskedastic linear models.

The paper starts in Sect. 2 with numerical investigation of designs for a first-order model with two continuous explanatory variables. The numerical results suggest an extension of the "General Equivalence Theorem" of [9] which is presented in Sect. 3 along with references to related results. Some discussion of numerical algorithms is in Sect. 4. The paper concludes in Sect. 5 with consideration of extensions including that to Generalized Linear Models and a discussion of experimental design in the motivating medical example.

A.C. Atkinson (✉)
Department of Statistics, London School of Economics, London WC2A 2AE, UK
e-mail: a.c.atkinson@lse.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_3

2 A Simple Response Surface Example

The simple response surface model in two variables is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} + \epsilon_i, \tag{1}$$

where the independent errors ϵ_i have constant variance σ^2 and the design region $\mathscr{X}$ is the unit square $[-1, 1]^2$. Estimation of β is by least squares.

As is standard in the theory of optimum experimental design, an experimental design ξ places a fraction w_i of the experimental trials at the conditions x_i. A design with n points of support is written as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \ldots & x_n \\ w_1 & w_2 & \ldots & w_n \end{array} \right\}, \tag{2}$$

where $w_i > 0$ and $\sum_{i=1}^{n} w_i = 1$. Any realisable experimental design for a total of N trials will require that the weights are ratios of integers, that is $w_i = r_i/N$, where r_i is the number of replicates at condition x_i. The mathematics of finding optimal experimental designs and demonstrating their properties is greatly simplified, as in this paper, by the consideration of continuous designs in which the integer restriction is ignored.

In general, the linear model (1) is written

$$y_i = \beta^T f(x_i) + \epsilon_i. \tag{3}$$

The parameter vector β is $p \times 1$, with $f(x_i)$ a known function of the explanatory variables x_i.

The information matrix for the design ξ with n support points is written

$$M(\xi) = \sum_{i=1}^{n} w_i f(x_i) f(x_i)^T = F^T W F, \tag{4}$$

where F is the $n \times p$ extended design matrix, with ith row $f^T(x_i)$ and W is a diagonal matrix of weights.

D-optimum designs, minimizing the generalized variance of the estimates of β, maximize the determinant $|F^T W F|$ over the design region $\mathscr{X}$ through choice of the optimum design ξ^*. For the two variable model (1) the D-optimum design is the 2^2 factorial with support at the corners of $\mathscr{X}$, so that $w_i = 0.25, (i = 1, \ldots, 4)$.

That this design is D-optimum can be shown by use of the "general equivalence theorem" for D-optimality [9] which provides conditions for the optimality of a design ξ which depend on the sensitivity function

$$d(x, \xi) = f^T(x) M^{-1}(\xi) f(x). \tag{5}$$

Table 1 D-optimum design for two variable model with sets of two points

Obs.	Set	x_1	x_2	w_i	w_i^{SET}	$d(x_i)$	$d_{\text{AVE}}(x_i)$
1	1	−1.0	−1.0	0.125	0.25	4.42	4.00
2	1	−0.9	−0.9	0.125	0.25	3.58	4.00
3	2	1.0	−0.9	0.125	0.25	4.44	4.00
4	2	0.9	−0.8	0.125	0.25	3.56	4.00
5	3	−0.7	1.0	0.125	0.25	3.96	4.00
6	3	−0.9	0.8	0.125	0.25	4.04	4.00
7	4	1.0	0.9	0.125	0.25	4.69	4.00
8	4	0.8	0.8	0.125	0.25	3.31	4.00
9	5	0.3	0.3	0.0	0.0	1.22	1.16
10	5	0.2	0.2	0.0	0.0	1.09	1.16
11	6	0.4	−0.4	0.0	0.0	1.43	1.31
12	6	0.4	−0.1	0.0	0.0	1.19	1.31

For the optimum design $\bar{d}(x, \xi^*)$, the maximum value of the sensitivity function over $\mathscr{X}$, equals p, the number of parameters in the linear predictor. These values occur at the points of support of the design x_i. For the 2^2 factorial it is straightforward to show that these maximum values are four.

Instead of single observations, now suppose that the experimental design consists of the choice of pairs of experimental conditions. An example is in Table 1. There are twelve observations, grouped into the six sets in column 2. The design problem is to find the six weights for these sets that give the D-optimum design.

The structure of the design is exhibited in Fig. 1. The black dots are close to the four design points of the 2^2 factorial, optimum for single observations. The four crosses, combined with the conditions at their nearest dot, form the first four sets in the table. The remaining two sets of points, represented by open circles, have been chosen to be at conditions close to the centre points typically used in response surface work for model checking.

The observation numbers are in the first column of Table 1, with the membership of the sets of two observations in column 2. Columns 3 and 4 give the values of the two explanatory variables. The D-optimum weights at the 12 points of potential observation are given in the fifth column of the table, with the weights for the sets in column six. The design, like the design for individual observations, has four points of support, the last two sets having zero weight. The weights for the four sets included in the design are 0.25, as they are for the 2^2 factorial for single observations.

The most interesting results are the values of the sensitivity functions in the last two columns of the table. There are four parameters in the model, so the D-optimum design for individual observations had a value of four for the sensitivity function at the points of the optimum design. Here, for three of the first four sets, the near optimum point had a value slightly greater than four, with the related point, represented by a cross, having a value slightly less than four. The implication is, if it

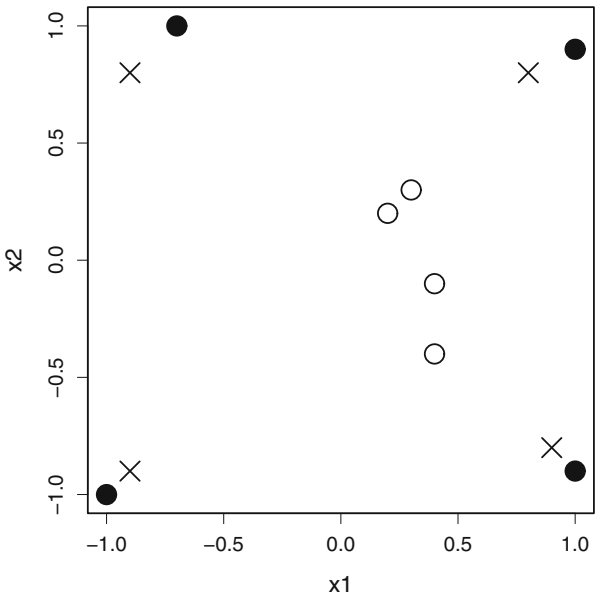

Fig. 1 Sets of points: the *black circles* are close to or at the support points of the 2^2 factorial, with the nearby X the second in each set of observations. The *unfilled circles* denote two further pairs of points

were possible, that the 'crosses' should be moved closer to the 'dots'. The exception is set three, where both values of the sensitivity function are close to four since the 'dot' and the 'cross' are a similar distance from the support point of the 2^2 factorial. The values of the sensitivity functions for the other two sets are not much above one, an indication that readings close to the centre point are not informative about parameters other than β_0.

The last column gives the average values of the sensitivity functions for each set. These are exactly four for the four sets which are included in the optimum design. The implications for a generalization of the equivalence theorem are considered in the next section.

3 Equivalence Theorem

The numerical results for designs with sets of points suggest that an equivalence theorem applies that is an extension of that for individual observations.

Some notation is needed. Let S_i denote the ith set of observations, taken at points $x_{i1}, x_{i2}, \ldots, x_{is}$ and let

$$d_{\text{AVE}}(i, \xi) = \sum_{j \in S_i} d(x_{ij}, \xi)/s. \tag{6}$$

Further, let $\bar{d}_{\text{AVE}}(\xi)$ be the maximum over $\mathscr{X}$ of $d_{\text{AVE}}(i, \xi)$.

Then the **Equivalence Theorem** states the equivalence of the following three conditions on ξ^*:

1. The design ξ^* maximizes $|M(\xi)|$;
2. The design ξ^* minimizes $\bar{d}_{\text{AVE}}(\xi)$;
3. The value of $\bar{d}_{\text{AVE}}(\xi^*) = p$, this maximum occurring at the points of support of the design.

As a consequence of 3, we obtain the further condition:

4. For any non-optimum design the value of $\bar{d}_{\text{AVE}}(\xi) > p$.

The proof of this theorem follows from the additive nature of the information matrix. Standard proofs of the equivalence theorem, such as those in [10, §5.2] and [7, §2.4.2] depend on the directional derivative at a point in $\mathscr{X}$. Here, with the extension to a set of observations, the directional derivative is the sum of the derivatives for the individual observations.

The result also follows immediately by considering the s observations in each set as a single multivariate observation. In the customary multivariate experiment, observation i consists of measurements of s different responses taken at the point x_i. Here the same response is measured at the set of s conditions defined by S_i. However, standard results such as those in Theorem 1 and the first line of Table 1 of [4] not only prove the equivalence theorem but show how to handle correlation between observations in the same set.

The assumption in this paper is that all sets contain the same number, s, of design points. With sets containing different numbers of observations, standardization by the number of design points allows comparison of the efficiency of individual points in a set, as in Table 1. However, this aspect of optimality is not always the major concern.

In a pharmacokinetic experiment described by [7, §7.3.1] interest is in the effect of sampling at fewer than the total possible number of time points, in their case 16. Dropping a few non-optimal design points will move the normalized design ξ_{16} towards optimality. But the variances of the parameter estimates from fewer than 16 observations will be increased. The question is by how much? Standardization by the number of sampling points is then appropriate. In this example, an eight-point design leads to only a slight increase in the variances of virtually all of the parameter estimates and results in reduced sampling costs. Such costs can be introduced explicitly; [6] formulated optimum design criteria when costs are included in experiments with individual observations. Fedorov and Leonov [7, Chapter 7] presents several pharmacokinetic applications.

4 Algorithms

Numerical algorithms are essential for the construction of any but the simplest optimum designs. Much of the discussion in the literature, for example [7, Chapter 3], stresses the desirability of using algorithms that take account of the specific structure of optimum designs. However, the design for this paper was found using a general purpose numerical algorithm.

There are often two sets of constraints in the maximization problem of finding an optimum design. The first is on the design weights which must be non-negative and sum to one. The other is on the design points, which must be within $\mathscr{X}$. However, in the design of this paper, $\mathscr{X}$ contained six specified pairs of potential support points, so that only the weights had to be found. Atkinson et al. [3, §9.5] suggest search over an unconstrained space Ψ, using a transformation to polar co-ordinates to calculate weights w_i that satisfy the required constraints. Here use was made of a simpler approach.

The search variables are ψ_i. Taking

$$w_i = \psi_i^2 / \sum_{i=1}^{n} \psi_j^2 \tag{7}$$

provides weights that satisfy the required constraints. Of course, the w_i are in $n-1$ dimensions, so that (7) is not unique; the same weights are obtained when all ψ_i are replaced by $a\psi_i$, $(a \neq 0)$.

A related transformation takes advantage of the upper and lower constraints on variables in the R function `optim`. Now let

$$w_i = \psi_i / \sum_{i=1}^{n} \psi_j \tag{8}$$

with $0 \leq \psi_j \leq 1$, which again provides weights that satisfy the constraints. As before, the weights are not a unique function of ψ.

Comparisons for a small number of design problems, including generalized linear models, did not reveal situations in which the Quasi-Newton BGFS algorithm in `optim` had difficulty in converging. However, convergence with the weights (8) was always better than that with (7).

5 Extensions

There are several ways in which the results of this paper on sets of observations can be extended, for example through the use of other criteria of design optimality. A straightforward extension is to generalized linear models, where the design criteria

are weighted versions of those for regression. Some examples for logistic regression are given by [2] who includes plots of the design in the induced design region [8]. See [5] for a survey of recent results in designs for such models with individual observations.

However, the most important application may well be in simplifying the study of optimum designs in the medical experiment in deep-brain therapy that provided motivation for this paper. In this experiment there are two factors, stimulation at three levels and conditions at four levels. There are thus twelve treatment combinations. However, for safety reasons, it is not possible to expose each patient to all twelve. Instead, it was proposed to take measurements at only eight combinations; sixteen such sets were chosen. The design region thus contained sixteen distinct points, each of which would give a set of eight measurements from one patient.

A design question is, which of the sixteen sets should be used and in what proportions? Since the linear model for the factors contains only six parameters, it is unlikely that all sixteen points in the design region need to be included in the experiment. Even if an optimum design satisfying the equivalence theorem does include all sixteen, it may not be unique; there may be optimum designs requiring fewer distinct design points of which the sixteen-point design is a convex linear combination.

The equivalence theorem also provides a method of treatment allocation in clinical trials in which patients arrive sequentially. In the experiment in deep-brain therapy there is a prognostic factor, initial severity of the disease. The effect of this variable is not the focus of the trial, so that it would be considered a nuisance factor. Sequential construction of the D_S-optimum design for the treatment effects would aim for balance over the prognostic factor and lead to the most efficient inference about the treatments. However, such deterministic allocation rules are unacceptable in clinical trials, where they may lead to selection bias. A randomized rule based on D-optimality, such as those described by [1], should instead be used.

For such data, the assumption of independent errors might with advantage be replaced by a linear mixed model, as described in [11], that allows for correlation between observations from individual patients. Recent references on optimum design for such models can be found in [7, Chapter 7].

Acknowledgements I am grateful to Dr David Pedrosa of the Nuffield Department of Clinical Neurosciences, University of Oxford, for introducing me to the experimental design problem in deep-brain therapy that provided the motivation for this work.

I am also grateful to the referees whose comments strengthened and clarified the results of §3.

References

1. Atkinson, A.C.: Optimal model-based covariate-adaptive randomization designs. In: Sverdlov, O. (ed.) Modern Adaptive Randomized Clinical Trials: Statistical and Practical Aspects, pp. 131–154. Chapman and Hall/CRC Press, Boca Raton (2015)

2. Atkinson, A.C.: Optimum experiments for logistic models with sets of treatment combinations. In: Fackle-Fornius, E. (ed.) A Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday, pp. 44–58. Department of Statistics, Stockholm University, Stockholm (2015)
3. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)
4. Atkinson, A.C., Fedorov, V.V., Herzberg, A.M., Zhang, R.: Elemental information matrices and optimal experimental design for generalized regression models. J. Stat. Plan. Inference **144**, 81–91 (2014)
5. Atkinson, A.C., Woods, D.C.: Designs for generalized linear models. In: Dean, A., Morris, M., Stufken, J., Bingham, D. (eds.) Handbook of Design and Analysis of Experiments, pp. 471–514. Chapman and Hall/CRC Press, Boca Raton (2015)
6. Elfving, G.: Optimum allocation in linear regression theory. Ann. Math. Stat. **23**, 255–262 (1952)
7. Fedorov, V.V., Leonov, S.L.: Optimal Design for Nonlinear Response Models. Chapman and Hall/CRC Press, Boca Raton (2014)
8. Ford, I., Torsney, B., Wu, C.F.J.: The use of a canonical form in the construction of locally optimal designs for non-linear problems. J. R. Stat. Soc. Ser. B **54**, 569–583 (1992)
9. Kiefer, J., Wolfowitz, J.: The equivalence of two extremum problems. Can. J. Math. **12**, 363–366 (1960)
10. Pronzato, L., Pázman, A.: Design of Experiments in Nonlinear Models. Springer, New York (2013)
11. Verbeke, G., Molenberghs, G.: Linear Mixed Models for Longitudinal Data. Springer, New York (2000)

Design Keys for Multiphase Experiments

R.A. Bailey

Abstract Desmond Patterson introduced the design key in 1965 in the context of rotation experiments in agriculture. When there are many factors involved, the design key gives an algorithm for constructing the design and for keeping track of confounding. Here I extend the idea to multiphase experiments, using one design key for each phase.

1 Experiments with a Single Phase

In [12], Patterson introduced the design key as an aid to the design and evaluation of experiments in crop rotation. In such experiments, the set Γ of treatments and the set Ω of experimental units both have complicated factorial structure. The design key provides, first, an algorithm for constructing the design and, secondly, an algorithm for identifying the confounding of factorial treatment effects with factorial effects of experimental units.

In the simplest case, every factor has p levels, where p is prime. The levels are identified with the integers modulo p. Factors on the same set are combined in the way introduced by Fisher in [7], except that additive notation is used. As explained in [5, Chapter 1], there is a set $\mathscr{F}$ of treatment factors, giving one potential treatment for each combination of levels of the factors in $\mathscr{F}$. Similarly, there is one experimental unit for each combination of levels of the unit factors in set $\mathscr{G}$.

Each treatment factor F is converted by the design key Φ to a factor $\Phi(F)$ on the experimental units, which is a linear combination of the factors in $\mathscr{G}$, the coefficients being in $\mathbb{Z}_p$. For example, if $\mathscr{G} = \{G_1, \ldots, G_m\}$ then there are elements $\beta_1, \ldots, \beta_m$ of $\mathbb{Z}_p$ such that $\Phi(F) = \beta_1 G_1 + \cdots + \beta_m G_m$. Denote by $G_i(\omega)$ the level of factor G_i on experimental unit ω. The design key Φ specifies that the level of treatment factor F on ω is $\beta_1 G_1(\omega) + \cdots + \beta_m G_m(\omega)$. Applying this rule to all treatment factors in $\mathscr{F}$ gives the complete description of the treatment allocated to experimental unit ω.

R.A. Bailey (✉)
School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK
e-mail: rab24@st-andrews.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_4

Example 1 (Graeco-Latin square) The experimental units form a 5×5 square array of plots defined by two five-level factors, Rows and Columns, which are abbreviated to R and C. The 25 treatments are all combinations of levels of two other five-level factors, Variety of Wheat (W) and Quantity of Nitrogen (N). The convention is that each factor is represented by a single letter, and that these letters are all different.

The constraints on the design are that all combinations of levels of W and N should occur, and that W and N should both be orthogonal to both Rows and Columns. We use the design key Φ, where $\Phi(W) = R + C$ and $\Phi(N) = R + 2C$. Figure 1 shows how the plots are identified by the levels of R and C, how these levels are combined to give $W(\omega) = R(\omega) + C(\omega)$, shown as the first entry on each plot, and $N(\omega) = R(\omega) + 2C(\omega)$, shown as the second entry on each plot.

The second use of the design key is to identify confounding. Each non-zero linear combination of factors in $\mathscr{F}$ corresponds to $p - 1$ degrees of freedom (df) for contrasts among treatments, all for the same treatment effect: see [7]. Moreover, if one linear combination is a non-zero multiple of another, then they correspond to the same df; otherwise, the corresponding sets of contrasts are orthogonal to each other. When the experimental units have a poset block structure defined by the factors in $\mathscr{G}$, there is a similar result: each non-zero linear combination of factors in $\mathscr{G}$ corresponds to $p - 1$ df in the same stratum: see [1]. In Example 1, R gives 4 df for Rows, C gives 4 df for Columns, and the 16 df for the Rows-by-Columns interaction can be grouped into four orthogonal sets of four, one set corresponding to each of the linear combinations $R + C$, $R + 2C$, $R + 3C$ and $R + 4C$. The 24 df for treatment contrasts are broken down similarly, using linear combinations of W and N.

Suppose that $\mathscr{F} = \{F_1, \ldots, F_n\}$. Because all addition is done modulo p, Φ becomes a homomorphism from a group of order p^n to a group of order p^m. Thus if $\alpha_1 F_1 + \cdots + \alpha_n F_n$ is any linear combination of treatment factors then

$$\Phi(\alpha_1 F_1 + \cdots + \alpha_n F_n) = \alpha_1 \Phi(F_1) + \cdots + \alpha_n \Phi(F_n).$$

The design key method ensures that the $p - 1$ treatment df corresponding to the linear combination $\alpha_1 F_1 + \cdots + \alpha_n F_n$ are completely confounded with the $p - 1$ df for experimental units given by $\alpha_1 \Phi(F_1) + \cdots + \alpha_n \Phi(F_n)$.

Fig. 1 Graeco-Latin square given by $\Phi(W) = R + C$ and $\Phi(N) = R + 2C$; the first entry on each plot is the level of W, the second is the level of N

	C				
R	0	1	2	3	4
0	0,0	1,2	2,4	3,1	4,3
1	1,1	2,3	3,0	4,2	0,4
2	2,2	3,4	4,1	0,3	1,0
3	3,3	4,0	0,2	1,4	2,1
4	4,4	0,1	1,3	2,0	3,2

Table 1 Confounding in Example 1

Unit effect	Unit linear combination	df	Treatment linear combination	Treatment effect
Rows	R	4	$W + 2N$	Interaction
Columns	C	4	$W + 4N$	Interaction
Rows-by-Columns	$R + C$	4	W	Wheat-variety main
	$R + 2C$	4	N	Quantity-of-nitrogen main
	$R + 3C$	4	$W + 3N$	Interaction
	$R + 4C$	4	$W + N$	Interaction

In Example 1, $\Phi(W)$ and $\Phi(N)$ both have non-zero coefficients of both R and C, so the main effects of W and N are completely confounded with orthogonal parts of the Rows-by-Columns interaction. For part of the Wheat-by-Nitrogen interaction,

$$\Phi(W + 2N) = \Phi(W) + 2\Phi(N) = R + C + 2(R + 2C) = 3R \equiv R,$$

so these four df are totally confounded with the main effect of Rows. Here we write "$\equiv$" to mean "represents the same contrasts as". The confounding is shown in full in Table 1, where horizontal lines indicate the decomposition into factorial unit effects.

If a factor has p^r levels, where $r \geq 2$, then it is represented by r pseudofactors, each with p levels. The convention is that these pseudofactors are written with the same single letter and subscripts $1, \ldots, r$. A further complication is that the factors in $\mathscr{G}$ may not be completely crossed. To identify the factorial effect corresponding to a linear combination of factors or pseudofactors: (i) write down all the letters which occur, ignoring subscripts; (ii) if factor C is nested in factor D and letter C occurs then include letter D; (iii) remove any duplicate letters. The set of letters remaining gives the factorial effect (different authors have different conventions for punctuating this, such as CD or $C[D]$).

Example 2 (Factorial design in blocks) There are four 2-level treatment factors: S, T, U and V. The 16 experimental units are arranged in four blocks of four plots each, so there are pseudofactors B_1 and B_2 for blocks and P_1 and P_2 for plots within blocks. We want a design in which all treatment main effects, and as many treatment two-factor interactions as possible, are orthogonal to blocks. This can be achieved with the design key Φ, where

$$\Phi(S) = P_1, \quad \Phi(T) = P_2, \quad \Phi(U) = B_1 + P_1 + P_2, \quad \Phi(V) = B_2 + P_1 + P_2.$$

This gives the design in Fig. 2.

	Block 1				Block 2				Block 3				Block 4			
B_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
B_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
P_1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
P_2	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
S	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
T	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
U	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
V	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1

Fig. 2 Factorial design in incomplete blocks given by $\Phi(S) = P_1$, $\Phi(T) = P_2$, $\Phi(U) = B_1 + P_1 + P_2$ and $\Phi(V) = B_2 + P_1 + P_2$

Now $\Phi(S + T + U) = B_1$, $\Phi(S + T + V) = B_2$ and $\Phi(U + V) = B_1 + B_2$, so the treatment interactions S#T#U, S#T#V and U#V are confounded with Blocks while all the other treatment factorial effects are confounded with Plots within Blocks.

More information about design keys for experiments with a single phase can be found in [5, 6, 8, 10, 12, 14]. As these references show, experience with both design of experiments and use of the design key is needed in order to choose a suitable design key. Alternatively, the software **planor** [9] described in [11] may be used.

2 Two-Phase Experiments

In the first phase of a two-phase experiment, the treatments in Γ are allocated to the experimental units in Ω_1. In the second phase, produce from each unit in Ω_1 is allocated to one or more units in Ω_2, which is the set of experimental units for that phase. Typically, responses are measured on each unit of Ω_2, but not on units of Ω_1.

We assume that elements of Γ, Ω_1 and Ω_2 are defined by sets $\mathscr{F}$, $\mathscr{G}_1$ and $\mathscr{G}_2$ of factors or pseudofactors, each with p levels, in addition to any nesting relationships between factors on the same set. Each set of factors, with its nesting relationships, can be shown on a panel diagram like those introduced in [3].

Suppose that a design key Φ is used to allocate treatments to units in the first phase, and a design key Ψ is used to allocate first-phase units to second-phase units. Then the composite function $\Psi \circ \Phi$ becomes a design key between Γ and Ω_2, and the confounding rules already given make it surprisingly easy to find the confounding between factorial effects of treatments and those of experimental units in each phase. In turn, this shows how to decompose the ANOVA table as described in [4]. If both Ω_1 and Ω_2 give random effects with spectral components of variance defined by the randomization, as in [2], then the skeleton ANOVA table can be expanded to show expected mean squares, which are helpful for evaluating the design.

Example 3 (Proteomics) In a proteomics experiment described by Kathy Ruggiero of the University of Auckland, Γ consists of all combinations of two Interventions

with two types of Tissue. The first phase uses eight Cages of two Animals each. Interventions must be allocated to whole cages. After these have been given time to take effect, two body Positions are selected within each animal, and a tissue of each type extracted, one per position.

In the second phase, two samples from each tissue are analysed in the laboratory. Eight samples are processed in each of eight Runs of the machine. In each run, the samples are identified by eight different coloured dyes, which we here call Labels. Figure 3 shows the three panel diagrams.

The factors and pseudofactors, all with two levels, are I and T for $\mathscr{F}$; C_1, C_2, C_3, A and P for $\mathscr{G}_1$; and R_1, R_2, R_3, L_1, L_2 and L_3 for $\mathscr{G}_2$. There is no loss of generality in putting $\Phi(I) = C_1$. Amongst the normalized treatment effects, the main effect of I probably has the biggest variance in the first phase, so we should confound C_1 with the lowest-variance effect in the second phase, which is probably the Runs-by-Labels interaction. To avoid losing df for the relevant residual for I, we should also confound C_2 and C_3 with this interaction. Thus we put $\Psi(C_i) = R_i + L_i$ for $i = 1, 2, 3$. Then $\Psi(\Phi(I)) = R_1 + L_1$. Finally, we would like Tissues and the Interventions-by-Tissues interaction to be orthogonal to both Runs and Labels. This can be achieved by putting $\Psi(A) = R_1$, $\Psi(P) = L_2$ and $\Phi(T) = C_3 + P$. Then $\Phi(I + T) = C_1 + C_3 + P$, $\Psi(\Phi(T)) = R_3 + L_2 + L_3$ and $\Psi(\Phi(I + T)) = R_1 + R_3 + L_1 + L_2 + L_3$.

The skeleton ANOVA is in Table 2. Here Animals[C]$_1$ denotes the single df for A, which is confounded with R_1. The single df for $C_1 + A$ is also part of the effect

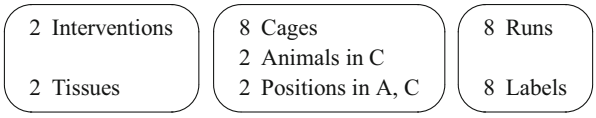

Fig. 3 Panel diagrams in Example 3

Table 2 Skeleton analysis of variance in Example 3

Lab units		Animal-bits		Treatments		
Source	df	Source	df	Source	df	EMS
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(\mathrm{I})$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{T})$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(\mathrm{IT})$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

of Animals within Cages; it is confounded with L_1 and is denoted Animals[C]$_2$. The remaining six df for Animals within Cages are confounded with the Runs-by-Labels interaction and are denoted Animals[C]$_3$. The effect of Positions within Animals and Cages is decomposed into three pieces in a similar way.

Full horizontal lines show the decomposition into factorial effects of the second-phase units. Each of these is further decomposed by the factorial effects of the first-phase units, as shown by incomplete horizontal lines. For example $R_1 = \Psi(A)$, $R_2 = \Psi(C_2 + P)$ and $R_1 + R_2 = \Psi(C_2 + A + P)$ while the remaining contrasts for Runs are orthogonal to the effects of first-phase units are so are denoted Residual. Some second-phase effects are even further decomposed by the factorial treatment effects.

The final column shows expected mean squares (EMS). Here ξ_R, ξ_L and ξ_{RL} are the spectral components of variance associated with Runs, Labels and Runs-by-Labels in Ω_2, while η_C, η_{CA} and η_{CAP} are the analogous spectral components of variance for Ω_1. Appropriate quadratic forms in the fixed main effects of Interventions and Tissues, and their interaction I#T, are shown as $q(I)$, $q(T)$ and $q(IT)$.

Example 4 (Field then laboratory) Ruth Butler of the New Zealand Institute for Crop and Food Research Limited discussed another example where the output from the first phase is analysed in the laboratory. In the first phase, 27 Varieties of cereal are grown in 81 Plots, arranged in three Rows crossed with three Columns. In the second phase, one sample of grain is taken from each plot; the samples are analysed in Batches of nine. Suitable factors and pseudofactors, all with three levels, are V_1, V_2 and V_3 for $\mathscr{F}$; R, C, P_1 and P_2 for $\mathscr{G}_1$; and B_1, B_2, S_1 and S_2 for $\mathscr{G}_2$.

In the first phase, no treatment effect should be confounded with Rows or Columns, but then at least two treatment df must be confounded with the interaction R#C. Without loss of generality, $\Phi(V_1) = P_1$, $\Phi(V_2) = P_2$ and $\Phi(V_3) = R + C$.

Likewise, in the second phase, at least two treatment df must be confounded with Batches. Table 3 shows three possible design keys for this phase, with the corresponding skeleton ANOVA tables, which help us to assess these designs. The notation is similar to that in Table 2, and $V \vdash V_3$ means all variety effects orthogonal to V_3.

There can never be more than two residual df to compare the effect of V_3 with. The first design key gives a larger EMS for this residual. In the other two designs, V_1 also has few residual df, both with the same EMS. There are more residual df for V_1 in the last design, so this should be preferred.

Table 3 Three possible designs keys, with consequent skeleton ANOVA tables, in Example 4

Design key	Skeleton ANOVA table						
	Samples		Plots		Varieties		
	Source	df	Source	df	Source	df	EMS
$\Psi(R) = B_1$	Batches	8	Rows	2			$\xi_B + \eta_R$
$\Psi(C) = B_2$			Columns	2			$\xi_B + \eta_C$
$\Psi(P_1) = S_1$			R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
$\Psi(P_2) = S_2$					Residual	2	$\xi_B + \eta_{RC}$
	Samples[B]	72	Plots[R,C]	72	$V \vdash V_3$	24	$\xi_{BS} + \eta_{RCP} + q(V \vdash V_3)$
					Residual	48	$\xi_{BS} + \eta_{RCP}$
	Samples		Plots		Varieties		
	Source	df	Source	df	Source	df	EMS
	Batches	8	Rows	2			$\xi_B + \eta_R$
$\Psi(R) = B_1$			Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
$\Psi(C) = S_1$					Residual	4	$\xi_B + \eta_{RCP}$
$\Psi(P_1) = B_2$	Samples[B]	72	Columns	2			$\xi_{BS} + \eta_C$
$\Psi(P_2) = S_2$			R#C	4	V_3	2	$\xi_{BS} + \eta_{RC} + q(V_3)$
					Residual	2	$\xi_{BS} + \eta_{RC}$
			Plots[R,C]$_2$	66	$V \vdash V_1 \vdash V_3$	22	$\xi_{BS} + \eta_{RCP} + q(V \vdash V_1 \vdash V_3)$
					Residual	44	$\xi_{BS} + \eta_{RCP}$
	Samples		Plots		Varieties		
	Source	df	Source	df	Source	df	EMS
	Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
$\Psi(R) =$					Residual	6	$\xi_B + \eta_{RCP}$
$B_1 + S_2$	Samples[B]	72	Rows	2			$\xi_{BS} + \eta_R$
$\Psi(C) = S_1$			Columns	2			$\xi_{BS} + \eta_C$
$\Psi(P_1) = B_2$			R#C	4	V_3	2	$\xi_{BS} + \eta_{RC} + q(V_3)$
$\Psi(P_2) = S_2$					Residual	2	$\xi_{BS} + \eta_{RC}$
			Plots[R,C]$_2$	64	$V \vdash V_1 \vdash V_3$	22	$\xi_{BS} + \eta_{RCP} + q(V \vdash V_1 \vdash V_3)$
					Residual	42	$\xi_{BS} + \eta_{RCP}$

3 Generalizations

One possible generalization increases the number of phases. If, in each phase i, material from the units in phase $i-1$ is allocated to the new experimental units using a design key but no further treatments are applied, then the foregoing ideas can be applied recursively, and no new concepts are involved.

A different generalization keeps two phases but has two sets of treatments Γ_1 and Γ_2 defined by sets $\mathscr{F}_1$ and $\mathscr{F}_2$ of factors and pseudofactors. In the first phase, treatments in Γ_1 are allocated to units in Ω_1 by a design key Φ. In the second phase,

not only is produce from units in Ω_1 allocated to units in Ω_2 by a design key Ψ, but treatments in Γ_2 are also allocated to units in Ω_2 by a design key Θ. Then treatment effects from Γ_1 and Γ_2 may be either totally confounded or orthogonal, depending on their confounding with effects of second-phase units, and it is also possible to identify interactions between Γ_1 and Γ_2.

The third generalization permits more than one prime to be involved. Unlike [13], we split each design key into separate keys, one for each prime, so that the algorithm for identifying confounding in Sect. 1 still works for each prime separately. Suppose that the primes include $p_1, \ldots, p_k$. For $i = 1, \ldots, k$, let T_i be a linear combination of treatment factors or pseudofactors with p_i levels. Then T_i belongs to an effect defined by a subset $\mathscr{S}_i$ of the initial letters of the genuine treatment factors. It can be shown the $\prod_{i=1}^{k}(p_i - 1)$ degrees of freedom for the interaction between $T_1, \ldots, T_k$ all belong to the effect defined by the subset $\mathscr{S}_1 \cup \cdots \cup \mathscr{S}_k$.

Example 5 (Potato storage) In the first phase of an experiment described in [3], all combinations of four potato Cultivars and three Fungicides are grown in a complete-block design with three Blocks of twelve Units each. In the second phase, potatoes are harvested and stored on Pallets on Shelves, there being twelve pallets on each of twelve shelves. Potatoes from each unit in the first phase are put onto four different pallets on the same shelf, where they are stored for four different lengths of Time. See the panel diagrams in Fig. 4.

The left-hand side of Table 4 shows one way of allocating pseudofactors to the two primes involved. Then the design in [3] can be made with the design keys in the right-hand side of Table 4, where the subscripts show the primes. Now $\Psi_2(\Phi_2(C_i)) = S_i$ for $i = 1, 2$, so Cultivars are confounded with Shelves but orthogonal to Blocks, which are also confounded with Shelves. Also, $\Psi_3(\Phi_3(F)) = P_3$ and $\Theta_2(T_i) = P_i$ for $i = 1, 2$: therefore Fungicides, Times and their interaction are confounded with Pallets within Shelves and are also orthogonal to Blocks. Continuing like this, the skeleton ANOVA table can easily be calculated.

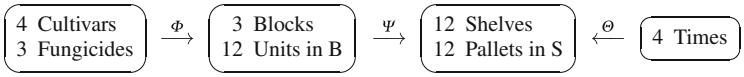

Fig. 4 Panel diagrams and design keys in Example 5

Table 4 Factors, pseudofactors and design keys in Example 5

Prime	Γ_1	Ω_1	Ω_2	Γ_2	Φ	Ψ	Θ
2	C_1, C_2	U_1, U_2	S_1, S_2, P_1, P_2	T_1, T_2	$\Phi_2(C_1) = U_1$	$\Psi_2(U_1) = S_1$	$\Theta_2(T_1) = P_1$
					$\Phi_2(C_2) = U_2$	$\Psi_2(U_2) = S_2$	$\Theta_2(T_2) = P_2$
3	F	B, U_3	S_3, P_3		$\Phi_3(F) = U_3$	$\Psi_3(B) = S_3$	
						$\Psi_3(U_3) = P_3$	

References

1. Bailey, R.A.: Patterns of confounding in factorial designs. Biometrika **64**, 597–603 (1977)
2. Bailey, R.A., Brien, C.J.: Randomization-based models for multitiered experiments: I. A chain of randomizations. Ann. Stat. (2016) (in press)
3. Brien, C.J., Bailey, R.A.: Multiple randomizations. J. R. Stat. Soc. Ser. B **68**, 571–609 (2006)
4. Brien, C.J., Bailey, R.A.: Decomposition tables for experiments I. A chain of randomizations. Ann. Stat. **37**, 4184–4213 (2009)
5. Cheng, C.-S.: Theory of Factorial Design. Chapman and Hall, Boca Raton (2014)
6. Cheng, C.-S., Tsai, P.-W.: Templates for design key construction. Stat. Sin. **23**, 1419–1436 (2013)
7. Fisher, R.A.: The theory of confounding in factorial experiments in relation to the theory of groups. Ann. Eugen. **11**, 341–353 (1942)
8. Kobilinsky, A.: Confounding in relation to duality of finite Abelian groups. Linear Algebra Appl. **70**, 321–347 (1985)
9. Kobilinsky, A., Bouvier, A., Monod, H. **planor**: Generation of regular factorial designs. R package version 0.2–3. http://CRAN.R-project.org/package=planor (2015)
10. Kobilinsky, A., Monod, H.: Experimental design generated by group morphisms: an introduction. Scand. J. Stat. **18**, 119–134 (1991)
11. Kobilinsky, A., Monod, H., Bailey, R.A.: Automatic generation of generalised regular factorial designs (2015, submitted for publication)
12. Patterson, H.D.: The factorial combination of treatments in rotation experiments. J. Agric. Sci. **65**, 171–182 (1965)
13. Patterson, H.D.: Generation of factorial designs. J. R. Stat. Soc. Ser. B **38**, 175–179 (1976)
14. Patterson, H.D., Bailey, R.A.: Design keys for factorial experiments. J. R. Stat. Soc. Ser. C **27**, 335–343 (1978)

On Designs for Recursive Least Squares Residuals to Detect Alternatives

Wolfgang Bischoff

Abstract Linear regression models are checked by a lack-of-fit (LOF) test to be sure that the model is at least approximatively true. In many practical cases data are sampled sequentially. Such a situation appears in industrial production when goods are produced one after the other. So it is of some interest to check the regression model sequentially. This can be done by recursive least squares residuals. A sequential LOF test can be based on the recursive residual partial sum process. In this paper we state the limit of the partial sum process of a triangular array of recursive residuals given a constant regression model when the number of observations goes to infinity. Furthermore, we state the corresponding limit process for local alternatives. For specific alternatives designs are determined dominating other designs in respect of power of the sequential LOF test described above. In this context a result is given in which e^{-1} plays a crucial role.

1 Introduction

In order to guarantee the quality of each delivery of contract goods, companies take samples to decide whether the quality is or is not constant. If the goods are sequentially produced this problem can be modelled by the regression model

$$Y(t) = g(t) + \epsilon(t), \quad t \in [0, 1], \tag{1}$$

where g is the true, but unknown mean function of the quality, $\epsilon(t)$ is a real random variable with expectation 0 and variance $\sigma^2 > 0$, and $[0, 1]$ is the period of production. Since our results keep true when σ^2 is replaced by a consistent estimator for σ^2 we can put $\sigma^2 = 1$ without loss of generality.

We consider the problem more generally. We want to test if the model (1) is a linear model with respect to known and linearly independent functions $f_1, \ldots, f_d$, i.e. if there exist suitable constants $\beta_1, \ldots, \beta_d \in \mathbb{R}$ such that $g(t) = \sum_{i=1}^{d} \beta_i f_i(t), t \in$

W. Bischoff (✉)
Mathematisch-Geographische Fakultaet, Katholische Universitaet Eichstaett-Ingolstadt,
D-85071 Eichstaett, Germany
e-mail: wolfgang.bischoff@ku.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_5

[0, 1]. Hence, we look for a test of the null hypothesis

$$H_0 : g = \sum_{i=1}^{d} \beta_i f_i = f^\top \beta \text{ for some } \beta = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{R}^d, \tag{2}$$

where $f^\top = (f_1, \ldots, f_d)$, against the alternative that model (1) is not a linear model with respect to $f_1, \ldots, f_d$, that is (2) is not fulfilled.

In quality control we are interested in testing

$$H_0 : g(t) = \beta = \mathbf{1}_{[0,1]}(t)\beta, \quad t \in [0, 1], \ \beta \in \mathbb{R} \text{ unknown constant}, \tag{3}$$

where $\mathbf{1}_{[0,1]}$ is the function identically 1 on [0, 1], against the alternative

$$K : g \neq \textit{constant}. \tag{4}$$

Such a function g under the alternative has typically the following form:

$$g(t) = \beta \text{ for } t \in [0, t_0], \quad g \text{ increasing or decreasing for } t \in (t_0, 1].$$

This form of g means that the quality keeps constant up to a fixed, known or unknown change–point $t_0 \in (0, 1)$, then the quality is getting worse or is getting better.

In the literature on 'detecting changes' in regression models, it is common to consider (recursive) residual partial sum processes or variants thereof; see for instance, Gardner [8], Brown, Durbin and Evans [5], Sen and Srivastava [17], MacNeill [13, 14], Sen [16], Jandhyala and MacNeill [9–11], Watson [19], Bischoff [1], Jandhyala, Zacks and El-Shaarawi [12], Bischoff and Miller [3], Xie and MacNeill [20], Bischoff and Somayasa [4], Bischoff and Gegg [2]. The asymptotics of the partial sum of recursive residuals is investigated by Sen [16] only. Sen [16], however, assumed a time series sampling for his asymptotic result. For our problem we need a triangular array approach.

In Sect. 2 we discuss some asymptotic results for the partial sum process of recursive least squares residuals. Assuming a constant regression model we state such a result for a triangular array of design points on one hand if the null hypothesis (3) is true and on the other hand under certain assumptions if a local alternative (4) is true. With this we are in the position to establish an asymptotic size α test to test the null hypothesis (3). Furthermore, in Sect. 3, we can discuss the power of this test for certain alternatives. There, we determine designs that have uniformly more power than other designs. For one of the results e^{-1} occurs as crucial constant.

2 Recursive Residuals

Recursive (least squares) residuals were described in Brown et al. [5], for some history see Farebrother [7]. Brown et al. considered recursive residuals for a linear regression model given a time series sampling. To this end let $t_1 < t_2 < \dots$ be a sequence of (design) points (in time), let $\varepsilon_1, \varepsilon_2, \dots$ be iid real random variables with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = 1$, let $n \in \mathbb{N}, n > d$, be the number of observations where d is the number of known regression functions. Moreover, we put

$$X_i = (f(t_1), \dots, f(t_i))^\top,\ d \le i \le n,$$

where $f = (f_1, \dots, f_d)^\top$. So we get $n - d + 1$ linear models, namely for the first i, $d \le i \le n$, observations each

$$\mathbf{Y}_i = (Y_1, \dots, Y_i)^\top = X_i\boldsymbol{\beta} + (\epsilon_1, \dots, \epsilon_i)^\top,\ d \le i \le n.$$

Let $t_1, \dots, t_d$ be chosen in such a way that $rank(X_d) = d$. Then, for each $i \ge d$ we estimate β by the least squares estimate $\hat{\beta}_i$ using the first i observations $\mathbf{Y}_i$. Now we can define $n - d$ recursive residuals

$$e_i = \frac{Y_i - f(t_i)^\top \hat{\beta}_{i-1}}{\left(1 + f(t_i)^\top (X_{i-1}^\top X_{i-1})^{-1} f(t_i)\right)^{1/2}},\ i = d+1, \dots, n.$$

To state Sen's and our result it is convenient to define the partial sum operator $T_n : \mathbb{R}^n \longrightarrow C[0,1], \mathbf{a} = (a_1, \dots, a_n)^\top \mapsto T_n(\mathbf{a})(z),\ z \in [0,1]$, where

$$T_n(\mathbf{a})(z) = \sum_{i=1}^{[nz]} a_i + (nz - [nz])a_{[nz]+1},\ z \in [0,1].$$

Here we used $[s] = \max\{n \in \mathbb{N}_0 \mid n \le s\}$ and $\sum_{i=1}^{0} a_i = 0$. Let $\mathbf{a} = (a_1, \dots, a_n)^\top \in \mathbb{R}^n, b_i = a_1 + \dots + a_i, i = 1, \dots, n$, then the function $T_n(\mathbf{a})(\cdot)$ is shown in Fig. 1.

By Donsker's Theorem the stochastic process $\frac{1}{\sqrt{n}} T_n((\boldsymbol{\epsilon_1}, \dots, \boldsymbol{\epsilon}_n)^\top)$ converges weakly to Brownian motion B for $n \to \infty$. For recursive residuals Sen [16] proved the following result.

Theorem 1 (Sen [16]) *For the regression model given in (2) let* $\mathbf{e}_n = (e_{d+1}, \dots, e_n)^\top$. *If* H_0 *given in (2) is true, then under certain assumptions*

$$\frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}_n)(z) \text{ converges weakly to } B(z),\ z \in [0,1], \text{ for } n \to \infty.$$

Sen, however, could not determine the limit process for a local alternative.

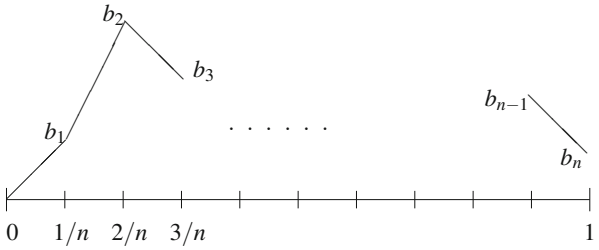

Fig. 1 The function $T_n(\mathbf{a})(\cdot)$

The time series sampling approach described above cannot be applied to problems of experimental design. Instead we need the asymptotic result for a triangular array of design points under the null hypothesis and under local alternatives. To this end let $n_0 \in \mathbb{N}$, $n_0 > d$, be the number of observations. We assume that the data are taken at the design points $0 \le t_{n_0 1} \le t_{n_0 2} \le \ldots \le t_{n_0 n_0} \le 1$. These design points can be embedded in a triangular array of design points: $0 \le t_{n1} \le t_{n2} \le \ldots \le t_{nn} \le 1,\ n \in \mathbb{N}$. Furthermore, let $\varepsilon_{n1}, \ldots, \varepsilon_{nn}, n \in \mathbb{N}$, be a triangular array of real random variables where $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ are iid with $E(\varepsilon_{ni}) = 0$ and $Var(\varepsilon_{ni}) = 1$ for each $n \in \mathbb{N}$. Accordingly, we get a corresponding triangular array of observations for model (1):

$$Y_{nj} = g(t_{nj}) + \epsilon_{nj},\ 1 \le j \le n,\ n \in \mathbb{N}.$$

We put

$$\varepsilon_i^n = (\varepsilon_{n1}, \ldots, \varepsilon_{ni})^\top,\ d \le i \le n,$$
$$X_i^n = (f(t_{n1}), \ldots, f(t_{ni}))^\top,\ d \le i \le n.$$

So we get $n-d+1$ linear models under the null hypothesis H_0 given in (2), namely for the first i, $d \le i \le n$, observations each

$$\mathbf{Y}_i^n = (Y_{n1}, \ldots, Y_{ni})^\top = X_i^n \beta + \varepsilon_i^n.$$

Let $t_{n1}, \ldots, t_{nd}$ for all $n \ge d$ be chosen in such a way that $rank(X_d^n) = d$. Moreover, let $\hat{\beta}_i^n, d \le i$, be the least squares estimate for β using the first i observations $\mathbf{Y}_i^n$. Then the $n-d$ recursive least squares residuals for the triangular array are defined by

$$e_{ni} = \frac{Y_{ni} - f(t_{ni})^\top \hat{\beta}_{i-1}^n}{\left(1 + f(t_{ni})^\top (X_{ni-1}^\top X_{ni-1})^{-1} f(t_{ni})\right)^{1/2}},\ i = d+1, \ldots, n.$$

Assuming the constant regression model the next result states the limit of the recursive residual partial sum process if H_0, see (3), is true and if a local alternative is true. When (3) is true, the location of the design points has no influence. So the result is true for any triangular array of design points. If a local alternative is true, we give the result for a uniform array of design points only to avoid further technical notation. This result will be sufficient for our purposes below.

Theorem 2 (Master Thesis Rabovski [15] under the supervision of the author and Frank Miller) *For the constant regression model let* $\mathbf{e}^n = (e_{nd+1}, \ldots, e_{nn})^\top$ *be the vector of the* $n-d$ *recursive residuals of a triangular array of design points.*

(a) If H_0 *given in (3) is true, then for any triangular array of design points*

$$\frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}^n)(z) \text{ converges weakly to } B(z),\ z \in [0,1], \text{ for } n \to \infty.$$

(b) Let $g : [0,1] \to \mathbb{R}$, $g \neq$ *constant, have bounded variation and let the triangular array of design points be given by a uniform design*

$$t_{n1} = 0, t_{n2} = \frac{1}{n-1}, t_{n3} = \frac{2}{n-1}, \ldots, t_{nn} = 1, n \in \mathbb{N}.$$

Then, if the local alternative $\frac{1}{\sqrt{n-d}} g$ *is true,*

$$\frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}^n)(z) \text{ converges weakly to } h(z) + B(z),\ z \in [0,1], \text{ for } n \to \infty,$$

where

$$h(z) = \int_0^z g(t)dt - \int_0^z \frac{1}{s} \int_0^s g(t)dtds,\ z \in [0,1].$$

Theorem 2 part (a) can be used to establish an asymptotic size α test of Kolmogorov(-Smirnov) or Cramér-von Mises type. As an example we state a one-sided test of Kolmogorov type, to detect a negative deviation h from Brownian motion.

Theorem 3 *For the constant regression model let* $\mathbf{e}^n = (e_{nd+1}, \ldots, e_{nn})^\top$ *be the vector of the* $n-d$ *recursive residuals of an arbitrary triangular array of design points. Then an asymptotic size* α *test is given by*

$$\text{reject } H_0 \text{ given in (3)} \iff \exists t \in [0,1] : \frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}^n)(t) < \Phi^{-1}(\frac{\alpha}{2}),$$

where $\Phi^{-1}(\frac{\alpha}{2})$ *is the* $\alpha/2$ *quantile of the standard normal distribution.*

Proof Since $P(\exists t \in [0,1] : B(t) < \Phi^{-1}(\frac{\alpha}{2})) = \alpha$, see, for instance, Shorack [18] p.314, the theorem is proved.

The above test is not constructed sequentially. The test statistic $\frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}^n)(t)$ can be calculated sequentially for each new recursive residual e_{ni} and so the null hypothesis can be rejected as soon as the test statistic is less than $\Phi^{-1}(\frac{\alpha}{2})$.

3 Designs for Detecting Alternatives

We look for designs being useful for the quality problem discussed in the introduction. Therefore we consider the constant regression model. Usually in the context of quality control certain properties of the alternative are often known.

We begin with the alternative

$$g_{t_0}(t) = g_{t_0;c_0,c_1}(t) = c_0 \mathbf{1}_{[0,t_0]}(t) + c_1 \mathbf{1}_{(t_0,1]}(t), \ t \in [0,1], \tag{5}$$

where $c_0, c_1 \in \mathbb{R}$ are unknown constants and the change-point $t_0 \in (0,1)$ is a known or unknown constant. We assume $c_0 > c_1$ to get a negative trend h, see Theorem 4. (Note that Theorem 3 states a test for detecting negative trends h.) Let a triangular array of design points be given with

$$q := \lim_{n\to\infty} \frac{\text{number of } \{t_{ni} | t_{ni} \le t_0, 1 \le i \le n\}}{n} \in (0,1).$$

We call such a triangular array of design points an asymptotic q-design. The proof of the following result is given in the next section.

Theorem 4 *For a constant regression model let* $\mathbf{e}^n = (e_{nd+1}, \ldots, e_{nn})^\top$ *be the vector of the* $n-d$ *recursive residuals of a triangular array of design points forming an asymptotic* q*-design. Let the alternative* g_{t_0} *given in (5) be true. Then, we have for the local alternative* $\frac{1}{\sqrt{n-d}} g_{t_0}$*:*

$$\frac{1}{\sqrt{n-d}} T_{n-d}(\mathbf{e}^n)(z) \text{ converges weakly to } h(z) + B(z), \ z \in [0,1], \text{ for } n \to \infty,$$

where

$$h(z) = q(c_1 - c_0)(\ln(z) - \ln(q)) \mathbf{1}_{(q,1]}(z), \ z \in [0,1].$$

For an asymptotic q-design the power of the test given in Theorem 3 with respect to the alternative (5) is given by

$$P\left(\exists z \in [0,1] : B(z) - q(c_0 - c_1)(\ln(z) - \ln(q))\mathbf{1}_{(q,1]}(z) \le \Phi^{-1}(\frac{\alpha}{2})\right).$$

Therefore, we call an asymptotic q^*-design uniformly better than an asymptotic q-design if for all $z \in (0,1]$

$$-q^*(\ln(z) - \ln(q^*))\mathbf{1}_{(q^*,1]}(z) \le -q\big(\ln(z) - \ln(q)\big)\mathbf{1}_{(q,1]}(z) \qquad (6)$$
$$\text{with } ' <' \text{ at least for one } z \in [0,1].$$

The proof of the following result is given in the next section.

Theorem 5 *Let the situation considered in Theorem 4 be given and let $q_1, q_2 \in [e^{-1}, 1)$. Then an asymptotic q_1-design is uniformly better than an asymptotic q_2-design, if $q_1 < q_2$.*

The author does not know whether the result stated above has some relation to the famous e^{-1}-law for the best choice problem, see, for instance, Bruss [6]. For n_0 design points we consider the design d^* with the fractional part of about e^{-1} design points as near as possible at 0 (let t_1^* be the largest of these design points) and the fractional part of about $1 - e^{-1}$ design points as near as possible at 1 (let t_2^* be the smallest of these design points). Then, by Theorem 5, d^* is asymptotically the uniformly best applicable design, if $t_1^* < t_2^*$ and $t_1^* \le t_0$.

Finally, we consider the alternative

$$g(t) = c_0\mathbf{1}_{[0,t_0]}(t) + (c_0 + c_1t_0 - c_1t)\mathbf{1}_{(t_0,1]}(t),\ t \in [0,1], \qquad (7)$$

where $t_0 \in [0,1)$ is a known or unknown constant and $c_0 \in \mathbb{R}, c_1 \in (0,\infty)$ are unknown constants. The last result follows in an analogous way as above.

Theorem 6 *The asymptotic 0-design which is uniformly distributed on $[t_0, 1]$ is uniformly better with respect to the alternative (7) than an asymptotic q-design, $q \in (0,1)$, whose fractional part of design points on $[t_0, 1]$ is uniformly distributed.*

For an unknown change-point t_0 the above result is of theoretical interest only.

4 Some Proofs

The following relation between an arbitrary design and a uniform design is crucial for the next proof. To this end let the alternative (5) and an arbitrary triangular array of design points $t_{n1}, \ldots, t_{nn}$ with $0 \le t_{n1} \le \ldots \le t_{ns} \le t_0 < t_{ns+1} \le \ldots \le t_{nn} \le 1$

be given. Moreover, let $q := s/n$. Then we have

$$g_{t_0}(t_{ni}) = g_q\left(\frac{i-1}{n-1}\right),\ i = 1, \ldots, n.$$

Thus instead of analyzing the alternative g_{t_0} and an arbitrary design with s design points equal to or less than t_0, we can analyze the alternative g_q with the change-point $q = s/n$ and a uniform design.

Proof (of Theorem 4) By the above considerations the limit process of the recursive residual partial sum process with respect to the local alternative $\frac{1}{\sqrt{n-d}}g_q$ and a uniform design coincides with the limit process with respect to the local alternative $\frac{1}{\sqrt{n-d}}g_{t_0}$ and an asymptotic q-design. The trend h given in Theorem 2 part (b) can be obtained for the local alternative $\frac{1}{\sqrt{n-d}}g_q$ and a uniform design after some calculations:

$$h(z) = 0, \qquad z \in [0, q],$$
$$h(z) = \int_q^z c_1 - \tfrac{1}{s}(qc_0 + (s-q)c_1)ds = q(c_1 - c_0)(\ln(z) - \ln(q)), \quad z \in (q, 1].$$

Proof (of Theorem 5) For $z = 1$ the expression $-q\big(\ln(z) - \ln(q)\big)\mathbf{1}_{(q,1]}(z)$ considered in (6) takes on its minimum for $q = e^{-1}$ and, furthermore, it is strictly increasing on $[e^{-1}, 1)$.

Let $e^{-1} \le q_1 < q_2 < 1$. Then we have for all $z \in (q_2, 1]$

$$\frac{d}{dz}\Big(-q_1\big(\ln(z) - \ln(q_1)\big)\Big) = -\frac{q_1}{z} > -\frac{q_2}{z} = \frac{d}{dz}\Big(-q\big(\ln(z) - \ln(q)\big)\Big).$$

This together with the first result of the proof justifies the statement of Theorem 5.

References

1. Bischoff, W.: A functional central limit theorem for regression models. Ann. Stat. **26**, 1398–1410 (1998)
2. Bischoff, W., Gegg, A.: Partial sum process to check regression models with multiple correlated response: with an application for testing a change point in profile data. J. Multivar. Anal. **102**, 281–291 (2011)
3. Bischoff, W., Miller, F.: Asymptotically optimal tests and optimal designs for testing the mean in regression models with applications to change-point problems. Ann. Inst. Stat. Math. **52**, 658–679 (2000)
4. Bischoff, W., Somayasa, W.: The limit of the partial sums process of spatial least squares residuals. J. Multivar. Anal. **100**, 2167–2177 (2009)
5. Brown, R.L., Durbin, J., Evans, J.M.: Techniques for testing the constancy of regression relationships over time. J. R. Stat. Soc. Ser. B **37**, 149–192 (1975)
6. Bruss, F.T.: A unified approach to a class of best choice problems with an unknown number of options. Ann. Probab. **12**, 882–889 (1984)

7. Farebrother, R.W.: An historical note on recursive residuals. J. R. Stat. Soc. Ser. B **40**, 373–375 (1978)
8. Gardner, L.A.: On detecting changes in the mean of normal variates. Ann. Math. Stat. **40**, 116–126 (1969)
9. Jandhyala, V.K., MacNeill, I.B.: Residual partial sum limit process for regression models with applications to detecting parameter changes at unknown times. Stoch. Process. Appl. **33**, 309–323 (1989)
10. Jandhyala, V.K., MacNeill, I.B.: Tests for parameter changes at unknown times in linear regression models. J. Stat. Plan. Inference **27**, 291–316 (1991)
11. Jandhyala, V.K., MacNeill, I.B.: Iterated partial sum sequences of regression residuals and tests for changepoints with continuity constraints. J. R. Stat. Soc. Ser. B **59**, 147–156 (1997)
12. Jandhyala, V.K., Zacks, S., El-Shaarawi, A.H.: Change–point methods and their applications: contributions of Ian MacNeill. Environmetrics **10**, 657–676 (1999)
13. MacNeill, I.B.: Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. Ann. Stat. **6**, 422–433 (1978)
14. MacNeill, I.B.: Limit processes for sequences of partial sums of regression residuals. Ann. Probab. **6**, 695–698 (1978)
15. Rabovski, O.: Asymptotische Tests basierend auf rekursiven Residuen von Regressionsmodellen, Diplomarbeit (2003)
16. Sen, P.K.: Invariance principles for recursive residuals. Ann. Stat. **10**, 307–312 (1982)
17. Sen, A., Srivastava, M.S.: On tests for detecting change in mean when variance is unknown. Ann. Inst. Stat. Math. **27**, 479–486 (1975)
18. Shorack, G.R.: Probability for Statisticans. Springer, New York [u.a.] (2000)
19. Watson, G.S.: Detecting a change in the intercept in multiple regression. Stat. Probab. Lett. **23**, 69–72 (1995)
20. Xie, L., MacNeill, I.B.: Spatial residual processes and boundary detection. S. Afr. Stat. J. **40**, 33–53 (2006)

A Multi-objective Bayesian Sequential Design Based on Pareto Optimality

Matteo Borrotti and Antonio Pievatolo

Abstract Complexity arises in different fields of application. The increasing number of variables and system responses used to describe an experimental problem limits the applicability of classical approaches from Design of Experiments (DOE) and Sequential Experimental Design (SED). In this situation, more effort should be put into developing methodological approaches for complex multi-response experimental problems. In this work, we develop a novel design technique based on the incorporation of the Pareto optimality concept into the Bayesian sequential design framework. One of the crucial aspects of the approach involves the selection method of the next design points based on current information and the chosen system responses. The novel sequential approach has been tested on a simulated case study.

1 Introduction

Current research and development at both academic and industrial level is tackling the design and characterisation of sophisticated systems, where the presence of a high number of factors and variables limits the complete experimental screening towards actual optimisation. A possible example is the separation process for recycling waste electrical and electronic equipments (WEEE) [4]. In order to plan the process operating conditions (i.e., variables) for a given multi-material mixture under treatment, several technological and economic criteria should be jointly considered. Therefore, the outcomes (i.e., system responses) of the process to be considered in order to find the best process operating conditions are more than one, leading to a multi-objective problem. Usually, a multi-objective problem involves minimising or maximising multiple system responses subject to a set of constraints. Examples can be found in all applications where an optimal solution (or a set of optimal solutions) is sought with trade-offs between two or more conflicting system responses.

M. Borrotti (✉) • A. Pievatolo
CNR-IMATI, Via Bassini 15, 20133 Milan, Italy
e-mail: matteo.borrotti@mi.imati.cnr.it; antonio.pievatolo@mi.imati.cnr.it

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_6

DOE is one of the most important tools in scientific research, and applications can be found in all industries. In fact, DOE enables investigators to conduct better experiments and collect data efficiently. However, current DOE techniques are based usually on linear models and one response variable, limiting their applicabilty in multi-objective problems. Few examples of multi-objective approaches used in DOE can be found in literature for optimising different optimal criteria simultaneously [5, 7, 8].

In this work, we propose a Bayesian sequential DOE approach for multi-objective designs. The novelty resides in the combination of the Bayesian paradigm and the unprocessed multiple responses. The classical theory of optimal experimental design has been completely transferred into the Bayesian world, where possible. For example D-optimal, I-optimal, A-optimal designs, and others, all have their Bayesian counterpart [2]. As far as Bayesian design is concerned, the latest developments involve sequential methods, with new methodology mainly motivated by clinical trials (adaptive design) [3], whereas frequent applications are motivated by specific industrial problems [6].

Our approach has been tested in a simulated example related to the optimisation of electrostatic separation processes in recycling [1], showing promising initial results.

2 Some Concepts from Multi-objective Optimisation

Multiple objectives in design optimisation pose the problem of identifying the trade-offs between competing system responses. The result of multi-objective optimisation is no longer a single optimal design point (i.e., optimal combination of operating conditions), as in single-objective optimisation, but rather a set of design points that are not worse than any other design point and strictly better in at least one of the system responses.

More precisely, for the bi-objective optimisation problem of maximising two objectives, candidate design points are evaluated according to an objective function vector $\phi = (\phi_1, \phi_2)$. ϕ_1 determines the value for the first system response (y_1) given a specific design point and ϕ_2 the value for the second system response (y_2). Given two design points s and s^*, we say that s dominates s^* iff $\phi_1(s) \geq \phi_1(s^*)$ and $\phi_2(s) \geq \phi_2(s^*)$. If no s^* exists such that s^* dominates s, the design point s is called Pareto-optimal. In this context, the goal of the optimisation is to determine (or approximate) the set of all Pareto-optimal design points, whose image in the bi-objective space is called the Pareto front [5]. In Fig. 1, an example of a Pareto front is shown.

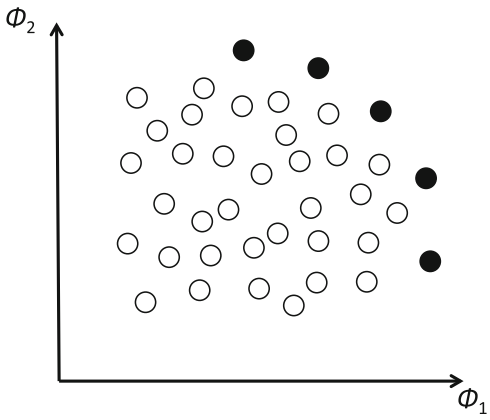

Fig. 1 Example of Pareto front (*black dots*) in which both objectives have to be maximised

3 The Multi-objective Bayesian Sequential DOE Approach

The main goal of our approach is to sequentially design an experiment in order to find the best compromise between d competitive system responses. More precisely, in this first attempt to develop a multi-objective Bayesian sequential DOE approach, we consider a general problem described by k experimental variables, $\mathbf{x} = (x_1, \ldots, x_k)$, and d objective functions, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_d)$, each of which is used to calculate the respective system response forming the vector $\mathbf{y} = (y_1, \ldots, y_d)$. The ultimate aim is to simultaneously maximise all the d objective functions. In order not to not put emphasis on one special objective function, φ_i, while doing multi-objective optimization, it may be necessary to normalise the system response values. In our example they all lie in the interval [0, 1] so no further action is required.

In this setting, the proposed approach can be described as follows:

(1) Select a first initial design matrix X composed by n design points and evaluate them. In this first attempt, the initial set of design points is randomly selected. Randomness allows the exploration of the space in areas not anticipated by prior knowledge but where interesting new effects may possibly arise.
(2) Create the matrix K composed by the initial design matrix X and the related matrix of system responses Y. At this step, matrix K has n rows and $k + d$ columns.
(3) Calculate the d Bayesian one-dimensional utility functions (one for each system response) for the whole search space, based on the marginal predictive distribution of the response conditional on K.
(4) Identify the Pareto front based on the d Bayesian one-dimensional utilities.
(5) Select the design point with the minimun Euclidean distance from the *utopia* point [5]. For instance, the utopia point is the ideal point in a bi-objective space with the maximum value of the two system responses in the Pareto front.
(6) Evaluate the selected design point and add it to the matrix K with the related vector of system responses.

(7) Repeat steps 3, 4, 5 and 6 until a certain stopping criterion is satisfied. A usual technique to determine a stopping criterion is to calculate the maximum number of design points allowed by the available budget.

3.1 The Proposed Bayesian Utility Function

We will refer to $U^i(\mathbf{x}_{n+1}|X)$ as the i-th component of the multidimensional Bayesian utility function, where $i = 1, \ldots, d$ and $\mathbf{x}_{n+1}$ is the feasible design point in the design space that may be chosen for the next experimental step.

The use of $U^i(\mathbf{x}_{n+1}|X)$ has been inspired by the work of Verdinelli and Kadane [9].

$$U^i(\mathbf{x}_{n+1}|X) = \int [\alpha\ y_{n+1,i} + \beta\ ln\ p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})] \times \\ \times p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1}) d\mathbf{y}_{n+1} d\theta,$$

where $(y)_{1:(n+1)}$ collects $n+1$ d-dimensional system responses and θ parameterises the likelihood. Furthermore, α and β are non-negative weights, which express the relative contribution that the experimenter is willing to attach to the two components of U^i. The number of system responses, d, is not higher than the dimension of $\mathbf{y}$, which is the vector of response values. Generally, the posterior distribution of θ depends on all y.

The i-th utility function can be expanded as follows:

$$\begin{aligned} U^i(\mathbf{x}_{n+1}|X) &= \alpha \int\ y_{n+1,i}\ p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i} + \\ &\quad + \beta\ ln\ p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})\ \times \\ &\quad \times p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i}\ d\theta \\ &= \alpha\ E[y_{n+1,i}|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1}]\ + \\ &\quad + \beta \int\ ln\ p(\theta|\mathbf{y}_{1:(n+1),i}, X, \mathbf{x}_{n+1})\ \times \\ &\quad \times p(y_{n+1,i}, \theta|\mathbf{y}_{1:n,i}, X, \mathbf{x}_{n+1})\ d\mathbf{y}_{n+1,i}\ d\theta. \end{aligned}$$

The second term is related to the expected gain in Shannon information, obtained from adding the new design point to X, conditional on $\mathbf{y}_{1:n}$. The simplest model for $\mathbf{y}_{1:n,i}$ is the following:

$$\mathbf{y}_{1:n,i} = X\ \theta_i\ +\ \boldsymbol{\epsilon}_{1:n,i}, \tag{1}$$

where $\theta_i \sim \mathcal{N}(\theta_0, \sigma^2 R_0^{-1})$, σ^2 is known and $\boldsymbol{\epsilon}_{1:n,i}$ is a vector of *iid* $\mathcal{N}(0, \sigma^2)$ random variables.

We know there are d system responses, each with its own U^i. However, without loss of generality, we can consider just one system response in the following steps. We then obtain:

$$y_{1:n} = X\,\theta \;+\; \epsilon_{1:n}. \tag{2}$$

Notice that the entries of Y are conditionally independent given θ.

We can now calculate the joint distribution of y_{n+1} and θ:

$$p(y_{n+1}, \theta | \mathbf{y}_{1:n}, X, \mathbf{x}_{n+1}) \;=\; p(y_{n+1}|\theta, \mathbf{x}_{n+1})\; p(\theta|\mathbf{y}_{1:n}, X), \tag{3}$$

and

$$\theta|\mathbf{y}_{1:n}, X \;\sim\; \mathcal{N}\{(X^T X + R_0)^{-1}\;(X^T \mathbf{y}_{1:n} + R_0\theta_0), \sigma^2\;(X^T X + R_0)^{-1}\}, \tag{4}$$

which is the posterior distribution after observing $\mathbf{y}_{1:n}$.

From Eqs. 3 and 4, we obtain the two terms of the utility function. The factor multiplying β is:

$$\begin{aligned}\int\; & ln\; p(\theta|\mathbf{y}_{1:(n+1)}, X, \mathbf{x}_{n+1}) \quad\; p(y_{n+1}, \theta|\mathbf{y}_{1:n}, X, \mathbf{x}_{n+1})\; dy_{n+1}\; d\theta \\ & = -\frac{k}{2}\; ln(2\pi) \;-\; \frac{k}{2} + \frac{1}{2}\; ln\; det\{\sigma^{-2}\;(\mathbf{x}_{(n+1)}\mathbf{x}_{(n+1)}^T \;+\; R)\},\end{aligned}$$

where $R \;=\; (X^T X + R_0)$.

The predictive mean multiplying α is:

$$\begin{aligned} E(y_{n+1}|\mathbf{y}_{1:n}) &= E(\theta^T x_{n+1} + \epsilon_{n+1} \mid \mathbf{y}_{1:n}) \\ &= E(\theta \mid \mathbf{y}_{1:n})^T\; x_{n+1} \\ &= [\{(X^T X) + R_0\}^{-1}\;(X^T \mathbf{y}_{1:n} + R_0\theta_0)]^T\; x_{n+1}. \end{aligned}$$

4 Simulated Case Study

Our simulated case study is based on the work of Borrotti et al. [1]. The authors consider the problem of separating metal and nonmetal particles derived from WEEE by using the Corona electrostatic separation (CES) process. CES is mainly used to separate conductors from insulators, like copper from plastics, in shredded multi-material mixtures. Briefly, in a CES machine, particles are transported by a feeder on a rotating drum, and are charged as they pass through an electrostatic field.

At this point, particles receive a discharge of electricity, which gives the nonmetals a high surface charge, causing them to be attracted to the rotor surface until they are brushed down into a bin, ideally on one side of the machine. Metal particles do not become charged, as the charge rapidly dissipates through the particles to the earthed rotor, so they fall into the bin on the other side of the machine under the effect of the centrifugal and gravitational forces. Ideally, if the separation process were perfectly accurate, all the particles would be correctly classified. However, because of the presence of mixed non-liberated particles and due to random particle impacts, a non-classified material flow is generated (middlings) consisting of particles that drop into a central bin.

A typical industrial CES machine with fixed design parameters depends on the following controllable parameters (i.e., variables):

(1) Electrostatic potential (x_1), or voltage, which ranges between -35.000 to -25.000 V;
(2) Drum speed (x_2), or simply speed, which ranges between 32 to 128 rpm;
(3) Feed rate (x_3), which ranges between 0.0028 to 0.028 kilograms per second (kg/s).

As system responses of the process we consider the recovery of conductive products ($R_{c,c}$) and the grade of conductive products ($G_{c,c}$) in the collecting box defined as follows:

$$R_{c,c} = \frac{m_{c,c}}{m_{c,c} + m_{c,m} + m_{c,nc}}, \tag{5}$$

where $m_{c,c}$ is the mass of conductive products in its collecting box, $m_{c,m}$ is the mass of conductive products in the middling box and $m_{c,nc}$ is the mass of conductive products in the non-conductive box. Instead, the grade is:

$$G_{c,c} = \frac{m_{c,c}}{m_{c,c} + m_{nc,c}}, \tag{6}$$

where $m_{nc,c}$ is the mass of non-conductive products in the conductive box. Both the system responses lie in the range $[0, 1]$ and should be simultaneously maximised in order to optimise the performance of a CES process.

From available data, we estimated two multiple regression models, ϕ_1 and ϕ_2, in order to simulate the responses $R_{c,c}$ and $G_{c,c}$ over the whole search space. Given a discretisation of the domain of the variables x_1, x_2 and x_3, we determined a search space of size 10^3 and deploying ϕ_1 and ϕ_2 we calculated the two system responses, from now on $\hat{y}_1$ and $\hat{y}_2$.

At this point, it was possible to calculate the real Pareto front of the whole search space obtaining a Pareto front of size 10. This result has been used to evaluate the performance of the multi-objective Bayesian sequential DOE approach.

4.1 Performance Indicators

When comparing the performance of multi-objective approaches, classical indicators are not suitable for understanding which approach is reaching the best Pareto front. However some possible indicators are: *hypervolume distance* and *equal number of points index*.

The hypervolume distance (*hd*) is the Euclidean distance between two hypervolume indicators. A hypervolume indicator measures the area of the response space that is dominated by the considered Pareto front. The larger this area, the better the set of design points in the Pareto front. If the real Pareto front is available, it is possible to calculate the hypervolume distance between the real Pareto front and the optimised Pareto front.

The equal number of points (*enp*) is the number of points lying in both the real Pareto front and the optimised one.

4.2 Results

Since we consider a randomly chosen initial design matrix, we computed $B = 50$ Monte-Carlo runs. The sample size n of the initial design was fixed to 20. A run was stopped after $T = 30$ iterations. Furthermore, we set R_0 as the identity matrix and σ^2 to 1. The values of α and β were varied in order to understand the contribution of the two terms in the Bayesian utility functions, $U^i(\mathbf{x}_{n+1}|X)$. In Table 1, the performance indicators are reported. The two indicators are the average values of *hd* and *enp* calculated over the 50 Monte-Carlo runs.

From Table 1, we notice that the larger the value of α the better the performance of the multi-objective Bayesian sequential DOE approach. This behaviour indicates that the first term of $U^i(\mathbf{x}_{n+1}|X)$, which is the one devoted to maximising the expected system responses, is more important than the second term, which is related to the expected gain in Shannon information [2]. When $(\alpha, \beta) = (1.0, 0.0)$, we obtain an average equal number of points close to 8 out of 10 and an average hypervolume indicator close to 0. However, this behaviour can lead to an early convergence on locally optimal solutions with more complex search spaces. A compromise between α and β is probably then preferable.

Table 1 Average performance indicators with Monte-Carlo standard deviation in parentheses

(α, β)	$\overline{hd}$	$\overline{enp}$
(0.00, 1.00)	0.06 (0.03)	1.48 (0.79)
(0.20, 0.80)	0.06 (0.02)	1.70 (1.76)
(0.50, 0.50)	0.01 (0.01)	1.38 (0.87)
(0.80, 0.20)	0.01 (0.01)	2.38 (0.78)
(0.90, 0.10)	0.01 (0.01)	3.80 (1.03)
(1.00, 0.00)	0.01 (0.01)	8.40 (3.32)

5 Conclusions

In this work, we proposed a multi-objective Bayesian sequential DOE approach and, more precisely, we concentrated our effort on a bi-objective problem derived from a simulated case study concerning the problem of separating metal and nonmetal particles recycled from waste electrical and electronic equipments (WEEE).

The novel approach is applied in a sequential manner and is based on the combination of the Bayesian paradigm and the Pareto front concept. We proposed a Bayesian one-dimensional utility function, which represents a trade-off between the expected system responses and the expected gain in Shannon information for the posterior distribution. At each step, the Bayesian optimal utilities are used to compute the Pareto front. The design point with the minimum Euclidean distance to the utopia point is selected for the next experiment.

The multi-objective Bayesian sequential DOE approach has shown promising results but it still needs further improvement. The Bayesian framework should be generalised to a non-linear setting. Furthermore, the sampling technique for the initial design and the selection method of the next design points need to be investigated in greater depth.

From the results, we observed that the tuning of the two non-negative weights α and β is important. Furthermore, we need to compare the multi-objective Bayesian sequential DOE approach with other methods in order to understand the real power of the method.

Acknowledgements This work has been carried out as a part of the FIDEAS project (Fabbrica Intelligente per la Deproduzione Avanzata e Sostenibile) co-funded within the Framework Agreement between Regione Lombardia and CNR.

References

1. Borrotti, M., Pievatolo, A., Critelli, I., Degiorgi, A., Colledani, M.: A computer-aided methodology for the optimization of electrostatic separation processes in recycling. Appl. Stoch. Models Bus. (2015, in press)
2. Chaloner, K., Verdinelli, I.: Bayesian experimental design: a review. Stat. Sci. **10**, 273–304 (1995)
3. Cheng, Y., Shen, Y.: Bayesian adaptive designs for clinical trials. Biometrika **92**, 633–646 (2005)
4. Grübler, A.: Technology and Global Change. Cambridge University Press, New York (2003)
5. Lu, L., Anderson-Cook, C.M., Robinson, T.J.: Optimization of design experiments based on multiple criteria utilizing a Pareto frontier. Technometrics **53**, 353–365 (2011)
6. Pauwels, E., Lajaunie, C., Vert, J.P.: A Bayesian active learning strategy for sequential experimental design in systems biology. BMC Syst. Biol. **8**, 1–11 (2014)
7. Müller, W.G., Pronzato, L., Rendas, J., Waldl, H.: Efficient prediction designs for random fields. Appl. Stoch. Models Bus. **31**, 178–194 (2015)
8. Sambo, F., Borrotti, M., Mylona, K.: A coordinate-exchange two-phase local search algorithm for the D- and I-optimal designs of split-plot experiments. Comput. Stat. **71**, 1193–1207 (2014)
9. Verdinelli, I., Kadane, J.B.: Bayesian design for maximizing information and outcome. J. Am. Stat. Assoc. **87**, 510–515 (1992)

Optimum Design via I-Divergence for Stable Estimation in Generalized Regression Models

Katarína Burclová and Andrej Pázman

Abstract Optimum designs for parameter estimation in generalized regression models are usually based on the Fisher information matrix (cf. Atkinson et al. (J Stat Plan Inference 144:81–91, 2014) for a recent exposition). The corresponding optimality criteria are related to the asymptotic properties of maximum likelihood (ML) estimators in such models. However, in finite sample experiments there can be problems with identifiability, stability and uniqueness of the ML estimate, which are not reflected by information matrices. In Pázman and Pronzato (Ann Stat 42:1426–1451, 2014) and in Chap. 7 of Pronzato and Pázman (Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties. Springer, New York, 2013) is discussed how to solve some of these estimability issues at the design stage of an experiment in standard nonlinear regression. Here we want to extend this design methodology to more general models based on exponential families of distributions (binomial, Poisson, normal with parametrized variances, etc.). The main tool is the information (or Kullback-Leibler) divergence, which is closely related to ML estimation.

1 Introduction

With each design point $x \in \mathcal{X}$, the design space, we associate an observation y (a random variable or vector), which is distributed according to the density of an exponential form

$$f(y \mid \theta, x) = \exp\left\{-\psi(y) + t^{\top}(y)\,\gamma - \kappa(\gamma)\right\}_{\gamma=\gamma(x,\theta)}, \tag{1}$$

with the unknown parameter θ taking values from a given parameter space $\Theta \subset \mathbb{R}^p$. This density is taken with respect to a measure $\nu(\cdot)$ on Y, the sample space of y. Usually $Y \subset \mathbb{R}^s$ and ν is the Lebesgue measure, or Y is finite or countable and $\nu(\{y\}) = 1$ for every $y \in Y$; then $f(y \mid \theta, x)$ is simply the probability of y.

K. Burclová (✉) • A. Pázman
Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovak Republic
e-mail: katarina.burclova@fmph.uniba.sk; pazman@fmph.uniba.sk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_7

Well known examples are the one-dimensional normal density

$$f(y \mid \theta, x) = \frac{1}{(2\pi)^{1/2}\,\sigma(x,\theta)} \exp\left[-\frac{\{y-\mu(x,\theta)\}^2}{2\sigma^2(x,\theta)}\right]; \quad y \in \mathbb{R}$$

or the binomial probability distribution

$$f(y \mid \theta, x) = \binom{n}{y} \pi(x,\theta)^y \{1-\pi(x,\theta)\}^{n-y}; \quad y \in \{0,1,\ldots,n\}\ . \tag{2}$$

Consider an exact design $X = (x_1, \ldots, x_N)$, where $x_i \in \mathcal{X}$ and the observations $y_{x_1}, \ldots, y_{x_N}$ are independent. The ML estimator for θ is $\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^N \ln f(y_{x_i} \mid \theta, x_i)$. For large N, and under some regularity assumptions, $\hat{\theta}$ is approximately normally distributed with mean θ and covariance matrix $M^{-1}(X,\theta)$, where $M(X,\theta) = \sum_{i=1}^N M(x_i,\theta)$, and $M(x,\theta) = E_\theta\left\{-\frac{\partial^2 \ln f(y|\theta,x)}{\partial\theta\,\partial\theta^\top}\right\}$ is the elementary information matrix at x. Hence within this asymptotic approximation, a design X is (locally) optimal if it maximizes $\Phi\left\{M\left(X,\theta^0\right)\right\}$, where θ^0 is a guess for the true (but unknown) value of θ. Here $\Phi(\cdot)$ stands for $\det^{1/p}(\cdot)$ in case of D-optimality, etc. This is the standard way to optimize designs in generalized regression models.

Alternatively to the information matrix, we take here for design purposes the I-divergence (the information or Kullback-Leibler divergence, cf. [4]), which for any two points $\theta^0, \theta \in \Theta$ is equal to $I_X\left(\theta^0,\theta\right) = \sum_{i=1}^N I_{x_i}\left(\theta^0,\theta\right)$ with the elementary I-divergence defined by

$$I_x\left(\theta^0,\theta\right) = E_{\theta^0}\left\{\ln\frac{f\left(y \mid \theta^0, x\right)}{f(y \mid \theta, x)}\right\}\ . \tag{3}$$

As is well known, $I_x\left(\theta^0,\theta\right) \geq 0$ and is equal to zero if and only if $f\left(y \mid \theta^0, x\right) = f(y \mid \theta, x)$. In general, the I-divergence measures well the sensitivity of the data y to the shift of the parameter from the value θ^0 to the value θ, even when θ and θ^0 are distant, while the information matrix $M\left(x,\theta^0\right)$ is doing essentially the same, but only for θ which is close to θ^0 (see Sect. 3). Hence the I-divergence may allow a more global characterization of the statistical properties of the model than the information matrix. An important fact is also that we can compute it easily (avoiding integrals) in models given by (1). Notice that for the normal model with $\sigma^2(x,\theta) \equiv 1$, one has $2I_X\left(\theta^0,\theta\right) = \sum_{i=1}^N \left\{\mu\left(x_i,\theta^0\right) - \mu(x_i,\theta)\right\}^2$, an expression, which is largely used in [6] and in Chap. 7 of [7].

We note that in [5] the I-divergence has been used in optimal design for model discrimination, which, however, corresponds to a different objective than the one considered here.

2 Basic Properties of Model (1)

It is clear that $t(y)$ is a sufficient statistic in model (1), so we can suppose, at least in theory, that we observe $t(y)$ instead of y. Denote by $\eta(x,\theta)$ its mean. For $\gamma = \gamma(x,\theta)$ we have

$$\eta(x,\theta) = \int_Y t(y) \exp\left\{-\psi(y) + t^\top(y)\gamma - \kappa(\gamma)\right\}_{\gamma=\gamma(x,\theta)} d\nu(y) = \left\{\frac{\partial \kappa(\gamma)}{\partial \gamma}\right\}_{\gamma=\gamma(x,\theta)} . \tag{4}$$

To be able to do this derivative at any γ, we suppose that the set $\{\gamma : \int_Y \exp\{-\psi(y) + t^\top(y)\gamma\} d\nu(y) < \infty\}$ is open. Then the model (1) is called regular, and regular models are standard in applications. Taking the second order derivative in (4) we obtain $\frac{\partial^2 \kappa(\gamma)}{\partial \gamma \partial \gamma^\top} = Var_\gamma\{t(y)\}$, which for $\gamma = \gamma(x,\theta)$ will be denoted by $\Sigma(x,\theta)$. By a reduction of the linearly dependent components of the vector $t(y)$ one can always achieve that $\Sigma(x,\theta)$ is nonsingular, and we obtain from (4) that $\frac{\partial \eta(x,\theta)}{\partial \theta^\top} = \Sigma(x,\theta) \frac{\partial \gamma(x,\theta)}{\partial \theta^\top}$.

The functions $\theta \in \Theta \to \eta(x,\theta)$ (the mean-value function) and $\theta \in \Theta \to \gamma(x,\theta)$ (the canonical function) are useful dual representations of the family of densities (1) (cf. [3]). From (1) and (4) it follows that $E_{x,\theta}\left\{\frac{\partial \ln f(y|x,\theta)}{\partial \theta}\right\} = 0$, and consequently the elementary information matrix is equal to

$$M(x,\theta) = Var_{x,\theta}\left\{\frac{\partial \ln f(y \mid x,\theta)}{\partial \theta}\right\} = \frac{\partial \eta^\top(x,\theta)}{\partial \theta} \Sigma^{-1}(x,\theta) \frac{\partial \eta(x,\theta)}{\partial \theta^\top} . \tag{5}$$

The elementary I-divergence is, according to (3) and (1),

$$I_x\left(\theta^0,\theta\right) = \eta^\top\left(x,\theta^0\right)\left\{\gamma\left(x,\theta^0\right) - \gamma(x,\theta)\right\} + \kappa\{\gamma(x,\theta)\} - \kappa\left\{\gamma\left(x,\theta^0\right)\right\} .$$

For more details on exponential families see [2].

3 Variability, Stability and I-Divergence

In this section we consider observations according to an "exact" design $X = (x_1,\ldots,x_N)$. The joint density of $\mathbf{y} = (y_{x_1},\ldots,y_{x_N})^\top$ is equal to $\widetilde{f}(\mathbf{y} \mid \theta) = \prod_{i=1}^N f(y_{x_i} \mid \theta, x_i)$.

The variability of the ML estimate $\hat{\theta}$ in the neighborhood of $\bar{\theta}$, the true value of θ, is well expressed by the information matrix $M\left(X,\bar{\theta}\right)$, since its inverse is the asymptotic covariance matrix of $\hat{\theta}$. But the same can be achieved by the I-divergence, since in model (1) it satisfies the following property.

Lemma 1 *If for any $x \in \mathcal{X}$ the third-order derivatives of $I_x\left(\bar{\theta}, \theta\right)$ with respect to θ are bounded on a neighborhood of $\bar{\theta}$, then we have*

$$I_X\left(\bar{\theta}, \theta\right) = \frac{1}{2}\left(\theta - \bar{\theta}\right)^{\top} M\left(X, \bar{\theta}\right)\left(\theta - \bar{\theta}\right) + o\left(\left\|\theta - \bar{\theta}\right\|^2\right) . \tag{6}$$

It is sufficient to prove this equality for the elementary I-divergence and elementary information matrix. We have $I_x\left(\bar{\theta}, \bar{\theta}\right) = 0$, $\frac{\partial I_x\left(\bar{\theta}, \theta\right)}{\partial \theta} \mid_{\theta=\bar{\theta}} = 0$, $\frac{\partial^2 I_x\left(\bar{\theta}, \theta\right)}{\partial \theta \partial \theta^{\top}} \mid_{\theta=\bar{\theta}} = M\left(x, \bar{\theta}\right)$, so by the Taylor formula we obtain (6).

On the other hand, we can have important instabilities of the ML estimate $\hat{\theta}$ when, with a large probability, $\ln \widetilde{f}\left(\mathbf{y} \mid \theta\right)$ is close to $\ln \widetilde{f}\left(\mathbf{y} \mid \bar{\theta}\right)$ for a point θ distant from $\bar{\theta}$. However, at the design stage we do not know the value of $\mathbf{y}$, so we cannot predict the value of the difference $\ln \widetilde{f}\left(\mathbf{y} \mid \theta\right) - \ln \widetilde{f}\left(\mathbf{y} \mid \bar{\theta}\right)$. But we can at least predict its mean.

Lemma 2 *For any $\theta \in \Theta$ we have $E_{\bar{\theta}}\left\{\ln \widetilde{f}\left(\mathbf{y} \mid \bar{\theta}\right) - \ln \widetilde{f}\left(\mathbf{y} \mid \theta\right)\right\} = I_X\left(\bar{\theta}, \theta\right)$.*

This equality is evident from (3).

As a consequence of Lemmas 1 and 2 we have that the I-divergence $I_X\left(\bar{\theta}, \theta\right)$ can express simultaneously both the variability of the ML estimate $\hat{\theta}$ in a neighborhood of $\bar{\theta}$ and the danger of instability of the ML estimation due to the possibility of "false" estimates which are very distant from the true value $\bar{\theta}$.

4 Extended Optimality Criteria

According to the principle mentioned in Sect. 3, the design based on the I-divergence should minimize the variability of $\hat{\theta}$ (related to the information matrix) and protect against instabilities coming from a value θ which is distant from the true value $\bar{\theta}$. This requirement can be well reflected by extended optimality criteria (cf. [6, 7] for classical nonlinear regression). To any design (design measure or approximate design) ξ on a finite design space $\mathcal{X}$ we define the extended criteria by a form

$$\phi_{ext}\left(\xi, \theta^0\right) = \min_{\theta \in \Theta} \sum_{x \in \mathcal{X}} 2 I_x\left(\theta^0, \theta\right) \left\{ \frac{1}{\rho^2\left(\theta^0, \theta\right)} + K \right\} \xi\left(x\right) , \tag{7}$$

where $K \geq 0$ is a tuning constant chosen in advance, θ^0 is a guess for the (unknown) value of θ, $\rho\left(\theta^0, \theta\right)$ is a distance measure (a norm or a pseudonorm) not depending on the design ξ. When $\rho\left(\theta^0, \theta\right) = \left\|\theta^0 - \theta\right\|$, the Euclidean norm, we have the extended E-optimality criterion, denoted by $\phi_{eE}\left(\xi, \theta^0\right)$. When $\rho\left(\theta^0, \theta\right) = \left|h\left(\theta^0\right) - h\left(\theta\right)\right|$, with $h\left(\theta\right) \in \mathbb{R}$, a given function of θ, we have the extended c-optimality criterion, denoted by $\phi_{ec}\left(\xi, \theta^0\right)$. When $\rho\left(\theta^0, \theta\right) = \max_{x \in \mathcal{X}} \left|\alpha\left(x, \theta^0\right) - \alpha\left(x, \theta\right)\right|$, with $x \in \mathcal{X} \rightarrow \alpha\left(x, \theta\right) \in \mathbb{R}$, the regression function of interest, we have the

extended G-optimality criterion, denoted $\phi_{eG}\left(\xi, \theta^0\right)$. Notice that usually $\alpha\left(x, \theta\right)$ is equal to $\gamma\left(x, \theta\right)$ or to $\eta\left(x, \theta\right)$, but not necessarily. The names of the criteria are justified by the following theorem.

Theorem 1 *Let $\mathscr{B}\left(\theta^0, \delta\right)$ be a sphere centred at θ^0 with diameter δ, and $M\left(\xi, \theta^0\right) = \sum_{x \in \mathscr{X}} M\left(x, \theta^0\right) \xi(x)$. Then*

$$\lim_{\delta \to 0} \min_{\theta \in \mathscr{B}\left(\theta^0, \delta\right)} \sum_{x \in \mathscr{X}} 2 I_x\left(\theta^0, \theta\right) \left\{ \frac{1}{\rho^2\left(\theta^0, \theta\right)} + K \right\} \xi(x)$$

is equal to

- $\lambda_{\min}\left\{M\left(\xi, \theta^0\right)\right\}$, *the minimal eigenvalue, when* $\rho\left(\theta^0, \theta\right) = \left\|\theta^0 - \theta\right\|$,
- $\left\{c^\top M^{-1}\left(\xi, \theta^0\right) c\right\}^{-1}$ *with* $c = \left\{\frac{\partial h(\theta)}{\partial \theta}\right\}_{\theta = \theta^0}$ *if* $M\left(\xi, \theta^0\right)$ *is nonsigular and if* $\rho\left(\theta^0, \theta\right) = \left|h\left(\theta^0\right) - h(\theta)\right|$,
- $\left\{\max_{x \in \mathscr{X}} f^\top(x) M^{-1}\left(\xi, \theta^0\right) f(x)\right\}^{-1}$ *with* $f(x) = \left\{\frac{\partial \alpha(x, \theta)}{\partial \theta}\right\}_{\theta = \theta^0}$ *if* $M\left(\xi, \theta^0\right)$ *is nonsingular and if* $\rho\left(\theta^0, \theta\right) = \max_{x \in \mathscr{X}} \left|\alpha\left(x, \theta^0\right) - \alpha(x, \theta)\right|$.

When the model is normal, linear, with unit variances of observations, $h(\theta)$ is a linear function of θ and $\alpha(x, \theta) = \eta(x, \theta)$ is linear, then $\phi_{ext}\left(\xi, \theta^0\right)$ coincides with the corresponding well known optimality criterion in linear models.

Proof The proof follows from Lemma 1 and from known expressions: $\lambda_{\min}\{M\} = \min_u \left\{u^\top M u / u^\top u\right\}$, $c^\top M^{-1} c = \max_{u:\, u \neq 0} \left\{\left(c^\top u\right)^2 / u^\top M u\right\}$ (if M is nonsingular). In a normal linear model with unit variances we use that $\gamma(x, \theta) = \eta(x, \theta) = f^\top(x)\theta$ and $\sum_{x \in \mathscr{X}} 2 I_x\left(\theta^0, \theta\right) \xi(x) = \left(\theta - \theta^0\right)^\top \left\{\sum_{x \in \mathscr{X}} f(x) f^\top(x) \xi(x)\right\} \left(\theta - \theta^0\right)$. □

If K is chosen very large, then the extended criterion gives the same optimum design as the classical criterion. On the other hand, when K is very small, then $1/\rho^2\left(\theta^0, \theta\right)$ is the dominating term within the braces in (7), and (7) becomes small simply because θ and θ^0 are very distant points.

Remark 1 We can write (7) in the form

$$\phi_{ext}\left(\xi, \theta^0\right) = \min_{\theta \in \Theta} \sum_{x \in \mathscr{X}} H\left(x, \theta^0, \theta\right) \xi(x) \ , \tag{8}$$

with an adequately chosen $H\left(x, \theta^0, \theta\right)$. It follows that the function $\xi \to \phi_{ext}\left(\xi, \theta^0\right)$ is concave, hence it has a directional derivative, and the "equivalence theorem" can be formulated, exactly as in [6].

Remark 2 When the design space $\mathscr{X}$ is finite, $\mathscr{X} = \left\{x^1, \ldots, x^k\right\}$, we can consider the task of computing the optimum design, $\xi^* = \arg\max_\xi \phi_{ext}\left(\xi, \theta^0\right)$ as an "infinite-dimensional" linear programming (LP) problem. Namely, we have to find

the vector $\left(t, \xi\left(x^1\right), \ldots, \xi\left(x^k\right)\right)$, which maximizes t under the linear constraints

$$\sum_{x \in \mathscr{X}} H\left(x, \theta^0, \theta\right) \xi(x) \geq t \text{ for every } \theta \in \Theta\,,$$

$$\sum_{x \in \mathscr{X}} \xi(x) = 1,\ \xi(x) \geq 0 \text{ for every } x \in \mathscr{X}\,.$$

As in [6], the solution can be approached by solving a sequence of LP problems, including a stopping rule (see the 'Numerical Example' below).

Illustrative Example Consider the binomial model (2), which can be written in the exponential form (1):

$$f(y \mid \theta, x) = \exp\left[\ln\binom{n}{y} + y\gamma(x, \theta) - n \ln\left\{1 + \mathrm{e}^{\gamma(x,\theta)}\right\}\right]$$

with $\gamma(x, \theta) = \ln[\pi(x, \theta) / \{1 - \pi(x, \theta)\}]$, and with the mean of $y = t(y)$ equal to $\eta(x, \theta) = n\pi(x, \theta) = n\mathrm{e}^{\gamma(x,\theta)} / \left\{1 + \mathrm{e}^{\gamma(x,\theta)}\right\}$ (the logistic function). In the example we take $n = 10$ and consider the regression model (similar to that in [6])

$$\gamma(x, \theta) = 2\cos(t - u\theta)\,;\ x = (t, u)^\top\,,$$

with two observations, one at $x_1 = (0, u)^\top$ and the second at $x_2 = (\pi/2, u)^\top$, where $u \in \left[0, \frac{11}{6}\pi\right]$ is to be chosen optimally for the estimation of the unknown parameter $\theta \in [0, 1]$. For the case of $u = \frac{11}{6}\pi$ we can see the circular "canonical surface" $\left\{(\gamma(x_1, \theta), \gamma(x_2, \theta))^\top\,;\ \theta \in [0, 1]\right\}$ in Fig. 1a and the "expectation surface" $\left\{(\eta(x_1, \theta), \eta(x_2, \theta))^\top\,;\ \theta \in [0, 1]\right\}$ in Fig. 1b (which is no longer circular due to the nonlinearity of the logistic function). The information "matrix" $M_u(\theta) \equiv M(x_1, \theta) + M(x_2, \theta)$ is computed according to (5), and for $\theta^0 = 0$ it is equal to $M_u\left(\theta^0\right) = nu^2$. It follows that the classical locally optimal design maximizing

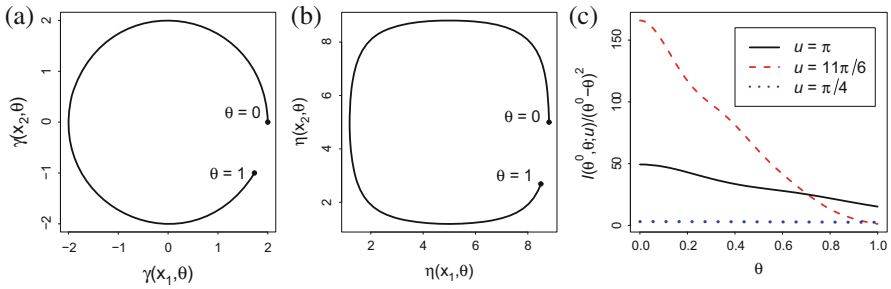

Fig. 1 **(a)** The canonical surface, **(b)** the expectation surface, **(c)** $I\left(\theta^0, \theta; u\right) / \left(\theta - \theta^0\right)^2$

$M_u\left(\theta^0\right)$ is obtained when $u = \frac{11}{6}\pi$. We see even from Fig. 1a, b that under this design, the ML estimate $\hat{\theta}\,(y)$ can be, with a large probability, in the neighborhood of $\theta = 1$, hence the estimator is unstable when $\theta^0 = \bar{\theta} = 0$.

On the other hand, take the I-divergence $I\left(\theta^0, \theta; u\right) \equiv I_{x_1}\left(\theta^0, \theta\right) + I_{x_2}\left(\theta^0, \theta\right)$ with

$$I_x\left(\theta^0, \theta\right) = n\left[\pi\left(x, \theta^0\right) \ln \frac{\pi\left(x, \theta^0\right)}{\pi\left(x, \theta\right)} + \left\{1 - \pi\left(x, \theta^0\right)\right\} \ln \frac{1 - \pi\left(x, \theta^0\right)}{1 - \pi\left(x, \theta\right)}\right] \tag{9}$$

and consider the extended criterion $\phi_{ext}\left(u, \theta^0\right) = \min_{\theta \in [0,1]} I\left(\theta^0, \theta; u\right) / \left(\theta - \theta^0\right)^2$. Numerical computation gives that $\arg\max_{u \in \left[0, \frac{11}{6}\pi\right]} \phi_{ext}\left(u, \theta^0\right) \doteq \pi$, and for this choice of u the probability of a false $\hat{\theta}\,(y)$ is negligible, because then the values of $\left(\gamma\left(x_1, \theta\right), \gamma\left(x_2, \theta\right)\right)^\top$ for $\theta = 0$ and $\theta = 1$ are as distant as possible. The same holds for $\left(\eta\left(x_1, \theta\right), \eta\left(x_2, \theta\right)\right)^\top$. Here we took the tuning constant $K = 0$, the optimal u would be between π and $\frac{11}{6}\pi$ for K positive. In Fig. 1c we present the dependence of $I\left(\theta^0, \theta; u\right) / \left(\theta - \theta^0\right)^2$ on θ for different values of u.

Numerical Example The aim of this example is to show that in case of a finite $\mathscr{X} = \left\{x^1, \ldots, x^k\right\}$, we can compute the extended E-optimum design by solving a sequence of LP problems—see Remark 2. The expected value of an observed variable is as in [6], Example 2, i.e. the mean of y_x, observed at the design point $x = \left(x_1, x_2\right)^\top$, is equal to

$$\eta\left(x, \theta\right) = n\pi\left(x, \theta\right) = \frac{n}{6}\left\{1 + \theta_1 x_1 + \theta_1^3\left(1 - x_1\right) + \theta_2 x_2 + \theta_2^2\left(1 - x_2\right)\right\}. \tag{10}$$

However, the error structure is quite different. Instead of an error component not depending on $\theta = \left(\theta_1, \theta_2\right)^\top$, we consider here a binomial model with y_x distributed according to (2), with $n = 10$ and $\pi\left(x, \theta\right)$ given by (10). Consequently, in the extended criterion (7) we use the binomial I-divergence (9) and $\rho^2\left(\theta^0, \theta\right) = \left\|\theta^0 - \theta\right\|^2$. We choose $\theta^0 = (1/8, 1/8)^\top$, $x \in \mathscr{X} = \{0, 0.1, \ldots, 0.9, 1\}^2$, $\theta \in \Theta = [-1, 1] \times [0, 2]$.

The iterative algorithm presented below follows the same lines as in [6].

1. Take any vector $\left(\xi^{(0)}\left(x^1\right), \ldots, \xi^{(0)}\left(x^k\right)\right)$ such that $\sum_{x \in \mathscr{X}} \xi^{(0)}(x) = 1$ and $\xi^{(0)}(x) \geq 0 \;\forall\; x \in \mathscr{X}$, choose $\epsilon > 0$, set $\Theta^{(0)} = \emptyset$ and $n = 0$. Construct a finite grid $\mathscr{G}^{(0)}$ in Θ.
2. Set $\Theta^{(n+1)} = \Theta^{(n)} \cup \left\{\theta^{(n+1)}\right\}$, where $\theta^{(n+1)} = \arg\min_{\theta \in \Theta} \sum_{x \in \mathscr{X}} H\left(x, \theta^0, \theta\right) \xi^{(n)}(x)$ is computed as follows:
   (a) Compute $\hat{\theta}^{(n+1)} = \arg\min_{\theta \in \mathscr{G}^{(n)}} \sum_{x \in \mathscr{X}} H\left(x, \theta^0, \theta\right) \xi^{(n)}(x)$.
   (b) Perform a local minimization over Θ initialized at $\hat{\theta}^{(n+1)}$, denote by $\theta^{(n+1)}$ the solution and set $\mathscr{G}^{(n+1)} = \mathscr{G}^{(n)} \cup \left\{\theta^{(n+1)}\right\}$.

Table 1 ϕ_{eE}-optimal designs for $K = 0$ and $K = 5$, and corresponding values of ϕ_{eE}, see (7), for $K = 0$ and $K = 5$

K	ξ^*	$\phi_{eE}\left(\xi^*, \theta^0\right)$ for $K = 0$	$\phi_{eE}\left(\xi^*, \theta^0\right)$ for $K = 5$
0	$\begin{Bmatrix} (0,0)^\top & (0,1)^\top & (1,1)^\top \\ 0.345 & 0.029 & 0.626 \end{Bmatrix}$	0.0215	0.0249
5	$\begin{Bmatrix} (0,0)^\top & (1,0)^\top & (0,1)^\top & (1,1)^\top \\ 0.247 & 0.072 & 0.197 & 0.484 \end{Bmatrix}$	0.0165	0.1972

3. Use an LP solver to find $\left(t^{(n+1)}, \xi^{(n+1)}\left(x^1\right), \ldots, \xi^{(n+1)}\left(x^k\right)\right)$ to maximize $t^{(n+1)}$ with the constraints:
   - $t^{(n+1)} > 0,\ \xi^{(n+1)}(x) \geq 0\ \forall\ x \in \mathscr{X},\ \sum_{x\in\mathscr{X}} \xi^{(n+1)}(x) = 1,$
   - $\sum_{x\in\mathscr{X}} H\left(x, \theta^0, \theta\right) \xi^{(n+1)}(x) \geq t^{(n+1)}\ \forall \theta \in \Theta^{(n+1)}.$
4. Set $\Delta^{(n+1)} = t^{(n+1)} - \phi_{ext}\left\{\xi^{(n+1)}, \theta^0\right\}$, if $\Delta^{(n+1)} < \epsilon$, take $\xi^{(n+1)}$ as an ϵ-optimal design and stop, else set $n + 1$ for n and continue from Step 2.

The computations were performed in Matlab computing environment and we used the linprog function with the default interior point algorithm to solve LP problems. When the grid $\mathscr{G}^{(0)}$ was taken as a random Latin hypercube design with 10,000 points renormalized to Θ, $\epsilon = 10^{-10}$, and $\xi^{(0)}$ put mass uniformly to each x in $\mathscr{X}$, the algorithm stopped after 20 iterations for $K = 0$ and after 47 iterations for $K = 5$. The numerical results are summarized in Table 1.

Acknowledgements The authors thank Slovak Grant Agency VEGA, Grant No. 1/0163/13, for financial support.

References

1. Atkinson, A.C., Fedorov, V.V., Herzberg, A.M., Zhang, R.: Elemental information matrices and optimal experimental design for generalized regression models. J. Stat. Plan. Inference **144**, 81–91 (2014)
2. Brown, L.D.: Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. IMS Lecture Notes—Monograph Series, vol. 9. Institute of Mathematical Statistics, Hayward (1986)
3. Efron, B.: The geometry of exponential families. Ann. Stat. **6**, 362–376 (1978)
4. Kullback, S.: Information Theory and Statistics. Dover Publications, Mineola, N.Y. (1997)
5. López-Fidalgo, J., Tommasi, C., Trandafir, P.C.: An optimal experimental design criterion for discriminating between non-normal models. J. R. Stat. Soc. B. **69**, 231–242 (2007)
6. Pázman, A., Pronzato, L.: Optimum design accounting for the global nonlinear behavior of the model. Ann. Stat. **42**, 1426–1451 (2014)
7. Pronzato, L., Pázman, A.: Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties. Springer, New York (2013)

On Multiple-Objective Nonlinear Optimal Designs

Qianshun Cheng, Dibyen Majumdar, and Min Yang

Abstract Experiments with multiple-objectives form a staple diet of modern scientific research. Deriving optimal designs with multiple-objectives is a long-standing challenging problem with few tools available. The few existing approaches cannot provide a satisfactory solution in general: either the computation is very expensive or a satisfactory solution is not guaranteed. There is need for a general approach which can effectively derive multi-objective optimal designs. A novel algorithm is proposed to address this literature gap. We prove convergence of this algorithm, and show in various examples that the new algorithm can derive the true solutions with high speed.

1 Introduction

With the development of computational technology, nonlinear models have become more feasible and popular. An optimal/efficient design can improve the accuracy of statistical inferences with a given sample size or reduce the sample size needed for a pre-specified accuracy.

A common approach to study optimal designs is to use locally optimal designs, which are based on the best guess of the parameters [3]. The selection of an optimal design depends on the goals of the experiment. In practice, it is common for an experimenter to have multiple objectives. An optimal design based on one specific objective could be a disaster for other objectives. As in a multiple comparison study, a design with maximum power for one of the hypotheses may not have good power for other tests.

There are two ways of formulating the multiple-objective optimal design problem. One is based on compound optimality criteria and the other is based on constrained optimality criteria. Here we focus on constrained optimization, which is to maximize one objective function subjected to all other objective functions satisfying certain efficiencies.

Q. Cheng (✉) • D. Majumdar • M. Yang
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, USA
e-mail: qcheng5@uic.edu; dibyen@las.uic.edu; minyang.stat@gmail.com

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_8

There are two numerical approaches in the literature for constrained optimization: the grid search and the sequential approach. For the grid search approach, the number of grid points increases exponentially with the number of objectives, and can be huge even for a moderate number of objectives. With three objectives, Huang and Wong [7] proposed a sequential approach for finding the weights. The basic idea is to consider the objective functions in pairs and sequentially add more constraints. While this seems to have given reasonable answers in their examples, there lacks theoretical justification. Consequently this approach will generally not yield satisfactory solution even for the three-objective problem.

The goal of this paper is to propose a novel algorithm for deriving optimal designs for multiple objective functions. For a given constrained optimization problem, if the solution exists, we prove that the new algorithm guarantees to find the optimum weights. The new algorithm is also fast.

2 Set Up and Notation

We adopt the same notation as that of [8]. Denote the information matrix of design ξ as $\mathbf{I}_\xi$. Let $\Phi_0(\xi), \ldots, \Phi_n(\xi)$ be the values of $n+1$ smooth objective functions for design ξ. Ideally, we hope we can find a ξ^* which can maximize $\Phi_0(\xi), \ldots, \Phi_n(\xi)$ simultaneously among all possible designs. An ideal solution does not exist in general. One commonly used way of formulating a multiple objective design problem is the constrained optimization approach, which specifies one objective as the primary criteria and maximizes this objective subject to constraints on the remaining objectives [4, 6]. Formally, this approach can be written as

$$\underset{\xi \in \Xi}{\text{Maximize}}\ \Phi_0(\xi) \text{ subject to } \Phi_i(\xi) \geq c_i,\ i = 1, \ldots, n, \tag{1}$$

where $\mathbf{c} = (c_1, \ldots, c_n)$ are user-specified constants which reflect minimally desired levels of performance relative to optimal designs for these n objective functions. To make this problem meaningful, throughout this paper, we assume there is at least one design satisfying the constrains, which means an optimal solution exists.

Unfortunately, in a restricted optimality set up, the constrained optimization problem cannot be solved directly. We have to solve (1) through the corresponding compound optimal design. Let

$$L(\xi, \mathbf{U}) = \Phi_0(\xi) + \sum_{i=1}^{n} u_i(\Phi_i(\xi) - c_i), \tag{2}$$

where $u_i \geq 0$, $i = 1, \ldots, n$. Let $\mathbf{U} = (u_1, \ldots, u_n)$.

To establish the relationship between constrained optimal design problems and compound optimal design problems, we need the following assumptions, which are adapted from [4]. Assume that

1. $\Phi_i(\xi)$, $i = 0, \ldots, n$ are concave on Ξ.
2. $\Phi_i(\xi)$, $i = 0, \ldots, n$ are differentiable and the directional derivatives are continuous in $\mathbf{x}$
3. If ξ_n converges to ξ, then $\Phi_i(\xi_n)$ converges to $\Phi_i(\xi)$, $i = 0, \ldots, n$
4. There is at least one design ξ in Ξ such that the constrains (1) are satisfied.

Clyde and Chaloner [4] generalized a result of [6] and showed the equivalence of the constrained optimization approach and the compound optimality approach. The result demonstrates that a numerical solution for the constrained design problem (1) can be derived by using an appropriate compound optimality criterion. In fact, almost all numerical solutions for constrained design problems use this strategy. The big challenge is how to find the desired $\mathbf{U}^*$ for a given constrained design problem (1). There are two approaches to handle this: a grid search and a sequential approach. Both approaches utilize the weighted optimal design problem, which is equivalent to a compound optimal design approach.

However, both a grid search and a sequential approach have their own problems. Consequently they cannot serve as a general solution for the constrained optimal design problem (1). A general and efficient algorithm should be based on the characterization of $\mathbf{U}^*$. To derive theoretical results, we need the following two assumptions. The first one is

$$\Phi_0 \text{ is a strong concave function on the set of information matrices,} \tag{3}$$

and the second one is

$$u_i^* \in [0, N_i) \text{ where } N_i \text{ is pre-specified, } i = 1, \ldots, n. \tag{4}$$

Due to space limitation, we skip the characterization of $\mathbf{U}^*$, which can be found in [2]. For easy presentation, we denote $\xi_{\mathbf{U}}$ as a design which maximizes the Lagrangian $L(\xi, \mathbf{U})$ for a given weight vector $\mathbf{U} = (u_1, \cdots, u_n)$ and $\hat{\Phi}_i(\xi)$ as $\Phi_i(\xi) - c_i$, $i = 1, \ldots, n$.

3 Algorithm

For a given constrained design problem (1), the new algorithm aims to find the desired $\mathbf{U}^*$. In each step, we need to derive an optimal design for a compound optimal design problem (2) with $\mathbf{U}$ being given, which can be derived based on the optimal weight exchange algorithm (OWEA) proposed by [8]. Here we only present the main algorithm.

The Main Algorithm

The strategy of the algorithm is to search from the simplest case (no constraint is active) to the most complicated case (all constraints are active). The algorithm can be described as following:

Step 1 Set $a = 0$, derive $\xi^* = \text{argmax}_\xi\{\Phi_0(\xi)\}$ and check whether $\Phi_i(\xi^*) \geq c_i$ for $i = 1, \cdots, n$. If all constrains are satisfied, stop and ξ^* is the desired design. Otherwise set $a = 1$ and go to Step 2.

Step 2 Set $i = 1$, consider $\xi^* = \text{argmax}_\xi\{\Phi_0(\xi) + u_i\Phi_i(\xi)\}$. Adjust the value of u_i using the bisection technique on $[0, N_i]$ to obtain u_i^* such that $\hat{\Phi}_i(\xi^*) = 0$. During the bisection process, the upper bound, instead of the median, of the final bisection interval will be picked as the right value for u_i^*. If $\hat{\Phi}_i(\xi^*) > 0$ when $u_i = 0$, set $u_i^* = 0$. If $\hat{\Phi}_i(\xi^*) < 0$ when $u_i = N_i$, set $u_i^* = N_i$. For $\xi^* = \text{argmax}_\xi\{\Phi_0(\xi) + u_i^*\Phi_i(\xi)\}$, check whether $\hat{\Phi}_j(\xi^*) \geq 0$ for $j = 1, \cdots, n$. If all constraints are satisfied, stop and ξ^* is the desired design; otherwise change i to $i + 1$ and repeat this process. After $i = n$ is tested and no desired ξ^* is found, then set $a = 2$ and proceed to Step 3.

Step 3 Find all subsets of $\{1, \cdots, n\}$ of size a, choose one out of these subsets. Denote it as S.

Step 4 Let $(s_1, \ldots, s_a)$ be the indexes of the elements in $\mathbf{U}_S$. To find the right value $\mathbf{U}_S^*$ for $\mathbf{U}_S$, we follow a recurrent process. For each time a given value of u_{s_a}, first use the bisection technique to find the corresponding $u_{s_1}, \cdots, u_{s_{a-1}}$. The full weight vector $\mathbf{U}$ can be constructed with $u_{s_1}, \cdots, u_{s_a}$ by setting all the other weight elements in $\mathbf{U}$ as 0's, which we later denote by $\mathbf{U} = \{\mathbf{U}_S, 0\}$. Then adapt the value of u_{s_a} as follows:

- If $\hat{\Phi}_{s_a}(\xi_\mathbf{U}) > 0$ when u_{s_a} is assigned as 0, set $u_{s_a}^* = 0$.
- If $\hat{\Phi}_{s_a}(\xi_\mathbf{U}) < 0$ when u_{s_k} is assigned as N_a, set $u_{s_a}^* = N_a$.
- Otherwise use the bisection technique to find $u_{s_a}^*$ such that $\hat{\Phi}_{s_a}(\xi_\mathbf{U}) = 0$.

Record $u_{s_a}^*$ and the corresponding values for $\{u_{s_1}^*, \cdots, u_{s_{a-1}}^*\}$ as U_S^*. For the bisection process in each dimension, the upper bound of the final bisection interval will be picked as the right value for the corresponding element in weight vector $\mathbf{U}_S^*$. Then the full weight vector $\mathbf{U}^*$ can be constructed using $\mathbf{U}^* = \{\mathbf{U}_S^*, 0\}$.

Step 5 For the U_S^* and $\xi_{\mathbf{U}^*}$ derived in Step 4, check $\hat{\Phi}_i(\xi_{\mathbf{U}^*})$, $i = 1, \ldots, n$. If all constraints are satisfied, stop and $\xi_{\mathbf{U}^*}$ is the desired design. Otherwise, pick another a-element subset in Step 3, and go through Step 4 to Step 5 again. If all a-element subsets are tested, go to Step 6.

Step 6 Change a to $a + 1$, go through Step 3 to Step 5, until $a = n$. If no suitable design $\xi_{\mathbf{U}^*}$ is found, the implication is that there is no solution for the constrained optimal design (1).

Table 1 Comparison of calculation cost

	Three objectives		Four objectives	
Mesh grid size	0.01	0.001	0.01	0.001
Grid search	5050	500,500	171,700	167,167,000
New algorithm	265	525	4320	12,058

[a] Numbers in the table are counts of weighted optimal designs calculated to solve the constrained design problem for each technique

3.1 Convergence and Computational Cost

Whether an algorithm is successful mainly depends on two properties: convergence and computational cost. We first establish the convergence of the proposed algorithm. The proof can be found in [2].

Theorem 1 *For the constrained optimal design problem (1), under Assumptions (3) and (4), the proposed algorithm converges to* ξ^*.

Table 1 shows the comparison of computational costs between the new algorithm and a grid search under different grid sizes and numbers of constraints.

As for the sequential approach, the computational cost is significantly less than those of grid search and the new algorithm. However, as we will demonstrate in the next section, the sequential approach in general cannot find a satisfactory solution.

4 Numerical Examples

In this section, we will compare the performances (accuracy and the computing time) of the new algorithm, the grid search and the sequential approach. The sequential approach was introduced in [7]. This approach first reorders $\Phi_0, \cdots, \Phi_n$ as $\Phi_{s_1}, \cdots, \Phi_{s_{n+1}}$ according to a robustness technique. In this paper, we test all possible orders and pick the best design. Certainly it includes the special choice in [7].

All three approaches utilize the OWEA algorithm to derive optimal designs for given weighted design problems.

Example I: Atkinson et al. [1] derived Bayesian designs for a compartmental model which can be written as

$$y = \theta_3(e^{-\theta_1 x} - e^{-\theta_2 x}) + \epsilon = \eta(x, \theta) + \epsilon. \tag{5}$$

where ϵ is assumed to follow the normal distribution with mean zero and variance σ^2 and y represents the concentration level of the drug at time point x. Clyde and Chaloner [4] derived multiple objective optimal designs under this model with parameter values $\theta^T = (\theta_1, \theta_2, \theta_3) = (0.05884, 4.298, 21.80)$ and design space X is $[0, 30]$. Interest is in estimating θ as well as the following quantities:

- Area under the curve (AUC),

$$h_1(\theta) = \frac{\theta_3}{\theta_1} - \frac{\theta_3}{\theta_2}$$

- Maximum concentration,

$$c_m = h_2(\theta) = \eta(t_{max}, \theta),$$

where $t_{max} = 1.01$.

Let $\xi_0^* = \underset{\xi \in \Xi}{\text{argmin}} \, |I^{-1}(\xi)|$, c_i be the gradient vector of $h_i(\theta)$ according to the parameter vector θ and $\xi_i^* = \underset{\xi \in \Xi}{\text{argmin}} \, \text{tr}(c_i^T I^{-1}(\xi) c_i)$, $i = 1, 2$. The corresponding objective functions can be written as:

$$\Phi_0(I(\xi)) = -\left(\frac{|I^{-1}(\xi)|}{|I^{-1}(\xi_0^*)|}\right)^{\frac{1}{3}}, \text{ and}$$

$$\Phi_i(I(\xi)) = -\frac{\text{tr}(c_i^T I^{-1}(\xi) c_i)}{\text{tr}(c_i^T I^{-1}(\xi_i^*) c_i)}, i = 1, 2.$$

Consider the following three-objective optimal design problem:

$$\begin{aligned} &\underset{\xi}{\text{Maximize}} && \text{Effi}_{\Phi_0(\xi)} \\ &\text{subject to} && \text{Effi}_{\Phi_i(\xi)} \geq 0.4, i = 1, 2. \end{aligned}$$

Utilizing the new algorithm, we find the corresponding compound optimal design is

$$L(\xi, \mathbf{U}^*) = \Phi_0 + 0.0916\Phi_1 + 0.0854\Phi_2.$$

Table 2 shows the efficiencies and computational time comparisons of the constrained optimal designs derived using grid search, sequential approach and new algorithm. The table clearly shows that both the new algorithm and the grid search produce a satisfactory solution. But the grid search takes around eighteen times

Table 2 Example I: relative efficiency of constrained optimal design based on different techniques

	Efficiency			
Techniques	Φ_0	Φ_1	Φ_2	Time cost (seconds)
Grid search	0.9761	0.4042	0.4009	1047
Sequential approach	Fails			
New algorithm	0.9761	0.4008	0.4046	59

calculation time as that of the new algorithm. On the other hand, for the sequential approach, all possible orders are tested and they all fail to provide a satisfactory solution.

5 Discussion

While the importance of multiple objective optimal designs is well recognized in scientific studies, their applications are still underdeveloped due to a lack of a general and efficient algorithm. The combination of the OWEA algorithm for compound optimal design problems and the new algorithm provides an efficient and stable framework for finding general multiple-objective optimal designs.

To guarantee convergence of the new algorithm, strict concavity of the objective function Φ_0 is required. However, various cases are tested and convergence holds for virtually all situations based on our experience. The new algorithm is implemented under local optimal designs context for all examples. It is possible to extend the results to other settings, like to the cases discussed in [5].

For optimal designs with no more than four objective functions, the new algorithm can efficiently derive the satisfactory solution. When there are five or more objective functions, it is unlikely that all constraints are active. If only less than four constraints are active, the new algorithm can still solve the optimal design problem efficiently. However, in the rare situation where there are four or more active constraints, the computational time can become lengthy. More research work is needed to deal with these cases.

Due to space limitation, many results in this paper are referred to [2]. The report can be accessed through the webpage:

http://homepages.math.uic.edu/~minyang/research.htm.

References

1. Atkinson, A.C., Chaloner, K., Juritz, J., Herzberg, A.M.: Optimum experimental designs for properties of a compartmental model. Biometrics **49**, 325–337 (1993)
2. Cheng, Q., Majumdar, D., Yang, M.: On multiple-objective nonlinear optimal designs. Technical report. Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago (2015)
3. Chernoff, H.: Locally optimal designs for estimating parameters. Ann. Math. Stat. **24**, 586–602 (1953)
4. Clyde, M., Chaloner, K.: The equivalence of constrained and weighted designs in multiple objective design problems. JASA **91**, 1236–1244 (1996)
5. Cook, R.D., Fedorov, V.V.: Constrained optimization of experimental design. Statistics **26**, 129–178 (1995)
6. Cook, R.D., Wong, W.K.: On the equivalence of constrained and compound optimal designs. JASA **89**, 687–692 (1994)

7. Huang, Y.C., Wong, W.K.: Sequential construction of multiple-objective optimal designs. Biometrics **54**, 1388–1397 (1998)
8. Yang, M., Biedermann, S., Tang, E.: On optimal designs for nonlinear models: a general and efficient algorithm. JASA **108**, 1411–1420 (2013)

Futher Reading

- Dette, H., Bretz, F., Pepelyshev, A., Pinheiro, J.: Optimal designs for dose-finding studies. JASA **103**, 1225–1237 (2008)
- Dette, H., Melas, V.B.: A note on the de la Garza phenomenon for locally optimal designs. Ann. Stat. **39**, 1266–1281 (2011)
- Dette, H., Schorning, K.: Complete classes of designs for nonlinear regression models and principal representations of moment spaces. Ann. Stat. **41**, 1260–1267 (2013)
- Huang, Y.C.: Multiple-objective optimal designs. Dr.P.H. dissertation, School of Public Health, Department of Biostatistics, University of California at Los Angeles (1996)
- Notari, R.E.: Biopharmaceutics and Clinical Pharmacokinetics. Introduction. Marcel Dekker, New York/Basel (1980)
- Rosenberger, W.F., Grill, S.E.: A sequential design for psychophysical experiments: an application to estimating timing of sensory events. Stat. Med. **16**, 2245–2260 (1997)
- Yang, M.: On the de la Garza phenomenon. Ann. Stat. **38**, 2499–2524 (2010)
- Yang, M., Stufken, J.: Support points of locally optimal designs for nonlinear models with two parameters. Ann. Stat. **37**, 518–541 (2009)
- Yang, M., Stufken, J.: Identifying locally optimal designs for nonlinear models: a simple extension with profound consequences. Ann. Stat. **40**, 1665–1681 (2012)

Design for Smooth Models over Complex Regions

Peter Curtis and Hugo Maruri-Aguilar

Abstract Smooth supersaturated models are a modelling alternative for computer experiments. They are polynomial models that behave like splines and allow fast computations. In this contribution we use a boxing approach together with Gram-Schmidt orthogonalization to model over complex regions. We then perform a two stage design and modelling strategy and apply our methodology in a complex example taken from the literature of soap film smoothing.

1 Introduction

Interpolating splines are the solution $y(x)$ that minimizes a measure of roughness $\Psi_m(y) = \int_a^b \{y^{(m)}(x)\}^2 dx$ when searching among all interpolating functions for a given data set. The solution $y(x)$ is an interpolating polynomial spline of degree $2m-1$, which for the case involving second derivatives is a cubic spline [6]. Thin-plate splines extend the theory of splines for multivariate x. They minimize an extension of the criterion Ψ_2 above, which in the bivariate case is $\Psi_2(y) = \int\int \left\{ \left(\frac{\partial^2 y}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 y}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 y}{\partial x_2^2}\right)^2 \right\} dx_1 dx_2$, where the integral is computed over all $\mathbb{R}^2$, see [9]. If interpolation is not required, spline smoothing may be used to model and the criterion to minimize is a linear combination of the roughness measure and the residual sum of squares. The literature for spline modelling and smoothing is vast, we refer the reader to [6, 14].

This paper is concerned with smooth supersaturated models (SSM) which are polynomial interpolators that minimize the same measure of roughness used for splines [4]. SSM are polynomials of high degree with more terms than the number of observations, and those extra degrees of freedom are used to achieve a spline-like behavior over a specified smoothing region $\mathscr{X}$. The smoothing region $\mathscr{X}$ for SSM can be arbitrarily defined to allow modelling over non-standard regions. Convergence of SSM to splines is guaranteed asymptotically [2], although in practice a good approximation is achieved after adding a few extra terms.

P. Curtis (✉) • H. Maruri-Aguilar
Queen Mary, University of London, Mile End Road, London E1 4NS, UK
e-mail: P.R.Curtis@qmul.ac.uk; H.Maruri-Aguilar@qmul.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_9

In this contribution we study SSM over a non-standard smoothing region $\mathscr{X}$. We compute over $\mathscr{X}$ defined as the union of rectangular intervals and build orthogonal bases using standard Gram-Schmidt method. This has the benefit of stabilizing computations. We design for smoothness using a response independent criterion. This paper is a first attempt to use SSM for non-standard regions, where spline and smoothing methodology is still largely under development, [12, 15, 16]. To the best of our knowledge no attempts on design have been produced for non-standard regions, and in this we differ from the literature on designs for spline models [8, 17].

The order of the paper is as follows. In Sect. 2 we review the multivariate SSM method. We discuss in Sect. 3 alternatives to tailor SSM for non-standard smoothing regions. We establish our design approach and illustrate it with an example in Sect. 4. This example is reworked to compare with soap film smoothing [16] where instead of Ψ_2, a distortion measure is averaged over a region of interest:

$$J(y) = \int\int \left(\frac{\partial^2 y}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2} \right)^2 dx_1 dx_2. \tag{1}$$

2 Smooth Supersaturated Model (SSM)

Consider the problem of finding an interpolator to n data points that minimizes a measure of roughness. Instead of searching over the space of functions with second derivatives, we use polynomial functions and thus the existence of second derivatives is guaranteed. Indeed the vector spaces implied by our models are Sobolev spaces [14]. They admit a seminorm induced by Ψ_2 as shown below.

2.1 Univariate Polynomial Formulation

The available data are $(x_1, y_1), \ldots, (x_n, y_n)$, where no two design points are repeated. Let the interpolator be the univariate polynomial in x with real coefficients $y(x) = f(x)^T\theta = \sum_{i=0}^{N-1} \theta_i x^i$, with $f(x)^T = (1, x, x^2, \ldots, x^{N-1})$ and $\theta^T = (\theta_0, \ldots, \theta_{N-1})$. We assume that $N > n$, i.e. the model has more terms than data points and the design-model matrix X is of size $n \times N$. To compute Ψ_2 we need the following definition.

Definition 1 Let $\mathscr{X} \subset \mathbb{R}$ be a closed, compact and bounded region; let $s(x)$ and $t(x)$ be univariate polynomial functions. We define $\langle s, t \rangle := \int_{\mathscr{X}} s''(x)t''(x)dx$.

The roughness of a polynomial $y(x)$ is $\Psi_2(y) = \langle y, y \rangle$. The following Lemma establishes the roughness of $y(x)$. We omit the proof.

Lemma 1 *Let $y(x), \theta$ and $f(x)$ be as above. Let K be the matrix of inner products $\langle x^i, x^j \rangle$ for $i, j = 0, \ldots, N-1$. Then*

1. *The roughness of* $y(x)$ *is* $\Psi_2(y) = \theta^T K \theta$, *and*
2. *The matrix* K *is of the form* $K = DDMD^T D^T$ *where* M *is the usual moment matrix of terms in* $f(x)$, *computed over the region* $\mathscr{X}$. *The matrix* D *is a square matrix of size* N *with zero entries apart from entry* $D_{i+1,i}$ *that equals* i *for* $i = 1, \ldots, N-1$.

2.2 Multivariate Polynomial Formulation

For multiple input variables, we use hierarchical polynomials, i.e. if a term is included in the polynomial, all its divisors are also included. For k input variables $x_1, \ldots, x_k$, a monomial is the power product $x^\alpha := x_1^{\alpha_1} \cdots x_k^{\alpha_k}$ where $\alpha_1, \ldots, \alpha_k$ are non-negative integer exponents collected in the exponent vector α. A hierarchical multivariate polynomial is written as

$$y(x) = \sum_{\alpha \in L} \theta_\alpha x^\alpha, \tag{2}$$

where L is a list of N exponents satisfying the hierarchy restriction and θ_α is the coefficient of monomial term x^α. For example, the list $L = \{(0,0), (1,0), (0,1), (1,1)\}$ implies $f(x)^T = (1, x_1, x_2, x_1 x_2)$ which with coefficients $\theta^T = (\theta_{00}, \theta_{10}, \theta_{01}, \theta_{11})$ defines the bivariate polynomial $f(x)^T \theta = \theta_{00} + \theta_{10} x_1 + \theta_{01} x_2 + \theta_{11} x_1 x_2$.

The measure of roughness is extended to

$$\Psi_2(y) = \int_{\mathscr{X}} tr\{H(y)^2\} dx \tag{3}$$

where $H(y)$ is the Hessian matrix of y and $\mathscr{X}$ is a closed, compact, fully dimensional and bounded region $\mathscr{X} \subset \mathbb{R}^k$. The restriction over the smoothing region $\mathscr{X}$ is required to use standard Riemann integration. As $\mathscr{X}$ is finite, the criterion $\Psi_2(y)$ corresponds to minimal energy splines rather than to thin-plate splines [10].

The inner product of Definition 1 generalizes to

$$\langle s, t \rangle := \int_{\mathscr{X}} \sum_{i,j} \frac{\partial^2 s(x)}{\partial x_i \partial x_j} \frac{\partial^2 t(x)}{\partial x_i \partial x_j} dx, \tag{4}$$

where the sum is performed over all pairs i, j taking values from $1, \ldots, k$. Lemma 1 can be generalized with K extending to inner products between terms in (2), i.e. the elements of K are $\langle x^\alpha, x^\beta \rangle$ for all pairs $\alpha, \beta \in L$; model roughness is the familiar $\Psi_2(y) = \langle y, y \rangle = \theta^T K \theta$ and K can be computed from moment matrix M.

The bilinear form $\langle \cdot, \cdot \rangle$ satisfies symmetry and non-negativity properties. However it is not a full scalar product over the space of polynomials as it is possible that this form equals zero despite using non-zero polynomial functions, e.g. $\langle 1, x_1 \rangle = 0$.

For our purposes, this fact does not mean that we cannot include terms with zero roughness, as the identifiability of such terms is guaranteed by the design-model matrix X under mild conditions.

2.3 Fitting the Interpolator to Data

The SSM is fitted to data $Y^T = (y_1, \ldots, y_n)$ by solving the linear system

$$\begin{pmatrix} 0 & X \\ X^T & K \end{pmatrix} \begin{pmatrix} \lambda \\ \theta \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix}. \tag{5}$$

Here the design has been extended using non replicated multivariate design points; the basis L is chosen so that the design-model matrix X is full rank and λ is a vector of Lagrange multipliers for the interpolation constraints $X\theta = Y$, see [2]. Existence of the solution of (5) is straightforward when K and X are full rank. When K is not full rank, a condition is imposed on the rank of X. This mild condition does not restrict the applicability of SSM. The inverse of the above matrix can be written as

$$\begin{pmatrix} 0 & X \\ X^T & K \end{pmatrix}^{-1} = \begin{pmatrix} -Q & H^T \\ H & P \end{pmatrix}, \tag{6}$$

where Q, H and P are matrices that satisfy $XH = I_n$, $XP = 0_{n,N}$ $X^T H^T + KP = I_N$, $X^T Q = KH$ and I_n, I_N are identity matrices of sizes n and N. The parameter estimates are $\theta^* = HY$; the fitted SSM is $y^*(x) = f(x)^T \theta^*$ and the minimum value of roughness achieved by this polynomial model is $\Psi_2^* = Y^T Q Y = \theta^{*T} K \theta^*$.

3 Smoothing over Complex Regions

Computer experiments often involve design regions which are products of simple intervals for each input so that the design region is a mutidimensional cube in $\mathbb{R}^k$. After linearly transforming individual variables to the unit interval, design and analysis is done over $[0, 1]^k$.

The SSM methodology requires the smoothing region $\mathscr{X}$ to be bounded and fully dimensional, but it does not require that the design and smoothing regions coincide. Indeed the SSM method can be used in cases when there is interest in smoothing over a different region from that where the experiment was performed. However we restrict attention to cases where we design and smooth over the same region. We briefly comment on possible approaches to deal with different scenarios of smoothing region $\mathscr{X}$. The region may consist of several disconnected components but each component is bounded and fully dimensional.

3.1 Boxing the Region

If $\mathscr{X}$ is not specified, a solution is to linearly transform each design variable to $[-1, 1]$ so the design lies in $\mathscr{X} = [-1, 1]^k$. We then replace each monomial x^α in the interpolator by the product of univariate Legendre polynomials with degrees given by entries of α. This linear change of basis from Equation (2) induces a checkerboard pattern in matrix K and has proved in practice to be numerically stable.

Alternatively, we compute K over each of a finite collection of non-overlapping "boxes" $\mathscr{X}_1, \ldots, \mathscr{X}_M$. Each box is a k-dimensional rectangle for which moments are computed using closed formulæ and the matrix K is the sum of the matrices computed for each box. If the region $\mathscr{X}$ is a union of boxes $\mathscr{X} = \bigcup_{i=1}^{M} \mathscr{X}_i$, then this method gives exact results. If $\mathscr{X}$ has an irregular shape, a cover of boxes can be used to approximate computations. In the simplest case, a single box would contain $\mathscr{X}$ and this coincides with translating to $[-1, 1]^k$. We advocate a coarse cover with a few boxes to include the smoothing region so that $\mathscr{X} \subset \bigcup_{i=1}^{M} \mathscr{X}_i$. Model oscillations close to the border of $\mathscr{X}$ can be dampened when smoothing over a larger region than the region of interest; after this initial approximation, there is no additional gain in refining the cover.

3.2 Polyhedral Regions

When $\mathscr{X}$ is a convex polyhedron, a triangulation method can be used so that $\mathscr{X}$ becomes the union of a finite number of simplices and computations over each simplex are added to obtain K. The use of triangulations is standard practice when fitting minimal energy splines [10] and a software implementation is available for polynomial integration over convex polyhedra, see [7]. When the region $\mathscr{X}$ is an arbitrary, non-convex polytope, triangulation may still be performed in some cases although this can be a much harder problem than over a convex region, see [1].

3.3 Gram-Schmidt Orthogonalization of SSM Bases

The linear system at the core of SSM can be improved depending on the choice of polynomial bases, and a more stable or sparse version of Equation (5) can be achieved by linearly transforming the basis. If the region can be transformed to $\mathscr{X} = [-1, 1]^k$, natural bases are products of Legendre polynomials. For a complex region such as a union of boxes, we create orthogonal polynomial bases using monomial terms of degree greater than or equal to two as input to a Gram-Schmidt algorithm that uses the inner product $\langle \cdot, \cdot \rangle$ of Equation (4). After orthogonalization, the matrix K becomes a diagonal matrix.

4 Designing for SSM in Complex Regions

Space filling designs such as Latin hypercubes are commonly used in computer experiments when the design region $\mathscr{X}$ can be transformed into the unit cube. Sampling from a larger cube and rejecting inadequate samples not in $\mathscr{X}$ can be useful when $\mathscr{X}$ is not very small relative to the cube generating the samples. A recent space-filling proposal for non-standard regions is to generate a large number of random points and the centroids of Ward's clustering method become design points. Designs built with this approach are termed FFF designs, see [11].

In order to design for smoothness criterion, recall that the observed roughness of the SSM is $\Psi_2^* = Y^T QY$ with Q as in (6). A simple proposal to design for smoothness is to minimize a function $\phi(Q)$ of the eigenvalues of the matrix Q and we suggest $\phi(Q) = \lambda_{max}(Q)$, i.e. to minimize the largest eigenvalue of Q, see [2]. The computation of matrix Q and its eigenvalues does not depend on the actual data values Y and there are at least two design alternatives. One is to select a design with points that minimize $\phi(Q)$. The alternative we pursued is that, given an initial design that is kept fixed, select a number of additional points that minimize $\phi(Q)$.

In the examples below we used the Ramsay horseshoe function [12, 16]. For our computations, the region $\mathscr{X}$ was covered with three coarse boxes, see Fig. 1 (left).

4.1 Design Using Roughness Ψ_2

An initial FFF design over $\mathscr{X}$ with $n = 35$ points was used to fit an SSM to data using $N = 140$ terms. The model we used has the 136 monomials of degree less than or equal to 15 and four monomials of degree 16. At this stage, 200 new FFF points were used to obtain an empirical measure of model fit with validation errors which produced RMSE $= 0.119$. This RMSE amounts to 1.43 % of the data range, well within an informal rule of thumb of 5 %.

To generate 5 new design points, a set of 200 FFF points, different from the validation set, was used as a pool to generate a sample of size five. The value

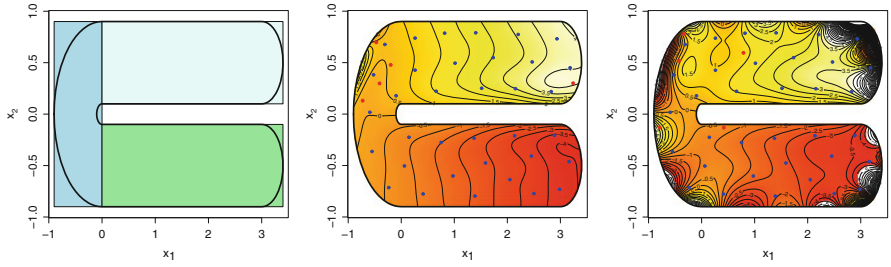

Fig. 1 Coarse boxing used and sequential fits Ψ_2 and J (from *left*)

of criterion $\phi(Q)$ was recorded and the procedure repeated 10,000 times and the best set of extra points was kept. The SSM with same number of terms and augmented design and data was validated using the same 200 validation points to yield RMSE $=$ 0.015, around 0.18 % of the data range. See Fig. 1 (centre) for depiction of the model at this stage.

4.2 Design for Distortion J

The distortion criterion $J(y)$ for soap film smoothing of Equation (1) defines a seminorm $\langle \cdot, \cdot \rangle$. We applied SSM techniques using this criterion with the same initial design and model terms as Sect. 4.1. The initial fit yielded validation RMSE $=$ 0.709, still relatively high as it is 8.5 % of data range. A search such as in Sect. 4.1 produced five new design points with RMSE $=$ 0.515 for the updated model (6.17 % of data range), see Fig. 1 (right).

A non-sequential soap film smoother [16] is superior with RMSE $=$ 0.0868 (1.04 % of data range). The performance of SSM under distortion J suffers from the fact that the matrix K for J has very low rank compared with the K matrix for roughness Ψ_2. This phenomenon creates aliasing of terms as very few terms are being used to minimize distortion and the method is thus both inefficient and prone to instability. In addition to this, our SSM-J procedure does not address the boundary conditions of soap film so this underperformance is to be expected. A potential solution to this is to use Hermite interpolators [5], which is possible when the boundary is an algebraic variety.

5 Discussion and Future Work

We provided a simple, response independent sequential design for smoothness. Our proposal uses orthonormal bases to enable the SSM method to be applied in non-standard regions. Coarse boxing provides the other element to achieve a simple and effective way of fitting SSM which worked well in the examples we tried.

The SSM method does not provide direct uncertainty quantification. An option is to perform polynomial regression under smoothness constraints [3]. Recent polynomial work by [13] could be alternatively adapted to embed SSM bases in a stochastic model, thus enabling quantification of uncertainty.

References

1. Baldoni, V., Berline, N., De Loera, J.A., Köppe, M., Vergne, M.: How to integrate a polynomial over a simplex. Math. Comput. **80**, 297–325 (2011)
2. Bates, R.A., Curtis, P.R., Maruri-Aguilar, H., Wynn, H.P.: Optimal design for smooth supersaturated models. J. Stat. Plan. Inference **154**, 3–11 (2014)
3. Bates, R.A., Giglio, B., Wynn, H.P.: A global selection procedure for polynomial interpolators. Technometrics **45**, 246–255 (2003)
4. Bates, R.A., Maruri-Aguilar, H., Wynn, H.P.: Smooth supersaturated models. J. Stat. Comput. Simul. **84**, 2453–2464 (2014)
5. Becker, T., Weispfenning, V.: The Chinese remainder problem, multivariate interpolation, and Gröbner bases. In: ISSAC '91: Proceedings of the 1991 International Symposium on Symbolic and Algebraic Computation, Bonn, pp. 64–69 (1991)
6. deBoor, C.: A Practical Guide to Splines. Volume 27 of Applied Mathematical Sciences, revised edn. Springer, New York (2001)
7. De Loera, J.A., Dutra, B., Köppe, M., Moreinis, S., Pinto, G., Wu, J.: Software for exact integration of polynomials over polyhedra. ACM Commun. Comput. Algebra **45**, 169–172 (2012)
8. Dette, H., Melas, V.B., Pepelyshev, A.: Optimal design for smoothing splines. Ann. Inst. Stat. Math. **63**, 981–1003 (2011)
9. Duchon, J.: Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. Rev. Francaise Automat. Informat. Recherche Operationnelle Ser. RAIRO Analyse Numerique **10**, 5–12 (1976)
10. Lai, M.-J.: Multivariate splines for data fitting and approximation. In: Neamtu, M., Schumaker, L.L. (eds.) Approximation Theory XII: San Antonio, pp. 210–228. Nashboro Press, Brentwood (2007)
11. Lekivetz, R., Jones, B.: Fast flexible space-filling designs for nonrectangular regions. Qual. Reliab. Eng. Int. **31**, 829–837 (2015)
12. Ramsay, T.: Spline smoothing over difficult regions. J. R. Stat. Soc. Ser. B Stat. Methodol. **64**, 307–319 (2002)
13. Tan, M.H.Y.: Stochastic polynomial interpolation for uncertainty quantification with computer experiments. Technometrics **57**, 457–467 (2015)
14. Wahba, G.: Spline Models for Observational Data. Volume 59 of CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1990)
15. Wang, H., Ranalli, M.G.: Low-rank smoothing splines on complicated domains. Biometrics **63**, 209–217 (2007)
16. Wood, S.N., Bravington, M.V., Hedley, S.L.: Soap film smoothing. J. R. Stat. Soc. Ser. B Stat. Methodol. **70**, 931–955 (2008)
17. Woods, D., Lewis, S.: All-bias designs for polynomial spline regression models. Aust. N. Z. J. Stat. **48**, 49–58 (2006)

PKL-Optimality Criterion in Copula Models for Efficacy-Toxicity Response

Laura Deldossi, Silvia Angela Osmetti, and Chiara Tommasi

Abstract In recent years, there has been an increasing interest in developing dose finding methods incorporating both efficacy and toxicity outcomes. It is reasonable to assume that efficacy and toxicity are associated; therefore, we need to model their stochastic dependence. Copula functions are very useful tools to model different kinds of dependence with arbitrary marginal distributions. We consider a binary efficacy-toxicity response with logit marginal distributions. Since the dose which maximizes the probability of efficacy without toxicity (P-optimal dose) changes depending on different copula functions, we propose a criterion which is useful for choosing between the rival copula models but also protects patients against doses that are far away from the P-optimal dose. The performance of this compromise criterion (called PKL) is illustrated for different choices of the parameter values.

1 Introduction

This work is focused on dose-finding designs based on joint models for both efficacy and toxicity. Such designs are known as phase I-II because they combine two phases of the clinical evaluation of a drug which are usually separate and consecutive. Phase I clinical trial designs aim at identifying the maximum tolerable dose of the drug on the basis of the toxicity. In phase I efficacy outcomes are observed but not used and only in phase II clinical trials, the efficacy of the drug is studied in order to decide whether further evaluation of the drug is warranted. In recent years, there has been an increasing interest in developing dose finding methods incorporating both efficacy and toxicity outcomes. According to this approach drug doses are acceptable only if they are safe and efficacious. Then, in general, the typical goal is to find the best dose which maximizes the probability of efficacy

L. Deldossi (✉) • S.A. Osmetti,
Dipartimento di Scienze statistiche, Università Cattolica del Sacro Cuore, L.go Gemelli 1, 20123 Milan, Italy
e-mail: laura.deldossi@unicatt.it; silvia.osmetti@unicatt.it

C. Tommasi,
Dipartimento DEMM, Università degli Studi di Milano, Via Conservatorio, 7, 20122 Milan, Italy
e-mail: chiara.tommasi@unimi.it

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_10

without toxicity. Herein, this dose is called P-optimal dose. In this context the relationship between the dose and the bivariate efficacy-toxicity response has to take into account the possible association between the two responses. For this purpose we explore copula functions which provide a rich and flexible class of dependence structures to obtain joint distributions for multivariate data, especially when standard multivariate distributions are not applicable. Differently from other dose-finding methods incorporating efficacy and toxicity outcome (see [3, 5, 10] among others) in this work we do not assume a specific dependence structure. We develop a robustness study to assess whether the P-optimal dose changes according to two different dependence structures. Since this occurs, we propose a compromise criterion (called PKL) which enables us to find doses which not only are good to discriminate between the rival copula models but also are not "far away" from the optimal dose for the patients.

The paper is organized as follows. In Sect. 2 the bivariate copula model is introduced and the main definitions are given. Section 3 describes the binary model for efficacy-toxicity response through copula functions. In Sect. 4, KL-optimality criterion is recalled and a compromise PKL-criterion is introduced. Finally, in Sect. 5 we study the robustness of the P-optimal dose with respect to a misspecified dependence structure and we compute the PKL-optimum design as the P-optimal dose is not robust.

2 Bivariate Copula-Based Model

Let (Y_1, Y_2) be a bivariate response variable with marginal distributions $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$, which depend on the unknown parameter vectors α and β, respectively. If there is an association between them, it is necessary to define a joint model for (Y_1,Y_2). In this paper a copula function is used to evaluate efficacy and toxicity simultaneously.

A bivariate copula is a function $C : I^2 \to I$, with $I^2 = [0,1]\times[0,1]$ and $I = [0,1]$, that, with an appropriate extension of the domain in R^2, satisfies all the properties of a cumulative distribution function (cdf). In particular, it is the cdf of a bivariate random variable (U_1, U_2), with uniform marginal distributions in $[0,1]$:

$$C(u_1,u_2;\theta) = P(U_1 \le u_1, U_2 \le u_2;\theta), \quad 0 \le u_1 \le 1 \quad 0 \le u_2 \le 1,$$

where θ is a parameter measuring the dependence between U_1 and U_2. The importance of copulae in statistical modelling stems from Sklar's theorem [8]. Let $\delta = (\alpha,\beta)$; according to Sklar's theorem, given a bivariate random variable (Y_1, Y_2) with joint cdf $F_{Y_1,Y_2}(y_1,y_2;\delta,\theta)$ and marginals $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$, there exists a copula function $C : I^2 \to I$ such that

$$F_{Y_1,Y_2}(y_1,y_2;\delta,\theta) = C\{F_{Y_1}(y_1;\alpha), F_{Y_2}(y_2;\beta);\theta\} \quad y_1,y_2 \in \mathbb{R}. \tag{1}$$

If $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$ are continuous functions then the copula $C(\cdot,\cdot;\theta)$ is unique. Conversely, if $C(\cdot,\cdot;\theta)$ is a copula function and $F_{Y_1}(y_1;\alpha)$ and $F_{Y_2}(y_2;\beta)$ are marginal cdfs, then $F_{Y_1,Y_2}(y_1,y_2;\delta,\theta)$ given in (1) is a joint cdf.

This theorem states that each joint distribution can be expressed in terms of marginal distributions and a function $C(\cdot,\cdot;\theta)$ that binds them together and so provides a general mechanism to construct new multivariate models in a straightforward manner. Thus, a copula captures the dependence structure between the marginal probabilities. This construction allows researchers to consider marginal distributions and the dependence between them as two separate but related issues. For each copula there exists a relationship between the parameter θ and the Kendall's τ or the Spearman ρ coefficients (see [8] pp. 158–170) and between θ and the lower and upper tail dependence parameters λ_l and λ_u (which measure the association in the tails of the joint distribution; see [8] pp. 214–216).

3 Binary Model for Efficacy and Toxicity

Let (Y_1, Y_2) be a binary efficacy-toxicity response variable; both Y_1 and Y_2 take values in $\{0, 1\}$ (1 denotes occurrence and 0 denotes no occurrence). Let $\pi_1(x;\alpha) = P(Y_1 = 1|x;\alpha)$ and $\pi_2(x;\beta) = P(Y_2 = 1|x;\beta)$ be the marginal success probabilities of efficacy and toxicity and $x \in \mathscr{X}$ denotes the dose of a drug. In this work, we consider a logistic model for both Y_1 and Y_2.

It is commonly accepted that efficacy and toxicity increase with dose. However, for efficacy, to allow a wide variety of possible dose-response relationships (including non-monotonic functions) a logistic model with a quadratic term is reasonably flexible [10]. Then, we assume the following logistic models for efficacy and toxicity:

$$\text{logit}\{\pi_1(x;\alpha)\} = \alpha_0 + \alpha_1 x + \alpha_2 x^2, \quad \alpha = (\alpha_0, \alpha_1, \alpha_2), \tag{2}$$

$$\text{logit}\{\pi_2(x;\beta)\} = \beta_0 + \beta_1 x, \quad \beta = (\beta_0, \beta_1). \tag{3}$$

A copula approach is applied to define a bivariate binary logistic model for the efficacy-toxicity response. Let

$$p^C_{y_1,y_2}(x;\delta,\theta) = P(Y_1 = y_1, Y_2 = y_2|x;\delta,\theta) \quad \text{for} \quad y_1, y_2 = 0, 1, \tag{4}$$

be the joint probability of (Y_1,Y_2) for a given experimental condition x. The superscript C denotes a copula family. From the copula representation (1), we define

$$p^C_{11}(x;\delta,\theta) = P(Y_1 = 1, Y_2 = 1|x;\delta,\theta) = C\{\pi_1(x;\alpha), \pi_2(x;\beta);\theta\}, \tag{5}$$

Table 1 Copula functions and corresponding lower and upper tail dependence parameters (λ_l, λ_u)

Copula	$C(u_1, u_2; \theta)$	$\theta \in \Theta$	λ_l, λ_u
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$	$(0, \infty)$	$\lambda_l = 2^{(-1/\theta)}, \lambda_u = 0$
Gumbel	$\exp\left(-\left[\{-\ln(u_1)\}^\theta + \{-\ln(u_2)\}^\theta\right]^{1/\theta}\right)$	$[1, \infty)$	$\lambda_l = 0, \lambda_u = 2 - 2^{(1/\theta)}$

where C is a copula function which models the dependence between $\pi_1(x; \alpha)$ and $\pi_2(x; \beta)$ given in (2) and (3), respectively. Equation (5) actually defines a class of models for the binary response: specifing $C(\cdot, \cdot; \theta)$ it provides a particular model.

From equation (5) we have that the probability of efficacy without toxicity, for a given experimental condition x, is

$$p_{10}^C(x; \delta, \theta) = P(Y_1 = 1, Y_2 = 0 | x; \delta, \theta) = \pi_1(x; \alpha) - p_{11}^C(x; \delta, \theta). \tag{6}$$

Clinicians are interested in finding the P-optimal dose, which maximizes the probability of efficacy without toxicity, i.e.

$$x_p^* = \arg\max_{x \in \mathcal{X}} p_{10}^C(x; \delta, \theta). \tag{7}$$

This is a deterministic problem that can be solved whenever the model for the data is known. For a fixed parameter vector δ (a scenario), however, the P-optimal dose could change for different copula functions and/or for different values of the dependence parameter θ.

Several bivariate copulae have been proposed in the literature (see for instance [8]). In this paper we consider only Clayton and Gumbel copulae, which have been already applied by [2] and [9] in the context of Optimal Design. Both these copulae allow only for positive association between variables even if they exhibit strong *left* and strong *right* tail dependence, respectively. Their main characteristics are summarized in Table 1.

Before finding the P-optimal dose, it is necessary at first to identify the correct dependence structure $C(\cdot, \cdot; \theta)$ in (5) and then to estimate the parameters of the chosen model. In the following section we propose a criterion to reach the first goal of model identification.

4 PKL-Optimality Criterion

An approximate design ξ is a discrete probability measure on a compact experimental domain $\mathcal{X}$; $\xi(x)$ represents the amount of experimental effort at the support point x. An optimal design maximizes a concave optimality criterion function which reflects an inferential goal.

Let (Cl, G) denote Clayton and Gumbel copulae and let (θ_{Cl}, θ_G) be the corresponding dependence parameters. From now on, we assume that nominal values

for both model and dependence parameters (i.e. δ and θ) are available; hence, we compute locally optimum designs. In order to discriminate between the two rival copulae, we could use the following geometric mean of KL-efficiencies:

$$\Phi_{KL}(\xi;\delta,\theta_{Cl},\theta_G) = \{\text{Eff}_{G,Cl}(\xi;\theta_{Cl})\}^{\gamma_1} \cdot \{\text{Eff}_{Cl,G}(\xi;\theta_G)\}^{1-\gamma_1} \quad 0 \le \gamma_1 \le 1,$$

where

$$\text{Eff}_{i,j}(\xi;\theta_j) = \frac{I_{i,j}(\xi;\theta_j)}{I_{i,j}(\xi^*_{i,j};\theta_j)}, \quad \xi^*_{ij} = \arg\max_{\xi} I_{i,j}(\xi;\theta_j), \quad i,j = Cl,G. \tag{8}$$

The function

$$I_{i,j}(\xi;\theta_j) = \inf_{\theta_i} \sum_{x\in\mathscr{X}} \mathscr{I}\{p^j_{y_1y_2}(x;\delta,\theta_j), p^i_{y_1y_2}(x;\delta,\theta_i)\}\,\xi(x), \tag{9}$$

is the KL-criterion proposed by [6]. Here $\mathscr{I}\{p^j_{y_1y_2}(x;\delta,\theta_j), p^i_{y_1y_2}(x;\delta,\theta_i)\}$ denotes the Kullback-Leibler divergence between the true model $p^j_{y_1y_2}(x;\delta,\theta_j)$ and the rival one $p^i_{y_1y_2}(x;\delta,\theta_i)$, with $i,j = Cl,G$.

Since the KL-criterion is invariant with respect to linear transformation of the design space (see [1]), we standardize the dose x as

$$d = \frac{x - (x_{min} + x_{max})/2}{(x_{max} - x_{min})/2}. \tag{10}$$

Hence, the design region $\mathscr{X}$ becomes the interval $\mathscr{D} = [-1, 1]$ and the intermediate dose x corresponds to $d = 0$. In clinical trials, x_{min} and x_{max} in (10) usually are the minimum effective dose (MED) and the maximum tolerated dose (MTD), i.e. the extremes of a therapeutic region (of course MED and MTD depend on the nominal value for model parameter vector δ). From now on we consider design in $\mathscr{D}$.

Doses which are optimal for discrimination purposes, however, could be very different from the best dose for the patients. Therefore, we propose to penalize criterion $\Phi_{KL}(\xi;\delta,\theta_{Cl},\theta_G)$ through the following geometric mean of P-efficiencies:

$$\Phi_P(\xi;\delta,\theta_{Cl},\theta_G) = \left\{\text{Eff}^P_{Cl}(\xi;\theta_{Cl})\right\}^{\gamma_2} \cdot \left\{\text{Eff}^P_{G}(\xi;\theta_G)\right\}^{1-\gamma_2}, \quad 0 \le \gamma_2 \le 1, \tag{11}$$

where a P-efficiency is defined as

$$0 \le \text{Eff}^P_C(\xi;\delta,\theta_C) = \frac{\Phi^P_C(\xi;\delta,\theta_C)}{\Phi^P_C(\xi^*_p;\delta,\theta_C)} = \frac{\Phi^P_C(\xi;\delta,\theta_C)}{p^C_{10}(d^*_p;\delta,\theta_C)} \le 1 \tag{12}$$

and

$$\Phi^P_C(\xi;\delta,\theta_C) = \sum_{d\in\mathscr{D}} p^C_{10}(d;\delta,\theta_C)\,\xi(d), \quad C = Cl,G$$

is the probability of efficacy without toxicity (see [7]). The P-efficiency $\mathrm{Eff}_C^P(\xi;\delta,\theta)$ measures the goodness of a design ξ with respect to $\xi_p^* = \arg\max_\xi \Phi_C^P(\xi;\delta,\theta_C)$; let us remark that $\xi_p^* = \xi_{d_p^*}$, that is the design with the whole mass at the P-optimal dose $d_p^* = \arg\max_{d\in\mathscr{D}} p_{10}^C(d;\delta,\theta_C)$ (this justifies the second equality in (12)).

In other terms, in order to discriminate between the two rival copulae, it is ethically more convenient to consider the following compromise criterion (called PKL):

$$\Phi(\xi;\delta,\theta_{Cl},\theta_G) = \{\Phi_{KL}(\xi;\delta,\theta_{Cl},\theta_G)\}^{\gamma_3} \cdot \{\Phi_P(\xi;\delta,\theta_{Cl},\theta_G)\}^{1-\gamma_3}, \quad 0 \le \gamma_3 \le 1. \tag{13}$$

The corresponding PKL-optimum design, i.e.

$$\xi^* = \arg\max_\xi \Phi(\xi;\delta,\theta_{Cl},\theta_G),$$

can be computed applying the first order algorithm (see [4], §3.2). The directional derivative of the PKL-criterion can be obtained using [6] and [7]. Mathematical details are available upon request by contacting the third author. The PKL-criterion should ensure that doses, which are good for discrimination purposes, are also "near" to the best dose for the patients.

If a researcher wants to consider both the problems of model discrimination and parameter estimation (see [9]) at the same design stage, then a DKL-criterion should be used (see [11]). Even in this case we suggest to penalize with respect to (11).

5 Simulation

In order to assess whether P-optimal doses depend on the assumed dependence structure, we apply (7) using Clayton and Gumbel copulae, respectively, and different values of the dependence parameter. Table 2 lists the considered values for θ, along with the corresponding Kendall τ and Spearman's ρ coefficients and the lower and upper tail dependence coefficients, λ_l and λ_u.

These computations are developed for 6 different choices of $\delta = (\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$. Scenario 1: $\delta = (-1, 1, 0, -2.5, 0.5)$; Scenario 2: $\delta = (-1, 3, 0, -1, 4)$; Scenario 3: $\delta = (-0.5, 1, 0, 0, 1)$; Scenario 4 : $\delta = (-1, 3, -5, -1.5, 3)$; Scenario 5: $\delta = (0.5, 3, -3, -1.5, 3)$; Scenario 6: $\delta = (3, 2, -1, 1, 1)$.

In this Section, we report only the results for Scenarios 2 and 3, for which the P-optimum dose d_p^* changes under the two rival copulae and the different values of the dependence parameter.

Let d_Π^* be the P-optimal dose corresponding to the product copula Π (independence copula). In order to assess the effect of the dependence structure on the P-optimal dose, in Table 3 we compute the P-efficiency of $\xi_{d_\Pi^*}$ with respect to the P-optimal doses under both the rival copulae.

Table 2 Copula parameter values and the related dependence and tail dependence coefficients

Copula	θ	τ	ρ	λ_l	λ_u
Clayton	2	0.500	0.683	0.707	0.000
	8	0.800	0.941	0.917	0.000
	20	0.909	0.987	0.966	0.000
Gumbel	2	0.500	0.683	0.000	0.586
	5	0.800	0.943	0.000	0.851
	10	0.900	0.986	0.000	0.928

Table 3 P-optimal dose d_p^*, success probability $p_{10}^C(\cdot)$ and efficiency $\text{Eff}_C^P(\xi_{d_{PI}^*})$

Copula	θ	Scenario 2			Scenario 3		
		d_p^*	$p_{10}^C(d_p^*;\delta,\theta)$	$\text{Eff}_C^P(\xi_{d_{PI}^*})$	d_p^*	$p_{10}^C(d_p^*;\delta,\theta)$	$\text{Eff}_C^P(\xi_{d_{PI}^*})$
Clayton	2	0.325	0.090	0.971	1	0.084	0.828
	8	−0.318	0.032	0.518	1	0.017	0.394
	20	−0.356	0.031	0.062	1	0.001	0.120
Gumbel	2	0.037	0.113	0.918	−0.187	0.076	0.959
	5	−0.156	0.052	0.580	−0.936	0.015	0.749
	10	−0.276	0.036	0.235	−0.999	0.002	0.433
Independence copula		0.199	0.221	1	0.250	0.192	1

Table 4 P-efficiency of d_p^* obtained under a misspecified copula function

		Scenario 2						Scenario 3					
Copula	false	Clayton			Gumbel			Clayton			Gumbel		
true	θ	2	8	20	2	5	10	2	8	20	2	5	10
Clayton	2	1	0.548	0.514	0.867	0.696	0.586	1	1	1	0.654	0.373	0.353
	8	0.493	1	0.995	0.656	0.905	0.993	0.999	1	1	0.202	0.063	0.058
	20	0.038	0.993	1	0.218	0.751	0.965	1	1	1	0.032	0.004	0.004
Gumbel	2	0.761	0.720	0.675	1	0.900	0.771	0.740	0.740	0.740	1	0.889	0.872
	5	0.351	0.922	0.888	0.864	1	0.956	0.477	0.477	0.477	0.890	1	0.999
	10	0.083	0.993	0.977	0.583	0.938	1	0.206	0.206	0.206	0.624	0.975	1

In addition, in order to study the "robustness" of the P-optimal dose when a wrong dependence structure is assumed, Table 4 provides the P-efficiency of the P-optimal dose obtained under the misspecified copula with respect to the P-optimal dose under the true one.

From Tables 3 and 4, we may conclude that there are losses in P-efficiency when we do not take into account the dependence structure and also when we misspecify it. As a consequence, it seems relevant to discriminate between the rival copulae. Since the support points of the KL-optimum design are frequently "far away" from the P-optimal dose, we compute the PKL-optimum design ξ^*. Table 5 reports ξ^* for different values of θ_{Cl} and θ_G (for which we have the same association level) and for $\gamma_1 = \gamma_2 = 1/2$ and $\gamma_3 = 1/3$. In addition, both the KL- and P-efficiencies of ξ^* are

Table 5 PKL-optimal designs and their efficiencies with respect to the KL- and P-optimum designs

θ_{Cl}	θ_G	Scenario 2					Scenario 3				
		ξ^*	$Eff_{G,Cl}$	$Eff_{Cl,G}$	Eff^P_{Cl}	Eff^P_G	ξ^*	$Eff_{G,Cl}$	$Eff_{Cl,G}$	Eff^P_{Cl}	Eff^P_G
2	2	$\begin{Bmatrix} -0.10 & 0.651 \\ 0.568 & 0.432 \end{Bmatrix}$	0.766	0.641	0.766	0.663	$\begin{Bmatrix} -1 & 1 \\ 0.409 & 0.591 \end{Bmatrix}$	0.911	0.998	0.735	0.794
8	5	$\begin{Bmatrix} -0.250 & 0.850 \\ 0.689 & 0.311 \end{Bmatrix}$	0.746	0.717	0.871	0.672	$\begin{Bmatrix} -1 & 1 \\ 0.435 & 0.565 \end{Bmatrix}$	0.927	0.973	0.590	0.704
20	10	$\begin{Bmatrix} -0.3 & 1 \\ 0.739 & 0.261 \end{Bmatrix}$	0.585	0.681	0.737	0.772	$\begin{Bmatrix} -1 & 1 \\ 0.481 & 0.519 \end{Bmatrix}$	0.985	0.967	0.521	0.589

given. We observe that the compromise optimum design ξ^* seems to perform quite well both to discriminate between dependence structures and to guarantee a quite high probability of efficacy without toxicity.

References

1. Aletti, G., May, C., Tommasi, C.: KL-optimum designs: theoretical properties and practical computation. Stat. Comput. (2014). doi:10.1007/s11222-014-9515-8
2. Denman, N.G., McGree, J.M., Eccleston, J.A., Duffull, S.B.: Design of experiments for bivariate binary responses modelled by copula functions. Comput. Stat. Data Anal. **55**, 1509–1520 (2011)
3. Dragalin, V., Fedorov, V., Wu, Y.: A two-stage design for dose-finding that accounts for both efficacy and safety. Stat. Med. **27**, 5156–5176 (2008)
4. Fedorov, V., Hackl, P.: Model-Oriented Design of Experiments. Springer, New York (1997)
5. Gao, L., Rosenberger, W.F.: Adaptive Bayesian design with penalty based on toxicity-efficacy response. In: Ucinski, D., Atkinson, A.C., Patan, M. (eds.) mODa 10 – Advances in Model-Oriented Design and Analysis, pp. 91–98. Springer International Publishing, Heidelberg (2013)
6. López-Fidalgo, L.J., Tommasi, C., Trandafir, P.C.: An optimal experimental design criterion for discriminating between non-normal models. J. R. Stat. Soc. B **69**(2), 231–242 (2007)
7. McGree, J.M., Eccleston, J.A.: Probability-based optimal design. Aust. N. Z. J. Stat. **50**, 13–28 (2008)
8. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (2006)
9. Perrone, E., Müller, W.G.: Optimal designs for copula models. In: IFAS Research Paper Series 2014–66, Department for Applied Statistics, Johannes Kepler University Linz. http://www.jku.at/ifas/content/e108280/e146255/e241527/ifas_rp66.pdf. Cited May 2014
10. Thall, P.F., Cook, J.D.: Dose-finding based on efficacy-toxicity trade-offs. Biometrics **60**, 684–693 (2004)
11. Tommasi, C.: Optimal designs for both model discrimination and parameter estimation. J. Stat. Plan. Inference **139**, 4123–4132 (2009)

Efficient Circular Cross-over Designs for Models with Interaction

Pierre Druilhet

Abstract Cross-over designs are widely used in many areas such as clinical trials or agricultural field experiments. In addition to the effect of the treatment applied to a given period, a residual effect of the previous treatment is often observed. In that case, the effect observed depends on both treatments. When the aim of the experiment is to select a single treatment, it is natural to consider the effect of a treatment preceded by itself, also called the total effect. We consider universally optimal or efficient circular cross-over designs for total effects under a model with interaction between treatment and carry-over effects. The optimal designs are obtained from the theory of approximate designs developed by Kushner (Ann Stat 25:2328–2344, 1997).

1 Introduction

In cross-over designs, each experimental unit or subject receives successively several treatments. The main advantage is to reduce or eliminate the subject variability in the estimation of treatment effects. However, each response may be perturbed by a residual or carry-over effect induced by the treatment applied to the previous period. Neglecting this interference in the analysis may cause bias or extra-variability under randomization in the estimation of direct treatment effects: [2] and [16] showed that cross-over designs with neighbor balance properties may reduce this bias or variability. There are several ways to incorporate carry-over effects in the model. The most parsimonious is through an additive effect. Many results on optimal designs under this model are available, e.g. [14, 17, 20, 21] or [26] for designs without pre-period; for circular designs, we can cite [3, 5, 9, 10, 15] or [8].

To be more accurate, one may incorporate in the model a specific effect when a treatment is preceded by itself, such as in [1, 6, 7, 18, 19] or [25]. For a full interaction between carry-over and treatment effects, [24] and [22] propose optimal designs for the estimation of direct treatment effects, i.e. the effect of a treatment

P. Druilhet (✉)
Laboratoire de Mathématiques, Université Blaise Pascal, UMR 6620 – CNRS, France
e-mail: pierre.druilhet@univ-bpclermont.fr

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_11

averaged over the interaction with all the carry-over effects. In that case, direct treatment effects depend on the other treatments through the interactions. When the aim of the experiment is to select a single treatment, it may be more relevant to consider total effects, i.e. effects of treatments preceded by themselves, which do not depend on the other treatments. Bailey and Druilhet [4], abbreviated by BD in the following, obtained optimal non-circular designs for the estimation of such effects under the same model.

In this paper, we consider circular designs that are optimal for the estimation of total effects. In Sect. 2, we give the main ideas of the method proposed by [20] and its generalizations. To make the logical flow easier to follow, we skip technical results. Instead, we refer to equivalent results in BD since the proofs are similar even if the permutation group involved in the circular case is slightly different than that of the non-circular one. In Sect. 3, we construct optimal designs by solving a minimax problem. In Sect. 4, we give some examples of optimal sequences and optimal or efficient designs.

2 Interaction Repeated Measurement Models and Information Matrices

Denote by $\Omega_{t,n,p}$ the set of circular cross-over designs with t treatments, n subjects and k periods. Let $d(i,j)$ be the treatment assigned by the design d in $\Omega_{t,n,k}$ in the jth period to the ith subject, $1 \leq i \leq n$ and $1 \leq j \leq k$. The pre-period corresponds to $j = 0$ and the circularity condition is given by $d(i,0) = d(i,k)$.

The interaction model for the response y_{ij} is:

$$y_{ij} = \beta_i + \xi_{d(i,j),d(i,j-1)} + \varepsilon_{ij}, \tag{1}$$

where β_i is the effect of subject i and ξ_{uv} is the effect of treatment u when preceded by treatment v. We omit here a period effect in the model, which will be shortly discussed in Sect. 3. The residual errors ε_{ij} are assumed to be independent identically distributed with expectation 0 and variance σ^2. Denote by ξ the vector

$$\xi = (\xi_{11}, \xi_{12}, .., \xi_{1t}, \xi_{21}, \ldots, \xi_{tt})'.$$

Let ϕ_u be the total effect of treatment u, that is the effect of treatment u which is preceded by itself. We have $\phi_u = \xi_{uu}$ and in vector notation

$$\phi = K'\xi,$$

where K is the $t^2 \times t$ matrix with entries $K^w_{uv} = 1$ if $u = v = w$ and 0 otherwise, for $u, v, w = 1, \ldots, t$, where w is the single index for the columns and uv is the double index for the rows, similar to the double index for the vector ξ.

In vector notation, Model (1) can be written:

$$Y = B\beta + X_d\xi + \varepsilon,$$

where Y is the nk-vector of responses with entries y_{ij} in lexicographic order, X_d is the $nk \times t^2$ design matrix, β is the n-vector of subject effects and $B = I_n \otimes 1_k$ with I_n the identity matrix of order n and 1_k is the k-vector of ones.

The information matrix $C_d[\xi]$ for the estimation of vector ξ is given by (see e.g. [13]):

$$C_d[\xi] = X_d'\,\omega_B^{\perp}\,X_d.$$

where $\omega_B^{\perp} = I_{nk} - B(B'B)^{-1}B' = I_{nk} \otimes Q_k$ with $Q_k = I_k - J_k/k$ and J_k the $k \times k$ matrix of ones. Note that $\omega_B^{\perp} X_d 1_{t^2} = \omega_B^{\perp} 1_{nk} = \mathbf{0}$, and so

$$C_d[\xi]\,1_{t^2} = \mathbf{0}. \tag{2}$$

Let S be the set of all possible sequences of k treatments. For s in S, we denote by X_s the design matrix associated to s and $C_s[\xi] = X_s'Q_kX_s$ the corresponding information matrix. One key idea of Kushner's [20] methods is to express $C_d[\xi]$ as a linear combination of $C_s[\xi]$:

$$C_d[\xi] = n\sum_{s\in S}\pi_d(s)\,C_s[\xi] = n\sum_{s\in S}\pi_d(s)\,X_s'\,Q_k\,X_s, \tag{3}$$

where $\pi_d(s)$ is the proportion of subjects that receive the sequence s of treatments. In regard to the information matrix, an exact design is characterized by the proportions $\pi_d(s)$ which are necessarily proportional to n^{-1}. The second key idea of Kushner's methods is to relax this constraint and to seek an optimal design among approximate designs d in $\Omega_{t,n,k}$ which are defined by their proportions $\pi_d(s)$, s in S, with $0 \le \pi_d(s) \le 1$ and $\sum_{s\in S}\pi_d(s) = 1$.

The information matrix $C_d[\phi]$ of ϕ may be obtained from $C_d[\xi]$ by the extremal representation (see [11] or [23, p. 62]):

$$C_d[\phi] = C_d\left[K'\xi\right] = \min_{L\in\mathscr{L}_K} L'C_d[\xi]\,L, \tag{4}$$

where $\mathscr{L}_K = \{L \in R^{t^2\times t} \mid L'K = I_t\}$ and the minimum is taken relative to the Loewner ordering. It can be shown that $C_d[\phi]1_t = 0$ (see Lemma 1 in BD). Denoting by L_d^* a matrix L that achieves the minimum in (4), we have

$$C_d[\phi] = L^{*\prime}\,C_d[\xi]\;L^* = n\sum_{s\in\mathscr{S}}\pi_d(s)\,L^{*\prime}\,C_s[\xi]\;L^*. \tag{5}$$

For a sequence s in S, we denote by s_σ the sequence obtained from s by permuting the treatment labels according to the permutation σ. A design d is said to be

symmetric if, for any sequence s and any permutation σ, $\pi_d(s_\sigma) = \pi_d(s)$. From a design d, the *symmetrized design* $\bar{d}$ is defined by

$$\pi_{\bar{d}}(s) = \frac{1}{t!}\sum_{\sigma\in S_t}\pi_d(s_\sigma), \quad \forall s \in \mathscr{S}, \tag{6}$$

where S_t is the set of all permutations on $\{1,\dots,t\}$. To a permutation σ of the treatment labels corresponds a permutation $\tilde{\sigma}$ of the indices of ξ where $\tilde{\sigma}(u,v) = (\sigma(u),\sigma(v))$. An important requirement to apply Kushner's methods is that the matrix K that defines the parameter of interest ϕ must satisfy $P_{\tilde{\sigma}}KP_\sigma = K$ which is the case here (note that this is also the case for direct treatment effects). A first consequence is that a symmetrized design $\bar{d}$ has an information matrix for ϕ which is completely symmetric and whose trace is at least as large as that of d (see Lemma 3 in BD). This implies that a universally optimal approximate design may be sought among symmetric designs. The second consequence is that the matrix L_d^* defined in (5) has a simple parameterized form when d is symmetric. This considerably simplifies the minimization problem. Indeed, for a symmetric design, the entries of L_d^* can be chosen constant on the triple orbits of the permutation group S_t on $\{1,\dots,t\}$, i.e. $L^w_{uv} = L^{\sigma(w)}_{\sigma(u)\sigma(v)}$, for all $\sigma \in S_t$ (see Lemma 4 in BD). So, L_d^* can be written as a linear combination of 5 matrices:

$$L_d^* = \sum_{q=1}^{5} x_{dq}\, L_{(q)} \tag{7}$$

where $L^u_{(1)uu} = 1$ for $u = 1,\dots,t$ and 0 otherwise; $L^u_{(2)uv} = 1$, $L^v_{(3)uv} = 1$ and $L^v_{(5)uu} = 1$ for $u,v = 1,\dots,t$, $u \neq v$ and 0 otherwise; $L^w_{(4)uv} = 1$ for $u,v,w = 1,\dots,t$, $u \neq v \neq w \neq u$ and 0 otherwise. The constraint $L_d^{*'}K = I_t$ can be written $x_{d1} = 1$ and $x_{d5} = 0$ and therefore L_d^* depends only on three free parameters.

3 Universally Optimal Approximate Designs

From [12], a sufficient condition for a design d to be universally optimal over a class D of design is that $C_d[\phi]$ is completely symmetric and maximizes the trace over D. We saw in Sect. 2 that a symmetric design with maximal trace will be universally optimal among all possible approximate designs. So, let d be a symmetric design. Combining (4), (5) and (7), we have

$$\mathrm{tr}(C_d[\phi]) = n \min_{x_2,\dots,x_4} \sum_{s\in S}\pi_d(s)\sum_{p=1}^{4}\sum_{q=1}^{4} x_p x_q\, c_{spq} \quad \text{with } x_1 = 1, \tag{8}$$

where $c_{spq} = \mathrm{tr}\left(L'_{(p)}C_s[\xi]L_{(q)}\right)$.

We say that two sequences are equivalent if one may be obtained from the other one by permutation of the treatment labels and/or by circular permutation of the periods. The coefficient s is the same for two sequences in the same equivalence class, so (8) can be written

$$\mathrm{tr}(C_d[\phi]) = n \min_{x_2,\ldots,x_4} \sum_{\ell\in\mathscr{L}} \pi_d(\ell) \sum_{p=1}^{4}\sum_{q=1}^{4} x_p x_q\, c_{\ell pq} \quad \text{with } x_1 = 1, \tag{9}$$

where $\mathscr{L}$ is the set of equivalence classes and $\pi_d(\ell)$ is the proportion of sequences in design d that belong to the equivalence class ℓ. So, a symmetric d^* is universally optimal if its proportions $\pi_{d^*}(\ell)$, $\ell \in \mathscr{L}$, maximize (9). To solve this problem we consider the following procedure derived from [20]. Let $h_\ell(\gamma) = \sum_{p=1}^{4}\sum_{q=1}^{4} x_p x_q c_{\ell pq}$ with $\gamma = (x_2, x_3, x_4)$ and $h^*(\gamma) = \max_\ell h(\ell)$.

Step 1 Find γ^* that minimizes the function $h^*(\gamma)$ and let $h^* = h^*(\gamma^*)$ denote the minimum.

Step 2 Select the classes ℓ of sequences such that $h_\ell(\gamma^*) = h^*$ and let $\mathscr{C}^*$ denote this set.

Step 3 Solve in $\{\pi_\ell \mid \ell \in \mathscr{C}^*\}$ the linear system, $\sum_{\ell\in\mathscr{C}^*} \pi_\ell \frac{dh_\ell}{d\gamma}(\gamma^*) = 0$, for $0 < \pi_\ell < 1$ and $\sum_{\ell\in\mathscr{C}} \pi_\ell = 1$; denote by $\pi^* = \{\pi_\ell^* \mid \ell \in \mathscr{C}^*\}$ the solution (not necessarily unique).

Step 4 Give the symmetric designs such that $\pi_\ell = \pi_\ell^*$ for $\ell \in \mathscr{C}^*$ and $\pi_\ell = 0$ otherwise; these designs are universally optimal.

Let now assume that there are period effects in the model. A design is said to be *strongly balanced* on the period if, for any given period, each treatment is preceded by itself equally often, and is preceded by any other treatment equally often. For a approximate design, "equally often" must be replaced by "in the same proportion". As in BD (Prop. 8 and following remark) a symmetric exact or approximate design which is optimal under the model without period effect and which is strongly balanced on the period is also optimal under the model with period effects.

4 Examples of Optimal Designs

We give here some examples of optimal sequences.

4.1 k = 4, t = 4

An optimal design is generated by the sequence (1 1 2 2). For example, the following design obtained by considering all the possible permutations of treatment labels from the optimal sequence is symmetric and strongly balanced on the periods,

and so is universally optimal

$$\begin{pmatrix} 1&1&1&2&2&2&3&3&3&4&4&4\\ 1&1&1&2&2&2&3&3&3&4&4&4\\ 2&3&4&1&3&4&1&2&4&1&2&3\\ 2&3&4&1&3&4&1&2&4&1&2&3 \end{pmatrix}.$$

Since the sequences $(a\ a\ b\ b)$ are equivalent, we can reduce the design to

$$\begin{pmatrix} 1&1&1&2&2&3\\ 1&1&1&2&2&3\\ 2&3&4&3&4&4\\ 2&3&4&3&4&4 \end{pmatrix}.$$

However, this design is no longer optimal under the model with period effects.

4.2 k = 5, t = 5

The optimal design is generated by the sequences (1 1 2 2 2) and (1 1 2 2 3) with proportions respectively $7/10$ and $3/10$. A symmetric design with these proportions needs 480 subjects. So, it may be preferable to consider a symmetric design generated by a single sequence. A symmetric design generated by the first sequence has an efficiency factor equal to 0.97 and requires at least 20 subjects. A symmetric design generated by the second sequence has an efficiency factor equal to 0.92 and requires 60 subjects.

4.3 k = 6, t = 6

The optimal design is generated by the sequences (1 1 1 2 2 2) and (1 1 2 2 3 3) with proportions respectively $3/77(63-16\sqrt{14}) \approx 0.12$ and $16/77(3\sqrt{14}-7) \approx 0.88$. A symmetric design generated by the first sequence has an efficiency factor equal to 0.92. For the second sequence, it is 0.999.

4.4 k = 7, t = 7

The optimal design is generated by the single sequence (1 1 1 2 2 3 3). A symmetric design generated by this sequence require 210 subjects. The following design has only 42 subjects and a completely symmetric information matrix. Its efficiency

factor for the estimation of the total effects is 0.989 for model (1) with or without period effects. The method of construction of such designs can be found in BD (p. 2296)

$$\begin{pmatrix}
1&1&1&1&1&1&2&2&2&2&2&2&3&3&3&3&3&3&4&4&4&4&4&4&5&5&5&5&5&5&6&6&6&6&6&6&7&7&7&7&7&7\\
1&1&1&1&1&1&2&2&2&2&2&2&3&3&3&3&3&3&4&4&4&4&4&4&5&5&5&5&5&5&6&6&6&6&6&6&7&7&7&7&7&7\\
2&3&4&5&6&7&3&4&5&6&7&1&4&5&6&7&1&2&5&6&7&1&2&3&6&7&1&2&3&4&7&1&2&3&4&5&1&2&3&4&5&6\\
2&3&4&5&6&7&3&4&5&6&7&1&4&5&6&7&1&2&5&6&7&1&2&3&6&7&1&2&3&4&7&1&2&3&4&5&1&2&3&4&5&6\\
3&5&7&2&4&6&4&6&1&3&5&7&5&7&2&4&6&1&6&1&3&5&7&2&7&2&4&6&1&3&1&3&5&7&2&4&2&4&6&1&3&5\\
3&5&7&2&4&6&4&6&1&3&5&7&5&7&2&4&6&1&6&1&3&5&7&2&7&2&4&6&1&3&1&3&5&7&2&4&2&4&6&1&3&5\\
3&5&7&2&4&6&4&6&1&3&5&7&5&7&2&4&6&1&6&1&3&5&7&2&7&2&4&6&1&3&1&3&5&7&2&4&2&4&6&1&3&5
\end{pmatrix}.$$

4.5 k = 8, t = 4

The optimal design is generated by the single sequence (1 1 1 2 2 2 3 3). In comparison, [24] consider optimal designs for the estimation of direct effects defined by $\frac{1}{t}\sum_{j=1}^{t}\xi_{ij}$, $i = 1,\ldots,t$. They propose optimal designs based on the sequence (1 4 2 3 3 2 4 1) with 4 subjects. A symmetric design based on this sequence has an efficiency of 40 % for the estimation of total effects.

References

1. Afsarinejad, K., Hedayat, A.S.: Repeated measurements designs for a model with self and simple mixed carryover effects. J. Stat. Plan. Inference **106**, 449–459 (2002)
2. Azaïs, J.-M., Druilhet, P.: Optimality of neighbour balanced designs when neighbour effects are neglected. J. Stat. Plan. Inference **64**, 353–367 (1997)
3. Bailey, R.A., Druilhet, P.: Optimality of neighbour-balanced designs for total effects. Ann. Stat. **32**, 1650–1661 (2004)
4. Bailey, R.A., Druilhet, P.: Optimal cross-over designs for full interaction models. Ann. Stat. **42**, 2282–2300 (2014)
5. Druilhet, P.: Optimality of neighbour balanced designs. J. Stat. Plan. Inference **81**, 141–152 (1999)
6. Druilhet, P., Tinsson, W.: Optimal repeated measurement designs for a model with partial interactions. Biometrika **96**, 677–690 (2009)
7. Druilhet, P., Tinsson, W.: Optimal cross-over designs for total effects under a model with self and mixed carryover effects. J. Stat. Plan. Inference **154**, 54–61 (2014)
8. Filipiak, K.: Universally optimal designs under an interference model with equal left- and right-neighbor effects. Stat. Probab. Lett. **82**, 592–598 (2012)
9. Filipiak, K., Markiewicz, A.: Optimality and efficiency of circular neighbor balanced designs for correlated observations. Metrika **61**, 17–27 (2005)
10. Filipiak, K., Różański, R., Sawikowska, A., Wojtera-Tyrakowska, D.: On the *E*-optimality of complete designs under an interference model. Stat. Probab. Lett. **78**, 2470–2477 (2008)
11. Gaffke, N.: Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. Ann. Stat. **15**, 942–957 (1987)
12. Kiefer, J.: Construction and optimality of generalized Youden designs. In: A Survey of Statistical Design and Linear Models. Proceedings of International Symposium, Colorado State University, Ft. Collins, 1973, pp. 333–353. North-Holland, Amsterdam (1975)

13. Kunert, J.: Optimal design and refinement of the linear model with applications to repeated measurements designs. Ann. Stat. **11**, 247–257 (1983)
14. Kunert, J.: Optimality of balanced uniform repeated measurements designs. Ann. Stat. **12**, 1006–1017 (1984)
15. Kunert, J.: Designs balanced for circular residual effects. Commun. Stat. A Theory Methods **13**, 2665–2671 (1984)
16. Kunert, J.: On the analysis of circular balanced crossover designs. J. Stat. Plan. Inference **69**, 359–370 (1998)
17. Kunert, J., Martin, R.J.: On the determination of optimal designs for an interference model. Ann. Stat. **28**, 1728–1742 (2000)
18. Kunert, J., Stufken, J.: Optimal crossover designs in a model with self and mixed carryover effects. J. Am. Stat. Assoc. **97**, 898–906 (2002)
19. Kunert, J., Stufken, J.: Optimal crossover designs for two treatments in the presence of mixed and self-carryover effects. J. Am. Stat. Assoc. **103**, 1641–1647 (2008)
20. Kushner, H.B.: Optimal repeated measurements designs: the linear optimality equations. Ann. Stat. **25**, 2328–2344 (1997)
21. Kushner, H.B.: Optimal and efficient repeated-measurements designs for uncorrelated observations. J. Am. Stat. Assoc. **93**, 1176–1187 (1998)
22. Park, D.K., Bose, M., Notz, W.I., Dean, A.M.: Efficient crossover designs in the presence of interactions between direct and carry-over treatment effects. J. Stat. Plan. Inference **141**, 846–860 (2011)
23. Pukelsheim, F.: Optimal Design of Experiments. Wiley, New York (1993)
24. Sen, M., Mukerjee, R.: Optimal repeated measurements designs under interaction. J. Stat. Plan. Inference **17**, 81–91 (1987)
25. Wilk, A., Kunert, J.: Optimal crossover designs in a model with self and mixed carryover effects with correlated errors. Metrika **78**, 161–174 (2015)
26. Zheng, W.: Universally optimal designs for two interference models. Ann. Stat. **43**, 501–518 (2015)

Survival Models with Censoring Driven by Random Enrollment

Valerii V. Fedorov and Xiaoqiang Xue

Abstract In clinical studies with time-to-event end points we face uncertainties caused by the enrollment process that can often be viewed as a stochastic process. The observed endpoints are randomly censored and the amount of gained information is random and its actual value is not known at the design stage but becomes known only after the study completion. To take this fact into account we develop a method that maximizes the average information. We derive the average elemental Fisher information matrices for a few scenarios to illustrate the approach, assuming that enrollment can be modeled by a Poisson process.

1 Introduction

The design of experiments in survival analysis attracted attention from the "optimal design of experiments" community relatively recently and examples of publications include: [13] considered optimal design for the proportional hazard models, for a two-parameter linear regression model with exponentially distributed survival times and uniformly distributed patient enrollment time, [12] continued with the proportional hazard model and introduced censoring due to dropout. Other examples include [10] and [14]. The latter publication contains a survey section that covers most of the developments in this area. All these papers are focused on specific cases and derive the respective equivalence theorems complemented either by numerical procedures or by analytic solutions for relatively simple scenarios. In this paper similarly to [1] and [5] we develop a rather universal approach that is based on the concept of the elemental information matrix. It allows us to generate optimal designs in a routine manner for a variety of hazard models accepted in clinical studies. Additionally, we try to combine the traditional optimal design with operational

V.V. Fedorov (✉)
ICONplc, 2100 Pennbrook Parkway, North Wales, PA 19454, USA
e-mail: valerii.fedorov@iconplc.com

X. Xue
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: xxue@live.unc.edu

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_12

aspects of clinical trials and, in particular, with the enrollment of subjects. In our notation and terminology we mainly follow [1] and [3].

2 Model

2.1 Assumptions and Notation

Let us start with a diagram of subject arrivals and the respective observations or censoring, see Fig. 1, compare with [3, Ch. 1] or [6, Ch. 10]. In this diagram t_j stands for the arrival time of the j-th subject, enrollment stops at T_e and follow up continued till T_s (study completion). Each subject j is observed during a time interval τ_j . We assume that all censoring happens at a $t = T_s$ and thus do not consider censoring related to dropouts either informative or not, cf. [2] and [11, Ch. 1].

Enrollment may be stopped either at the pre-fixed time $T_e = T_s - T_f$, where T_f is a fixed follow-up period, or at $T(n_s)$ when a required number n_s of subjects is enrolled, or at a moment $T(R_s)$ when the needed number R_s of events has occurred. Various combinations of these stopping rules were discussed in the literature on life-testing experiments, see [2] for a summary of results and further references. In this paper the focus is on the stopping by time case.

We assume that:

- Time-to-event is random and its probability density is $\varphi(\tau, \eta)$.
- Elemental parameters η depend on variables (controls) $x \in \mathscr{X}$, i.e. $\eta = \eta(x, \theta)$.
- Response functions $\eta(x, \theta)$ are given, finite for all $x \in \mathscr{X}$ and twice differentiable with respect to $\theta \in \Omega$.
- Design region $\mathscr{X}$ and Ω are compact.

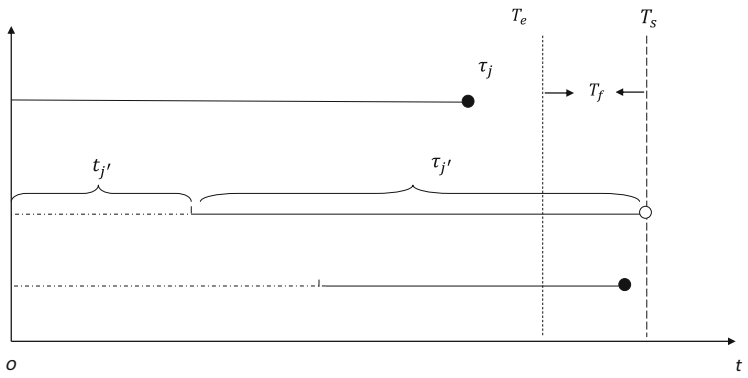

Fig. 1 Staggered entry, t_j stands for arrival time, τ_j exposure time; enrollment starts at $t = 0$ and stops at $t = T_e$; ∘ indicates censoring; • indicates an event; a trial stops at $t = T_s$; T_f is a minimal follow-up period

- Arrival times are generated by a homogeneous Poisson process with rate λ if it is not stated otherwise. "Total" enrollment $n(t)$ at moment t is Poisson distributed with parameter $\Lambda(t) = \lambda t$.
- On arrival subjects are randomized across doses $\{x_i\}_1^N$ with probabilities $\{p_i\}_1^N$, $\sum_{i=1}^N p_i = 1$.

To identify a subject and related outcomes we will use subscripts ij. At the study completion, the following values become known:

- Number of subjects n_i assigned to dose x_i and all arrival times t_{ij}.
- The outcomes $\{y_{ij}\}_1^{n_i} = \{\tau_{ij}, \delta_{ij}\}_1^{n_i}$, where $\delta_{ij} = 1$ if $\tau_{ij} \le T_s - t_{ij}$ and $\delta_{ij} = 0$ otherwise.

Most of the distributions that are popular in survival analysis, cf. [3], depend on one or two parameters. The objective of design is to find doses x_i^* and respective p_i^* that minimize a pre-selected function of the variance-covariance matrix of the maximum likelihood estimator $\hat{\theta}$ of unknown parameters. The more detailed and more accurate definition will be given later. To address the problem we follow the path introduced in [1] and [5, Ch. 2] with some additions that take into account censoring and randomness of the arrival times. To do this we are to define and find the analogue of the elemental information matrix that is essential for the approach.

In what follows we use the following notation:

- Distribution function: $\Phi(\tau, \eta) = \int_0^\tau \varphi(t, \eta)dt$.
- Survival function: $S(\tau, \eta) = 1 - \Phi(\tau, \eta)$.
- Hazard function: $h(\tau, \eta) = \varphi(\tau, \eta)/S(\tau, \eta) = -\partial \ln S(\tau, \eta)/\partial\tau$.
- Integrated hazard function: $H(\tau, \eta) = \int_0^\tau h(t, \eta)dt$.

In survival analysis, the selection of a hazard function is often a starting point of model building. Note that $S(\tau, \eta) = e^{-H(\tau,\eta)}$ and $\varphi(\tau, \eta) = h(\tau, \eta)e^{-H(\tau,\eta)}$.

2.2 *Elemental Information Matrix in Survival Analysis Setting*

We confine ourselves to maximum likelihood estimators and do not consider partial maximum likelihood estimators. There are two reasons: in general the use of partial maximum likelihood estimators leads to loss of information and the respective information matrices contain a random number of terms even in the case of a fixed sample size, cf. [3, Ch. 8.4 and 8.5]. The latter makes the design problem more complicated and is beyond of the scope of our paper.

To derive the respective Fisher information matrix (FIM) of a single observation for parameters θ, let us start with the FIM for the original/elemental parameters η, which in [1] was called the elemental Fisher information matrix (elemental FIM). For observations with the censoring time τ the elemental FIM can be calculated

using either one of two formulae:

$$\nu(\tau,\eta)=\int_0^\tau \frac{\partial \ln\varphi(t,\eta)}{\partial\eta}\frac{\partial \ln\varphi(t,\eta)}{\partial\eta\top}\varphi(t,\eta)dt+S(\tau,\eta)\frac{\partial \ln\Phi(\tau,\eta)}{\partial\eta}\frac{\partial \ln\Phi(\tau,\eta)}{\partial\eta\top} \tag{1}$$

or

$$\nu(\tau,\eta)=\int_0^\tau \frac{\partial \ln h(t,\eta)}{\partial\eta}\frac{\partial \ln h(t,\eta)}{\partial\eta^\top}\varphi(t,\eta)dt, \tag{2}$$

see, for instance, [7] and [8].

The role of elemental FIM can be appreciated if, as in [5, Ch. 1.6], we verify that the total FIM can be presented as

$$\underline{\mathrm{M}}(\{t_{ij}\},T_s,\theta)=\sum_{i=1}^{N}\sum_{j=1}^{n_i}F_i\nu(T_s-t_{ij},\eta(x_i,\theta))F_i^\top, \tag{3}$$

where $\tau_{ij}=T_s-t_{ij}$ and $F_i=F(x_i,\theta)=\partial\eta^\top(x_i,\theta)/\partial\theta$. Note that $F(x,\theta)$ is an $(m\times k)$ matrix, where $m=\dim\theta$, and $k=\dim\eta$. Thus, the knowledge of an elemental information matrix for a given density $\varphi(t,\eta)$ or hazard function $h(t,\eta)$ allows us to build the information matrix for $\hat{\theta}$ as soon as the function $\eta(x,\theta)$ is selected. Matrix $\mu(x,\tau,\theta)=F(x,\theta)\nu(\tau,\eta(x,\theta))F^\top(x,\theta)$ can be viewed as the FIM of a single observation performed at x with exposure time $\tau=T_s-t$. Values $\{n_i\}$ and $\{\tau_{ij}\}$ are realizations of random variables and become known at the completion of a study. Only their distributions are known at the design stage.

3 Random Subject Accrual

In a typical clinical trial arrival times $\{t_{ij}\}$ are known only after enrollment completion. Unfortunately one has to select doses and randomization rates prior to enrollment. In what follows we assume that arrival times can be modeled with a Poisson process with (rate) intensity $\lambda(t)$. The latter is assumed to be known at any time interval used in all formulae that follow. Let us assume for simplicity that the enrollment stops at predefined T_e.

Unlike those of the traditional design theory, the design $\xi=\{p_i,x_i\}_1^N$ is intrinsically connected to the operational aspects of clinical studies and, in particular, to the randomization that should carefully follow a study protocol. It is worth mentioning the two most popular :

- Complete randomization: at each arrival the subject is assigned with probability p_i to treatment x_i, $i=1,2,\ldots,N$.

- Permuted block design: randomization in blocks. For instance, for $p_1 = p_2 = 1/2$ the j-th subject is assigned either to dose x_1 or dose x_2 with probability $1/2$ and the $(j+1)$-th subject is assigned to the respective complementary dose.

While for larger sample sizes the randomization type is not important for the smaller sample sizes it may lead to a noticeable impact.

Let us confine ourselves to the first randomization type in this case. As arrival times are generated by a homogeneous Poisson process with rate λ, the randomized assignment of subjects to different doses splits this process in N Poisson processes and arrival times at each dose x_i follow the Poisson process with rate $\lambda_i = p_i\lambda$. This fact is an immediate corollary of the "Colouring Theorem", also known as the "Thinning Theorem", see [9, Ch. 5].

Arrival times and the number of enrolled subjects at each dose are not known at the design stage and we have to find a reasonable approximation of the total information matrix (3). Let us introduce

$$\Sigma_i = \sum_{j=1}^{n_i} v\left(T_s - t_{ij}, \eta(x_i, \theta)\right) = \sum_{j=1}^{n_i} v\left(T_e + T_f - t_{ij}, \eta(x_i, \theta)\right). \tag{4}$$

One may recall that T_f stands for a follow up time. Matrices Σ_i and $v\big(T_s - t_j, \eta(x_i)\big)$ are random non-negative definite. Using (2) one can verify that for any $0 \le t_j < T_e$

$$v\big(T_s - t_j, \eta(x_i, \theta)\big) \le v\big(T_s, \eta(x_i, \theta)\big), \tag{5}$$

where the ordering should be understood in the Loewner sense.

From Campbell's Theorem, see [9, Ch. 3.2], it follows that (for a while x_i, θ will be skipped):

$$\mathrm{E}[\Sigma_i] = \lambda_i \int_0^{T_e} v(T_e + T_f - t, \eta)dt = \lambda_i T_e \int_0^{T_e} v(T_e + T_f - t, \eta)\frac{1}{T_e}dt$$

and similarly

$$\mathrm{Var}\left[\Sigma_{i,\alpha\beta}\right] = \lambda_i \int_0^{T_e} v^2_{i,\alpha\beta}(T_e + T_f - t, \eta)dt = \lambda_i T_e \int_0^{T_e} v^2_{i,\alpha\beta}(T_e + T_f - t, \eta)\frac{1}{T_e}dt,$$

where $v_{\alpha\beta}$ is the element of matrix v that corresponds to η_α and η_β respectively. One may recall that $\lambda_i T_e$ is the expected number of patients that assigned to dose x_i.

Note that the relative standard deviation of elements of matrix Σ_i is of order $1/\sqrt{\lambda_i T_e}$ and for the majority of clinical trials $\lambda_i T_e$ is several hundreds. Therefore, the expected value of Σ is a "reasonable" replacement of Σ itself. The integral

$$\overline{v}(T_e, T_f, \eta) = \int_0^{T_e} v(T_e + T_f - t, \eta)\frac{1}{T_e}dt. \tag{6}$$

can be viewed as an expectation of the elemental information matrix when t is uniformly distributed on $[0, T_e]$, i.e.

$$\overline{v}(T_e, T_f, \eta) = \mathrm{E}\left[v(T_e + T_f - t, \eta)\right]. \tag{7}$$

A similar remark is valid for the integral (6). Combining (3), (4) and (6) we get

$$\mathfrak{M}(\xi, T_e, T_f, \theta) = \mathrm{E}\left[\underline{\mathrm{M}}(\{t_{ij}\}, T_e, T_f, \theta)\right] = \lambda T_e \sum_{i=1}^{N} p_i \overline{\mu}\left(T_e, T_f, \eta(x_i, \theta)\right), \tag{8}$$

where

$$\overline{\mu}\left(T_e, T_f, \eta(x, \theta)\right) = F(x, \theta)\overline{v}\left(T_e, T_f, \eta(x, \theta)\right) F^{\top}(x, \theta) \text{ and } \xi = \{p_i, x_i\}_1^N.$$

Introducing (average) normalized FIM

$$\lambda^{-1} T_e^{-1} \mathfrak{M}(\xi, T_e, T_f, \theta) = \mathrm{M}(\xi, T_e, T_f, \theta) = \sum_{i=1}^{N} p_i \overline{\mu}\left(T_e, T_f, \eta(x_i, \theta)\right), \tag{9}$$

we come to the following optimization (design) problem for given T_e, T_f, θ:

$$\xi^*(\lambda, T_e, T_f, \theta) = \arg\min_{\xi} \Psi\left[\lambda T_e \mathrm{M}(\xi, T_e, T_f, \theta)\right]. \tag{10}$$

For homogeneous criteria Ψ , cf. [5, Ch. 2.3], optimization problem (10) is equivalent to

$$\xi^*(T_e, T_f, \theta) = \arg\min_{\xi} \Psi\left[\mathrm{M}(\xi, T_e, T_f, \theta)\right] \tag{11}$$

and in this case the optimal design ξ^* does not depend on the enrollment rate λ. As soon as either matrix $\overline{v}$ or $\overline{\mu}$ is defined the design problem can be addressed in the framework of the classical (standard) design theory as in [1], [5, Ch.5.4] by replacing v or μ by $\overline{v}$ and $\overline{\mu}$ respectively. However there is one significant difference of (11) from the standard case: now locally optimal design $\xi^*(T_e, T_f, \theta)$ depends on T_e, T_f and θ. One can extend the optimization problem and look for the optimal enrollment duration and the optimal follow up period:

$$(T_e^*, T_f^*) = \arg\min_{T_e, T_f} \Psi\left[\mathrm{M}\left(\xi^*(T_e, T_f, \theta), T_e, T_f, \theta\right)\right], \quad C(\lambda, T_e, T_f) \le C^*, \tag{12}$$

where $C(\lambda, T_e, T_f)$ is a cost function that is defined by operational characteristics of a clinical study.

4 Example: Proportional Hazard Family

Let us consider the proportional hazard model, cf. [3, Ch.5.3], for this model the hazard function, the integrated hazard, the survivor function and density are:

$$h(\tau,\eta)=\eta h_0(\tau),\ \ H(\tau)=\eta H_0(\tau)=\eta\int_0^\tau h_0(t)dt, \tag{13}$$
$$S(\tau,\eta)=e^{-\eta H_0(\tau)},\ \ \varphi(\tau,\eta)=\eta h_0(\tau)e^{-\eta H_0(\tau)}.$$

From (2) and (13) it immediately follows that for the proportional hazard family the elemental FIM (actually it is a scalar in the case considered) for observations censored at τ is

$$v(\tau,\eta)=\frac{1}{\eta^2}[1-e^{-\eta H_0(\tau)}]=\frac{1}{\eta^2}\Phi(\tau,\eta), \tag{14}$$

and for the enrollment that can be described by a homogeneous Poisson process the average elemental FIM is

$$\bar{v}(T_e,T_f,\eta)=\frac{1}{\eta^2}\int_0^{T_e}\frac{1}{T_e}\Phi(T_e+T_f-t,\eta)dt=\frac{1}{\eta^2}\bar{\Phi}(T_e+T_f,\eta). \tag{15}$$

Let

$$\eta(x,\theta)=e^{\theta^\top f(x)}\ \text{and}\ F(x,\theta)=\eta(x,\theta)f^\top(x). \tag{16}$$

Combining (7), (8) and (14), (15) one can verify that the average normalized FIM (9) for this model is

$$\mathrm{M}(\xi)=\mathrm{M}(\xi,T_e,T_f,\theta)=\sum_{i=1}^{N}p_i\omega(x_i)f(x_i)f^\top(x_i), \tag{17}$$

where $\omega(x)=\bar{\Phi}(T_e+T_f,\eta(x,\theta))$. Thus, given T_e,T_f,θ matrix (17) has exactly the same structure as in the traditional case and the whole "optimal design" machinery may be implemented to build optimal designs. For example, for the D-criterion (i.e. minimization of $\ln|\mathrm{M}^{-1}(\xi)|$) the familiar inequality

$$\omega(x)d(x,\xi^*)=\omega(x)f^\top(x)\mathrm{M}^{-1}(\xi^*)f(x)\le m=\dim\theta,\ \ \text{for all}\ \ x\in\mathscr{X}$$

provides the necessary and sufficient condition that $\xi^*=\arg\min_\xi|\mathrm{M}^{-1}(\xi)|$. If one can verify that ξ^* has exactly m support points ($N^*=m$) then $p_i^*\equiv 1/m$, see [4, Ch 2], to the great relief of those who run clinical studies: equal randomization rates are operationally preferable.

While other analytic niceties from optimal design theory stay valid for our problem, the most important fact is that we can apply numerous and well developed numerical algorithms. However one should not forget that optimization problem (12) needs the multiple computations of (11).

5 Conclusion

We extended the results of [1] to the case of randomly censored observations that are typical for clinical trials with time-to-event endpoints. As in that paper the collection of results presented here will allow streamlining of the practical aspects of clinical study design and it will lead to the development of software that can incorporate multiple specific cases in one menu driven toolkit. We considered only one enrollment stopping rule: stopping by time. Stopping rules based on the number of enrolled subjects or on the number of events are beyond the scope of this paper and we hope to present the respective results in subsequent publications.

Acknowledgements We thank the referees for their many helpful comments and insightful suggestions leading to an improved paper.

References

1. Atkinson, A.C., Fedorov, V.V., Herzberg, A.M., Zhang, R.M.: Optimal experimental design for generalized regression models. JSPI **144**, 81–91 (2012)
2. Balakrishnan, N., Kundu, D.: Hybrid censoring: Models, inferential results and applications. Comput. Stat. Data Anal. **53**, 166–209 (2013)
3. Cox, D.R., Oakes, D.: Analysis of Survival Data. Chapman & Hall, London (1984)
4. Fedorov, V.V.: Theory of Optimal Experiment. Academic Press, New York (1972)
5. Fedorov, V.V., Leonov, S.: Optimal Design for Nonlinear Response Models. CRC Press, New York (2013)
6. Fridman, L.M., Furberg, C.D., DeMets, D.L.: Fundamentals of Clinical Trials, 4th edn. Springer, New York (2010)
7. Gertsbakh, I., Kagan, A.: Characterization of the Weibull distribution by properties of the Fisher information under type-I censoring. Stat. Probab. Lett. **42**, 99–105 (1999)
8. Gupta, R.D., Kundu, D.: On the comparison of Fisher information of the Weibull and GE distributions. JSPI **9**, 3130–3144 (2006)
9. Kingman, J.F.C.: Poisson Processes. Oxford University Press, New York (1993)
10. Konstantinou, M., Biedermann, S., Kimber, A.: Optimal designs for two-parameter nonlinear models for application to survival models. Statistica Sinica **24**, 415–428 (2014)
11. Little, R.J.A., Rubin, D.B.: Statistical Analysis with Missing Data. Wiley, New York (1987)
12. Lopez-Fidalgo, J., Rivas-Lopez, M.J.: Optimal experimental designs for partial likelihood information. Comput. Stat. Data Anal. **71**, 859–867 (2012)
13. Lopez-Fidalgo, J., Rivas-Lopez, M.J., Campo, R.D.: Optimal designs for Cox regression. Statistica Neerlandica **63**, 135–148 (2009)
14. Müller, C.H.: D-optimal designs for lifetime experiments with exponential distribution and consoring. Adv. Model-Oriented Des. Anal. 161–168 (2013)

Optimal Design for Prediction in Random Field Models via Covariance Kernel Expansions

Bertrand Gauthier and Luc Pronzato

Abstract We consider experimental design for the prediction of a realization of a second-order random field Z with known covariance function, or kernel, K. When the mean of Z is known, the integrated mean squared error of the best linear predictor, approximated by spectral truncation, coincides with that obtained with a Bayesian linear model. The machinery of approximate design theory is then available to determine optimal design measures, from which exact designs (collections of sites where to observe Z) can be extracted. The situation is more complex in the presence of an unknown linear parametric trend, and we show how a Bayesian linear model especially adapted to the trend can be obtained via a suitable projection of Z which yields a reduction of K.

1 Introduction

We consider a centered second-order random field $(Z_x)_{x\in\mathscr{X}}$ indexed over $\mathscr{X}$, a compact subset $\mathscr{X}$ of $\mathbb{R}^d$, $d \geq 1$, with covariance function $\mathsf{E}\{Z_x Z_{x'}\} = K(x, x') = K(x', x)$. We also consider a σ-finite measure μ on $\mathscr{X}$ which will be used to define the Integrated Mean Squared Error (IMSE) for prediction, see (1). We assume that $\mu(\mathscr{X}) = 1$ without any loss of generality. We suppose that $K : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is continuous and thus belongs to $L^2_{\mu\times\mu}(\mathscr{X} \times \mathscr{X})$. This framework is now classical for computer experiments, where $(Z_x)_{x\in\mathscr{X}}$ represents the output of a computer code with input variables $x \in \mathscr{X}$, in the sense that it forms a prior on the possible behavior of the code when x vary in $\mathscr{X}$, see in particular [10]. The design problem then consists in choosing n sites $x_1, \ldots, x_n$ where to observe Z_x *without error* (i.e., n inputs values for the computer code) in order to predict at best the realization of Z_x

B. Gauthier (✉)
ESAT-STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium
e-mail: bgauthie@esat.kuleuven.be

L. Pronzato
CNRS, Laboratoire I3S – UMR 7271 Université de Nice-Sophia Antipolis/CNRS, Nice, France
e-mail: pronzato@cnrs.fr

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_13

over $\mathcal{X}$. Note that the objective concerns prediction of *one instance* of the random field, based on a finite set of error-free observations.

The best linear predictor of Z_x, unbiased and in terms of mean squared error (the simple kriging predictor), based on the vector of observations $\mathbf{z}_n = (Z_{x_1}, \ldots, Z_{x_n})^\top$ obtained with the design $\mathcal{D}_n = (x_1, \ldots, x_n)$, is given by $\mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{z}_n$, where $\mathbf{k}(x) = (K(x_1, x), \ldots, K(x_n, x))^\top$ and the matrix $\mathbf{K}$ has elements $\{\mathbf{K}\}_{ij} = K(x_i, x_j)$. Its Mean Squared Error (MSE) is $\mathsf{MSE}(x; \mathcal{D}_n) = K(x, x) - \mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{k}(x)$.

The IMSE will be computed for the measure μ, which weighs the importance given to the precision of predictions in different parts of $\mathcal{X}$,

$$\mathsf{IMSE}(\mathcal{D}_n) = \int_{\mathcal{X}} \mathsf{MSE}(x; \mathcal{D}_n)\, \mathrm{d}\mu(x)\,. \tag{1}$$

Choosing a design $\mathcal{D}_n$ with minimum IMSE is then a rather natural objective; however, IMSE-optimal designs are considered as difficult to compute, see [10]. The introduction of the following integral operator on $L^2_\mu(\mathcal{X})$ yields a spectral representation of the IMSE and is at the core of the design approach presented in this paper.

For all $f \in L^2_\mu(\mathcal{X})$ and $x \in \mathcal{X}$, we define

$$T_\mu[f](x) = \int_{\mathcal{X}} f(t)K(x, t)\, \mathrm{d}\mu(t)\,.$$

From Mercer's theorem, there exists an orthonormal set $\{\widetilde{\phi}_k, k \in \mathbb{N}\}$ of eigenfunctions of T_μ in $L^2_\mu(\mathcal{X})$ associated with nonnegative eigenvalues, the eigenfunctions corresponding to non-zero eigenvalues being continuous on $\mathcal{S}_\mu$, the support of μ. Moreover, for $x, x' \in \mathcal{S}_\mu$, K has the representation $K(x, x') = \sum_{k=1}^{\infty} \lambda_k \widetilde{\phi}_k(x)\, \widetilde{\phi}_k(x')$, where the λ_k denote positive eigenvalues that we assume to be sorted by decreasing values, and the convergence is uniform on compact subsets of $\mathcal{S}_\mu$. As shown below, see (2), the integration over $\mathcal{X}$ in (1) can then be replaced by a summation over k, the index of eigenvalues. There is a slight technicality here, due to the fact that $\mathcal{S}_\mu$ may be strictly included in $\mathcal{X}$, so that the eigenfunctions $\widetilde{\phi}_k$, which are only defined μ-almost everywhere, may not be defined at a general design point $x_i \in \mathcal{X}$. This is the case for instance when μ is a discrete measure supported at a finite set of points $\mathcal{X}_Q \subset \mathcal{X}$ (thus in particular when the IMSE is approximated via a quadrature rule) and $\mathcal{D}_n$ is not a subset of $\mathcal{X}_Q$. A general way to avoid this difficulty is via the introduction of the canonical extensions

$$\phi_k = (1/\lambda_k) T_\mu[\widetilde{\phi}_k]\,, \quad \lambda_k > 0\,,$$

which are defined on the whole set $\mathcal{X}$ and satisfy $\phi_k(x) = \widetilde{\phi}_k(x)$ on $\mathcal{S}_\mu$. Note that the ϕ_k are continuous since we assumed that $K(\cdot, \cdot)$ is continuous. One may refer to

[4] for more details. Then, $K(x,x') = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(x')$ for all x, x' in $\mathscr{X}$ and

$$\mathsf{IMSE}(\mathscr{D}_n) = \tau - \sum_{k=1}^{\infty} \lambda_k^2 \underline{\phi}_k^\top \mathbf{K}^{-1} \underline{\phi}_k \,, \tag{2}$$

where $\tau = \sum_{k=1}^{\infty} \lambda_k$ and $\underline{\phi}_k = (\phi_k(x_1), \ldots, \phi_k(x_n))^\top$. A spectral truncation that uses only the n_{trc} largest eigenvalues yields an approximation of $\mathsf{IMSE}(\mathscr{D}_n)$,

$$\mathsf{IMSE}_{trc}(\mathscr{D}_n) = \tau_{trc} - \sum_{k=1}^{n_{trc}} \lambda_k^2 \underline{\phi}_k^\top \mathbf{K}^{-1} \underline{\phi}_k = \mathrm{trace}(\Lambda_{trc} - \Lambda_{trc} \Phi_{trc}^\top \mathbf{K}^{-1} \Phi_{trc} \Lambda_{trc}) \,, \tag{3}$$

with $\tau_{trc} = \sum_{k=1}^{n_{trc}} \lambda_k$, Φ_{trc} the $n \times n_{trc}$ matrix with elements $\{\Phi_{trc}\}_{ik} = \phi_k(x_i)$ and $\Lambda_{trc} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_{n_{trc}}\}$. This satisfies $\mathsf{IMSE}_{trc}(\mathscr{D}_n) \leq \mathsf{IMSE}(\mathscr{D}_n) \leq \mathsf{IMSE}_{trc}(\mathscr{D}_n) + \tau_{err}$, where $\tau_{err} = \tau - \tau_{trc} = \sum_{k>n_{trc}} \lambda_k$, see [4, 6].

Following [2, 13], we recall below (Sect. 2) how, using its Karhunen-Loève decomposition, $(Z_x)_{x\in\mathscr{X}}$ can be interpreted as a Bayesian Linear Model (BLM) *with correlated errors* having the eigenfunctions ϕ_k, $k = 1, \ldots, n_{trc}$, as regressors, and for which the IMSE of prediction coincides with $\mathsf{IMSE}_{trc}(\mathscr{D}_n)$. A similar decomposition has been used in [3], but for the estimation of trend parameters (or equivalently, for the prediction of the mean of Z_x). The IMSE for a BLM *with uncorrelated errors* takes the form of a Bayesian A-optimality criterion, which opens the way for the construction of approximate optimal designs (design measures) by convex optimization, see [8]. Exact designs $\mathscr{D}_n$, that is, collections of sites where to observe Z, can be extracted from these optimal measures, see Sect. 3. However, replacing correlated errors by uncorrelated ones, in order to be able to use the approximate design machinery, introduces an additional approximation (besides spectral truncation). Whereas the approximation in [13] relies on an homoscedastic model, the one we propose in Sect. 2 uses a more accurate heteroscedastic model. The situation is more complex when a linear parametric trend is present, and in Sect. 4 we show how a suitable projection of $(Z_x)_{x\in\mathscr{X}}$ on the linear subspace spanned by the trend, equivalent to a kernel reduction, yields a BLM with smaller errors than the direct approach of [13] that uses the original kernel K.

2 Bayesian Linear Models

Exact BLM The Karhunen-Loève decomposition of Z_x yields

$$Z_x = \sum_{k=1}^{n_{trc}} \beta_k \phi_k(x) + \varepsilon_x \,, \tag{4}$$

where ε_x, $x \in \mathscr{X}$, is a centered random field with covariance given by

$$\mathsf{E}\{\varepsilon_x \varepsilon_{x'}\} = K_{err}(x,x') = K(x,x') - K_{trc}(x,x'),$$

with $K_{trc}(x,x') = \sum_{k=1}^{n_{trc}} \lambda_k \phi_k(x)\phi_k(x')$ and where the β_k are mutually uncorrelated centered r.v., orthogonal to ε_x for all $x \in \mathscr{X}$, with $\mathsf{var}(\beta_k) = \lambda_k$. No approximation is involved at this stage, and this BLM gives an exact representation of $(Z_x)_{x\in\mathscr{X}}$. Consider the predictor $\widehat{Z}_x = \sum_{k=1}^{n_{trc}} \hat{\beta}_k \phi_k(x) = \underline{\phi}_{trc}^\top(x)\widehat{\underline{\beta}}$, with $\underline{\phi}_{trc} = (\phi_1(x), \ldots, \phi_{n_{trc}}(x))^\top$ and $\widehat{\underline{\beta}} = (\Phi_{trc}^\top \mathbf{K}_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1} \Phi_{trc}^\top \mathbf{K}_{err}^{-1} \mathbf{z}_n$ the estimator that minimizes the regularized LS criterion $J(\underline{\beta}) = (\mathbf{z}_n - \Phi_{trc}\underline{\beta})^\top \mathbf{K}_{err}^{-1}(\mathbf{z}_n - \Phi_{trc}\underline{\beta}) + \underline{\beta}^\top \Lambda_{trc}^{-1} \underline{\beta}$ ($\widehat{\underline{\beta}}$ coincides with the posterior mean of $\underline{\beta}$ when Z_x is Gaussian). Direct calculation shows that the IMSE of $\widehat{Z}_x$ equals $\mathsf{IMSE}_{trc}(\mathscr{D}_n)$ given by (3) and can also be written as

$$\mathsf{IMSE}_{trc}(\mathscr{D}_n) = \mathrm{trace}[(\Phi_{trc}^\top \mathbf{K}_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1}]\,.$$

Approximate BLM with uncorrelated errors In [13], the correlated errors ε_x of (4) are replaced by uncorrelated homoscedastic errors having variance $\sigma^2 = \tau_{err} = \tau - \tau_{trc}$. A more accurate approximation of the exact model (4) is obtained when using uncorrelated but heteroscedastic errors, with the same variance $\sigma^2(x) = K_{err}(x,x)$ as ε_x. We shall call this model the *heteroscedastic BLM* and denote by Σ_{err} the diagonal matrix $\mathrm{diag}\{K_{err}(x_1,x_1), \ldots, K_{err}(x_n,x_n)\}$. The curve (solid line) on the top of Fig. 3-Right shows $\sigma^2(x)$, $x \in [0,1]$, for the Matérn 3/2 kernel $K(x,x') = C_{3/2,10}(|x-x'|)$, where

$$C_{3/2,\vartheta}(t) = (\vartheta t + 1)\,\exp(-\vartheta t)\,, \tag{5}$$

with $n_{trc} = 10$. The first 4 eigenfunctions ϕ_k for this kernel are plotted in Fig. 2-Left. The strongly oscillating behavior of $\sigma^2(x)$, due to the form of eigenfunctions in K_{trc}, motivates the use of the heteroscedastic BLM instead of an approximate model with homoscedastic errors. This seems important within the framework of computer experiments, but would be less critical, however, in presence of additive uncorrelated measurement errors, as considered in [13]; in that case, the errors ε_x due to spectral truncation are even neglected in [3]. The IMSE for prediction with the heteroscedastic BLM is

$$\mathsf{IMSE}_{trc}^{hBLM}(\mathscr{D}_n) = \mathrm{trace}[(\Phi_{trc}^\top \Sigma_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1})^{-1}]\,. \tag{6}$$

3 Optimal Design

The IMSE (6), considered as a function of the exact design $\mathscr{D}_n$, has the form of an A-optimality criterion for the Bayesian information matrix $\Phi_{trc}^\top \Sigma_{err}^{-1} \Phi_{trc} + \Lambda_{trc}^{-1}$, see [8]. For ξ a probability measure on $\mathscr{X}$ (i.e., a design measure), let

$$\mathbf{M}(\xi) = \int_{\mathscr{X}} \sigma^{-2}(x) \underline{\phi}_{trc}(x) \underline{\phi}_{trc}^\top(x) \, \mathrm{d}\xi(x) \,,$$

which satisfies $\mathsf{IMSE}_{trc}^{hBLM}(\mathscr{D}_n) = \mathrm{trace}[(n\mathbf{M}(\xi_n) + \Lambda_{trc}^{-1})^{-1}]$, with ξ_n the empirical measure $(1/n)\sum_{i=1}^n \delta_{x_i}$ associated with $\mathscr{D}_n$. For any $\alpha \geq 0$, the Bayesian A-optimality criterion

$$\psi_\alpha(\xi) = \mathrm{trace}[(\alpha \mathbf{M}(\xi_n) + \Lambda_{trc}^{-1})^{-1}]$$

is a convex function of ξ which can easily be minimized using an algorithm for optimal design, see, e.g., [9, Chap. 9]. The solution is much facilitated when (i) μ is a discrete measure with finite support $\mathscr{X}_Q$ (a quadrature approximation is used to compute the IMSE) and (ii) only design measures ξ dominated by μ are considered (design points are chosen among quadrature points in $\mathscr{X}_Q$). Indeed, only the values $\widetilde{\phi}_k(x_j)$ for $x_j \in \mathscr{X}_Q$ are then used in all calculations, without having to compute the canonical extensions $\phi_k = (1/\lambda_k) T_\mu[\widetilde{\phi}_k]$, see Sect. 1, and the $\widetilde{\phi}_k(x_j)$ are obtained by the spectral decomposition of a $Q \times Q$ matrix, with $Q = |\mathscr{X}_Q|$, see [4, 5]. Once an optimal design measure ξ_α^* minimizing $\psi_\alpha(\cdot)$ has been determined, we still have to extract from it an exact design $\mathscr{D}_n$ with n points (and no repetitions since there are no observation errors). We have tried several procedures; all are based on the selection of n points among the support of ξ_α^*, $\mathscr{S}_{\xi_\alpha^*} = (x^{(1)}, \ldots, x^{(m)})$ say, and involve some trial and error for tuning values of α and n_{trc} in order to facilitate extraction of an n-point design when n is given *a priori*. They are listed below:

1. Keep the n points $x^{(i)}$ of $\mathscr{S}_{\xi_\alpha^*}$ with largest weight $\xi_\alpha^*(x^{(i)})$.
2. Perform a greedy (one-point-at-a-time) minimization of the IMSE using the finite set $\mathscr{S}_{\xi_\alpha^*}$ as design space.
3. Perform a minimum-spanning-tree clustering of $\mathscr{S}_{\xi_\alpha^*}$, with the metric $\Delta(x, x') = K(x, x) + K(x', x') - 2K(x, x')$ induced by K, with n clusters (the x_i then correspond to the barycenters of $x^{(j)}$ in clusters, with weights the $\xi_\alpha^*(x^{(j)})$).

Choosing α and n_{trc} of the same order of magnitude as n seems reasonable. Note that while 1. and 2. maintain the x_i within $\mathscr{X}_Q$ when μ has finite support $\mathscr{X}_Q$ and ξ_α^* is dominated by μ, this is not the case for 3. and canonical extensions $\phi_k = (1/\lambda_k) T_\mu[\tilde{\phi}_k]$ must then be computed (which requires summations over $\mathscr{X}_Q$). After extraction of a n-point design $\mathscr{D}_n$, a local minimization of $\mathsf{IMSE}(\mathscr{D}_n)$ (a problem in $\mathbb{R}^{n \times d}$) can be performed using any standard algorithm for unconstrained optimization (e.g., conjugate gradient, variable metric), since optimal points for

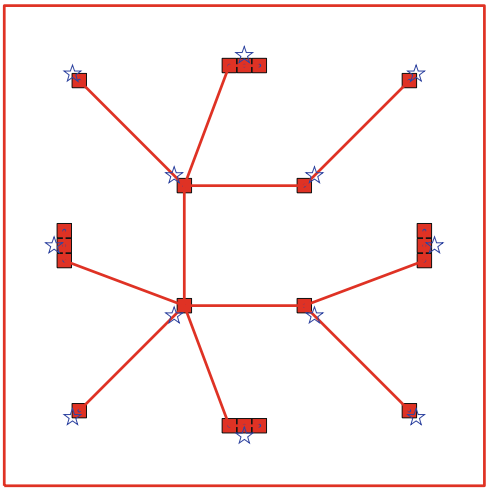

Fig. 1 The 20 squares correspond to the support of ξ_α^* when the vertex-exchange algorithm is stopped (the directional derivative of $\psi_\alpha(\xi)$ exceeds -10^{-4}), the associated minimum spanning-tree identifies $n = 12$ clusters, the stars indicate the optimal 12-point design obtained by local minimization of the IMSE

observation of Z_x need to be as correlated as possible with points within $\mathscr{X}$, and therefore lie in the interior of $\mathscr{X}$. Here also canonical extensions must be computed.

An illustration of procedure 3. is presented in Fig. 1 for $\mathscr{X} = [0,1]^2$ and $K(x,x') = C_{3/2,10}(\|x - x'\|)$, see (5), with $n_{trc} = \alpha = 10$; μ is the uniform measure on a regular grid $\mathscr{X}_Q$ of $Q = 33 \times 33$ points and a vertex-exchange algorithm with Armijo-type line search [1] is used to determine ξ_α^* supported on $\mathscr{X}_Q$.

4 Kernel Reduction for Models with Parametric Trend

Consider now the random field $Y_x = \mathbf{g}^\top(x)\underline{\theta} + Z_x$, where Z_x is as in Sect. 1, $\mathbf{g}(x) = (g_1(x), \cdots, g_p(x))^\top$ is a vector of (known) real-valued trend functions defined on $\mathscr{X}$ and where $\underline{\theta} = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$ is an unknown vector of parameters. The Best Linear Unbiased Predictor (BLUP) of Y_x (the universal kriging predictor), based on observations $\mathbf{y}_n = (Y_{x_1}, \ldots, Y_{x_n})^\top$, is $\mathbf{g}^\top(x)\widehat{\underline{\theta}} + \mathbf{k}^\top(x)\mathbf{K}^{-1}(\mathbf{y}_n - \mathbf{G}\widehat{\underline{\theta}})$, where $\widehat{\underline{\theta}} = (\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{G})^{-1}\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{y}_n$ with $\mathbf{G}$ the $n \times p$ matrix with elements $\{\mathbf{G}\}_{ij} = g_j(x_i)$ (we assume that $\mathbf{G}$ has full column rank). Its MSE is

$$\begin{aligned}\mathsf{MSE}(x;\mathscr{D}_n) = {} & K(x,x) - \mathbf{k}^\top(x)\mathbf{K}^{-1}\mathbf{k}(x) \\ & + [\mathbf{g}(x) - \mathbf{G}^\top\mathbf{K}^{-1}\mathbf{k}(x)]^\top(\mathbf{G}^\top\mathbf{K}^{-1}\mathbf{G})^{-1}[\mathbf{g}(x) - \mathbf{G}^\top\mathbf{K}^{-1}\mathbf{k}(x)]\,,\end{aligned}$$

see, e.g., [11, Chaps. 3,4], and $\mathsf{IMSE}(\mathscr{D}_n) = \int_{\mathscr{X}} \mathsf{MSE}(x;\mathscr{D}_n)\,\mathrm{d}\mu(x)$.

Similarly to Sect. 2, we can consider an exact BLM $Y_x = \mathbf{f}^\top(x)\underline{\gamma} + \varepsilon_x$, where $\mathbf{f}^\top(x) = (\mathbf{g}^\top(x), \underline{\phi}_{trc}^\top(x))$ and $\underline{\gamma} = (\underline{\theta}^\top, \underline{\beta}^\top)^\top$. The predictor of Y_x is then $\widehat{Y}_x = \mathbf{f}^\top(x)\widehat{\underline{\gamma}}$, where $\widehat{\underline{\gamma}} = (\mathbf{F}^\top\mathbf{K}_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{F}^\top\mathbf{K}_{err}^{-1}\mathbf{y}$, with $\mathbf{F} = (\mathbf{G}, \Phi_{trc})$ and $\Gamma^{-1} = \mathrm{diag}\{0, \ldots, 0, \lambda_1^{-1}, \ldots, \lambda_{n_{trc}}^{-1}\}$ where there are p zeros. Its IMSE coincides with the

truncated version of $\mathsf{IMSE}(\mathscr{D}_n)$ and is given by

$$\mathsf{IMSE}_{trc}(\mathscr{D}_n) = \mathrm{trace}[(\mathbf{F}^\top \mathbf{K}_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{U}]\,,$$

$$\text{where } \mathbf{U} = (\mathbf{f}|\mathbf{f}^\top)_{L^2_\mu} = \int_{\mathscr{X}} \mathbf{f}(x)\mathbf{f}^\top(x)\,\mathrm{d}\mu(x) = \begin{pmatrix} (\mathbf{g}|\mathbf{g}^\top)_{L^2_\mu} & (\mathbf{g}|\underline{\phi}_{trc}^\top)_{L^2_\mu} \\ (\underline{\phi}_{trc}|\mathbf{g}^\top)_{L^2_\mu} & \mathrm{Id}_{n_{trc}} \end{pmatrix},$$

with $\mathrm{Id}_{n_{trc}}$ the n_{trc}-dimensional identity matrix. Note that $\mathbf{U}$ does not depend on $\mathscr{D}_n$. Approximation of Y_x by a heteroscedastic BLM where uncorrelated errors with variance $\sigma^2(x) = K_{err}(x,x)$ are substituted for ε_x, gives a predictor with $\mathsf{IMSE}_{trc}^{hBLM}(\mathscr{D}_n) = \mathrm{tr}[(\mathbf{F}^\top \Sigma_{err}^{-1}\mathbf{F} + \Gamma^{-1})^{-1}\mathbf{U}]$, i.e., a Bayesian L-optimality criterion which can be minimized following the same lines as in Sect. 3. However, in general the realizations of $(Z_x)_{x\in\mathscr{X}}$ are not orthogonal to the linear subspace $\mathscr{T}$ spanned by the trend functions $\mathbf{g}$, so that, roughly speaking, this direct approach involves unnecessarily large error variances $\sigma^2(x)$.

Denote by $\mathfrak{p}$ the orthogonal projection of $L^2_\mu(\mathscr{X})$ onto $\mathscr{T}$: for any $f \in L^2_\mu(\mathscr{X})$, $\mathfrak{p}f = \mathbf{g}^\top(\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}(\mathbf{g}|f)_{L^2_\mu}$. Suppose for simplicity that the sample realizations of $(Z_x)_{x\in\mathscr{X}}$ belong to $L^2_\mu(\mathscr{X})$ and define $\mathfrak{p}Z_x = \mathbf{g}^\top(x)(\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}\int_{\mathscr{X}} \mathbf{g}(t)Z_t\,\mathrm{d}\mu(t)$, so that $Y_x = \mathbf{g}^\top(x)\underline{\theta} + Z_x = \mathbf{g}^\top(x)\underline{\theta} + \mathfrak{p}Z_x + \mathfrak{q}Z_x = \mathbf{g}^\top(x)\underline{\theta}' + \mathfrak{q}Z_x$, where $\mathfrak{q} = \mathrm{id}_{L^2_\mu} - \mathfrak{p}$ and $\underline{\theta}' = \underline{\theta} + (\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}\int_{\mathscr{X}} \mathbf{g}(t)Z_t\,\mathrm{d}\mu(t)$. The covariance kernel of $(\mathfrak{q}Z_x)_{x\in\mathscr{X}}$, called reduction of the kernel K in [12], is

$$K^{\mathfrak{q}}(x,x') = \mathsf{E}\{(\mathfrak{q}Z_x)(\mathfrak{q}Z_{x'})\} = K(x,x') + \mathbf{g}^\top(x)\mathbf{S}\mathbf{g}(x') - \mathbf{b}^\top(x)\mathbf{g}(x') - \mathbf{g}^\top(x)\mathbf{b}(x')\,,$$

where $\mathbf{S} = (\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}(T_\mu[\mathbf{g}]|\mathbf{g}^\top)_{L^2_\mu}(\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}$ and $\mathbf{b}(x) = (\mathbf{g}|\mathbf{g}^\top)^{-1}_{L^2_\mu}T_\mu[\mathbf{g}](x)$. A key property here is that, for a given design $\mathscr{D}_n$, the two random fields $\mathbf{g}^\top(x)\underline{\theta} + Z_x$ and $\mathbf{g}^\top(x)\underline{\theta} + \mathfrak{q}Z_x$ with unknown $\underline{\theta}$ yield the same BLUP and same IMSE. This can be related to results on intrinsic random functions, see [7]; notice, however, that the kernel reduction considered here applies to any linear parametric trend and is not restricted to polynomials in x.

Figure 2 presents the first four eigenfunctions for the kernels $K(x,x')$ (Left) and $K^{\mathfrak{q}}(x,x')$ (Right) for a one-dimensional random process on $[0,1]$ with Matérn 3/2 covariance function, see (5), and trend functions $\mathbf{g}(x) = (1,x,x^2)^\top$. The eigenfunctions oscillate at higher frequency for $K^{\mathfrak{q}}(x,x')$ than for $K(x,x')$: in some sense, kernel reduction has removed from K the energy which can be transferred to the slowly varying trend functions. This is confirmed by Fig. 3-Left which shows the first 20 eigenvalues for both kernels. One can thus expect that for the same number n_{trc} of eigenfunctions, an heteroscedastic BLM will be more accurate when based on $K^{\mathfrak{q}}$ than when based on K. This is confirmed by Fig. 3-Right where $\sigma^2(x) = K_{err}(x,x)$ is plotted for both kernels when $n_{trc} = 10$.

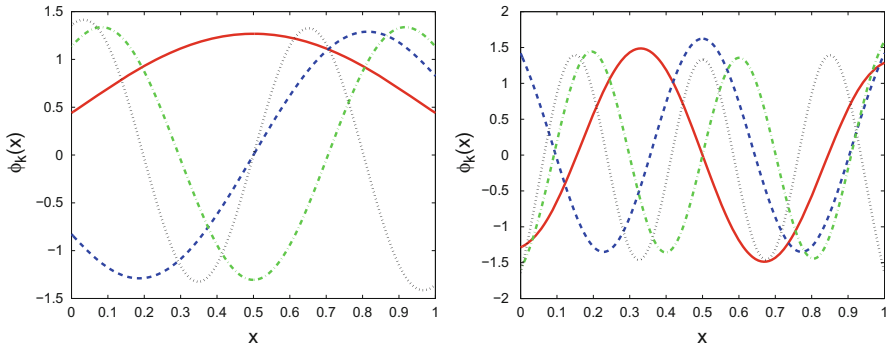

Fig. 2 First 4 eigenfunctions $\phi_k(x)$, $x \in \mathscr{X} = [0,1]$, for the kernel $K(x,x') = C_{3/2,10}(|x-x'|)$ (*Left*) and for $K^{\mathfrak{q}}(x,x')$ adapted to the trend functions $\mathbf{g}(x) = (1,x,x^2)^\top$ (*Right*)

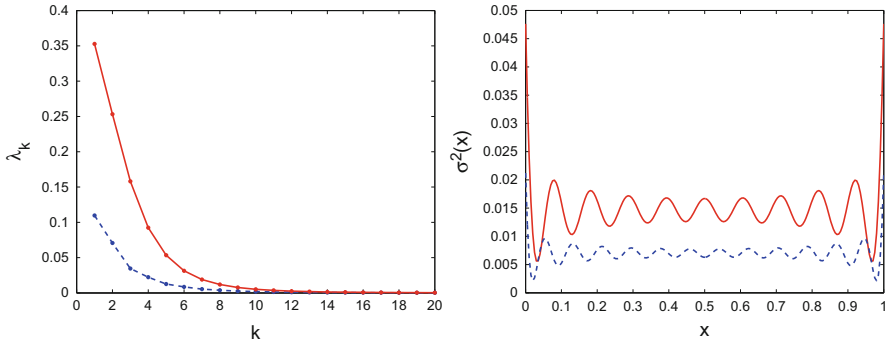

Fig. 3 *Left*: first 20 eigenvalues (sorted by decreasing values) for the kernel $K(x,x') = C_{3/2,10}(|x-x'|)$ (*top*, *solid line*) and for $K^{\mathfrak{q}}(x,x')$ adapted to the trend functions $\mathbf{g}(x) = (1,x,x^2)^\top$ (*bottom*, *dashed line*). *Right*: $\sigma^2(x)$, $x \in \mathscr{X} = [0,1]$, for the heteroscedastic BLM with $K(x,x')$ (*top*, *solid line*) and with $K^{\mathfrak{q}}(x,x')$ (*bottom*, *dashed line*); $n_{trc} = 10$ in both cases

Acknowledgements This work was partially supported by the ANR project DESIRE (DESIgns for spatial Random fiElds), nb. 2011-IS01-001-01, joint with the Statistics Dept. of the JKU Linz (Austria). B. Gauthier has also been supported by the MRI Dept. of EdF Chatou from Sept. to Dec. 2014.

References

1. Armijo, L.: Minimization of functions having Lipschitz continuous first partial derivatives. Pac. J. Math. **16**, 1–3 (1966)
2. Fedorov, V.V.: Design of spatial experiments: model fitting and prediction. In: Gosh, S., Rao, C. (eds.) Handbook of Statistics, vol. 13, chap. 16, pp. 515–553. Elsevier, Amsterdam (1996)

3. Fedorov, V.V., Müller, W.G.: Optimum design for correlated processes via eigenfunction expansions. In: López-Fidalgo, J., Rodríguez-Díaz, J.M., Torsney, B. (eds.) mODa'8 – Advances in Model–Oriented Design and Analysis, Proceedings of the 8th International Workshop, Almagro, pp. 57–66. Physica Verlag, Heidelberg (2007)
4. Gauthier, B., Pronzato, L.: Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. SIAM/ASA J. Uncertain. Quantif. **2**, 805–825 (2014)
5. Gauthier, B., Pronzato, L.: Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition. Commun. Stat. – Simul. Comput. (2015, to appear). doi:10.1080/03610918.2014.972518
6. Harari, O., Steinberg, D.M.: Optimal designs for Gaussian process models via spectral decomposition. J. Stat. Plan. Inference **154**, 87–101 (2014)
7. Matheron, G.: The intrinsic random functions and their applications. Adv. Appl. Probab. **5**(3), 439–468 (1973)
8. Pilz, J.: Bayesian Estimation and Experimental Design in Linear Regression Models. Teubner-Texte zur Mathematik, Leipzig (1983); also Wiley, New York (1991)
9. Pronzato, L., Pázman, A.: Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria and Small-Sample Properties, LNS 212. Springer, New York/Heidelberg (2013)
10. Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P.: Design and analysis of computer experiments. Stat. Sci. **4**, 409–423 (1989)
11. Santner, T.J., Williams, B.J., Notz, W.: The Design and Analysis of Computer Experiments. Springer, New York (2003)
12. Schaback, R.: Native Hilbert spaces for radial basis functions I. In: New Developments in Approximation Theory. International Series of Numerical Mathematics, vol. 32, pp. 255–282. Birkhäuser, Basel (1999)
13. Spöck, G., Pilz, J.: Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. Stoch. Environ. Res. Risk Assess. **24**, 463–482 (2010)

Asymptotic Properties of an Adaptive Randomly Reinforced Urn Model

Andrea Ghiglietti

Abstract In the design of experiments, urn models have been widely used as randomization devices to allocate subjects to treatments and incorporate ethical constraints. We propose a new adaptive randomly reinforced urn design, in a clinical trial context. The design consists of a randomly reinforced urn wherein a sequential allocation of patients to treatments is performed and the associated responses are collected. The model is based on two stochastic sequences representing random and time-dependent thresholds for the urn proportion process. These thresholds are defined as functions of the estimators of unknown parameters modeling the response distributions, so that they change any time a new response is collected. First and second-order asymptotic results under different conditions have been investigated. Specifically, we present the limit, the rate of convergence and the asymptotic distribution of the proportion of subjects assigned to the treatments.

1 Introduction

This paper presents some recent developments concerning the asymptotic properties of an important class of urn designs that can be adopted in several areas of application. In particular, the framework here considered is represented by clinical trials aimed at comparing two competing treatments ($\mathbb{T}_1$ and $\mathbb{T}_2$) where the subjects that sequentially enter the experiment are randomly assigned to one of the two treatments, according to the color of the balls sampled from an urn. The use of urn designs to construct randomized procedures is very common in applications to clinical trials, since their flexibility covers a large class of models and this guarantees the achievement of multiple objectives (e.g. see [6, 12, 13, 17]). Specifically, I consider *response-adaptive* urn models, in which the probability of assignment depends on the treatment performances (for a review, see [5, 12]). In the response-adaptive urn models the composition of the urn is sequentially updated taking into account the responses to treatments of the patients. The choice of the updating rule determines the type of urn model adopted in the experiment and hence the statistical

A. Ghiglietti (✉)
Università degli Studi di Milano, via Saldini 50, Milan, Italy
e-mail: andrea.ghiglietti@unimi.it

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_14

properties of the design. This paper focuses on *randomly reinforced* designs, in which only the composition of the color sampled from the urn is updated by a positive random quantity. First, I consider the randomly reinforced urn (RRU) design (see [14, 16]); then, I introduce an adaptive version of the RRU (ARRU) (see [4, 10]), that is able to target any desired allocation. For both these urn models, a complete description of the asymptotic behavior of the proportion of subjects assigned to the treatments is provided: results of consistency, rate of convergence and asymptotic distributions are presented for both equal and different means of the responses to the treatments.

2 Randomly Reinforced Urn Design

We now describe the RRU model. The responses of the subjects to treatments $\mathbb{T}_1$ and $\mathbb{T}_2$ are modeled by two sequences $\boldsymbol{\xi_1} = \{\xi_{1,n}; n \geq 1\}$ and $\boldsymbol{\xi_2} = \{\xi_{2,n}; n \geq 1\}$, respectively, of i.i.d. random variables. Denote by S a set that includes the supports of $\xi_{1,n}$ and $\xi_{2,n}$ and let $u : S \to [a, b]$, $0 < a \leq b < \infty$, be a utility function chosen to obtain the reinforcements from the response. Initially, consider an urn containing $y_{1,0} > 0$ red balls and $y_{2,0} > 0$ white balls, that in general may not be integers. At time $n = 1$, a ball is sampled from the urn and its color is observed. Let $X_1 = 1$ if the sampled ball is red, while $X_1 = 0$ if the sampled ball is white. We assume X_1 to be independent of the sequences $\boldsymbol{\xi_1}$ and $\boldsymbol{\xi_2}$ and hence X_1 is Bernoulli distributed with parameter $z_0 := y_0^{-1} y_{1,0}$, where $y_0 = y_{1,0} + y_{2,0}$. If $X_1 = 1$, the first subject receives treatment $\mathbb{T}_1$ and the response $\xi_{1,1}$ is observed; then, the extracted ball is returned on the urn together with $D_{1,1} = u(\xi_{1,1})$ new red balls. While if $X_1 = 0$, the first subject receives treatment $\mathbb{T}_2$, the response $\xi_{2,1}$ is observed and the extracted ball is returned on the urn together with $D_{2,1} = u(\xi_{2,1})$ new white balls. Formally, the sampled ball is always replaced in the urn together with $X_1 D_{1,1} + (1 - X_1) D_{2,1}$ new balls of the same color; hence, the urn composition is updated as follows

$$\begin{cases} Y_{1,1} = y_{1,0} + X_1 D_{1,1} \\ \\ Y_{2,1} = y_{2,0} + (1 - X_1) D_{2,1}; \end{cases}$$

thus, set $Y_1 = Y_{1,1} + Y_{2,1}$, $Z_1 = Y_1^{-1} Y_{1,1}$ and the process is repeated for all $n \geq 1$. Define the σ-algebra $\mathscr{F}_n$ generated by the allocations and the responses observed up to time n, i.e.

$$\mathscr{F}_n := \sigma\big(X_1, X_1 u(\xi_{1,1}) + (1 - X_1)\, u(\xi_{2,1}), \ldots, X_n, X_n u(\xi_{1,n}) + (1 - X_n)\, u(\xi_{2,n})\big).$$

At time $n + 1$ when a ball is sampled, let $X_{n+1} = 1$ if the ball is red and $X_{n+1} = 0$ otherwise. If $X_{n+1} = 1$, subject $n + 1$ receives treatment $\mathbb{T}_1$ and the response $\xi_{1,n+1}$ is observed; while, if $X_{n+1} = 0$, subject $n + 1$ receives treatment $\mathbb{T}_2$ and

the response $\xi_{2,n+1}$ is observed. Then, the ball is returned on the urn together with $X_{n+1}D_{1,n+1} + (1 - X_{n+1})\, D_{2,n+1}$ balls of the same color, where $D_{1,n+1} = u\,(\xi_{1,n+1})$ and $D_{2,n+1} = u\,(\xi_{2,n+1})$. Hence, the urn composition is updated as follows

$$\begin{cases} Y_{1,n+1} = y_{1,0} + \sum_{i=1}^{n+1} X_i D_{1,i}, \\ \\ Y_{2,n+1} = y_{2,0} + \sum_{i=1}^{n+1} (1 - X_i)\, D_{2,i}. \end{cases}$$

Finally, set $Y_{n+1} = Y_{1,n+1} + Y_{2,n+1}$ and $Z_{n+1} = Y_{1,n+1}/Y_{n+1}$. Note that X_{n+1}, conditionally on the σ-algebra $\mathscr{F}_n$, is Bernoulli distributed with parameter Z_n, since X_{n+1} is conditionally on $\mathscr{F}_n$ assumed to be independent of ξ_1 and ξ_2.

2.1 Asymptotic Results for an RRU Design

Here, the main asymptotic results concerning the proportion of subjects assigned to the treatments for an RRU design are presented. Let $N_{1n} := \sum_{i=1}^{n} X_i$ be the number of subjects that receives treatment $\mathbb{T}_1$ up to time n. Denote by $m_1 := \mathbf{E}[D_{1,1}]$ and $m_2 := \mathbf{E}[D_{2,1}]$ the reinforcement means. The convergence of the urn proportion for an RRU was proved in [7] for binary responses and extended in [16] for continuous responses as follows:

$$Z_n = \frac{Y_{1,n}}{Y_{1,n} + Y_{2,n}} \xrightarrow{a.s.} \begin{cases} 1 & \text{if } m_1 > m_2, \\ Z_\infty & \text{if } m_1 = m_2, \\ 0 & \text{if } m_1 < m_2, \end{cases} \tag{1}$$

where Z_∞ is a non-degenerate random variable with support $(0, 1)$. While for binary responses Z_∞ is Beta-distributed (see [7]), the properties of Z_∞ for general responses were studied in [1, 2]. Specifically, it was shown in [1] that, when $m_1 = m_2$, $\mathbf{P}(Z_\infty = x) = 0$ for any $x \in [0, 1]$. From (1), one can deduce the convergence of $N_{1,n}/n$ to the same limit of the urn proportion Z_n.

We now focus on the second-order asymptotic properties of $N_{1,n}/n$ for an RRU model. In the case $m_1 \neq m_2$, the rate of convergence and the limit distribution was established in [7] for binary responses and extended in [14] for continuous responses. Specifically, letting $m_1 > m_2$, the authors showed that

$$n^{1-\frac{m_2}{m_1}} \left(1 - \frac{N_{1n}}{n}\right) \xrightarrow{d} \eta^2,$$

where η^2 is a positive random variable.

In the case $m_1 = m_2$, it was established in [4] that the rate of convergence is $\sqrt{n}$ and the asymptotic distribution of $\sqrt{n}\left(\frac{N_{1n}}{n} - Z_\infty\right)$, conditionally on the limiting

proportion Z_∞, is Gaussian. Formally, this result is concerned with the concept of stable convergence (see [11]), that is recalled as follows: denoting by $\{\mathscr{X}_n; n \geq 1\}$ a random sequence on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$, we say that $\mathscr{X}_n \xrightarrow{d} \mathscr{X}$ (stably) if, for every point x of continuity for the cumulative distribution function of $\mathscr{X}$ and for every event $E \in \mathscr{F}$,

$$\lim_{n\to\infty} \mathbf{P}\left(\mathscr{X}_n \leq x, E\right) = \mathbf{P}\left(\mathscr{X} \leq x, E\right).$$

Hence, the central limit theorem for the RRU is expressed in the following result (see Theorem 2.6 in [4]), in which a stable convergence with $\mathscr{F} = \sigma(Z_\infty)$ is presented.

Theorem 1 *Consider an RRU model. Assume* $m_1 = m_2 = m$ *and let* $\sigma_1^2 := Var[D_{1,1}]$ *and* $\sigma_2^2 := Var[D_{2,1}]$. *Then,*

$$\sqrt{n}\left(\frac{N_{1n}}{n} - Z_\infty\right) \xrightarrow{d} \mathscr{N}(0, \Sigma), \quad (\textit{stably})$$

where

$$\Sigma := \left(1 + \frac{2\bar{\Sigma}}{m^2}\right) Z_\infty(1 - Z_\infty), \qquad \bar{\Sigma} := (1 - Z_\infty)\sigma_1^2 + Z_\infty\sigma_2^2. \tag{2}$$

It is worth highlighting that the limiting distribution obtained in Theorem 1 is Gaussian only conditionally on Z_∞, otherwise it represents a mixture of normal distributions weighted by the probability law of Z_∞. Note also that, in the case of deterministic reinforcements, i.e. Pòlya's urn, the asymptotic variance Σ in (2) reduces to $Z_\infty(1 - Z_\infty)$.

3 Adaptive Randomly Reinforced Urn Design

In an RRU model, when $\mathbb{T}_1$ ($\mathbb{T}_2$) is the superior treatment, i.e. $m_1 > m_2$ ($m_1 < m_2$), the asymptotic proportion of subjects assigned to treatment $\mathbb{T}_1$ ($\mathbb{T}_2$) is always 1. Although this represents a good ethical aspect, the statistical properties associated with such a design are typically quite poor (e.g. low power). For this reason, in [3] the RRU model has been modified by using two fixed thresholds, $0 < \rho_2 \leq \rho_1 < 1$, such that: if $Z_n < \rho_2$, no white balls are replaced in urn, while if $Z_n > \rho_1$, no red balls are replaced in the urn. The consistency for this modified RRU (MRRU) in the case $m_1 \neq m_2$ was established in [3]: $Z_n \xrightarrow{a.s.} \rho_1$ when $m_1 > m_2$, and $Z_n \xrightarrow{a.s.} \rho_2$ when $m_1 < m_2$. Moreover, a second order result for Z_n, namely the asymptotic distribution of Z_n after appropriate centering, was derived in [8]. An application of the MRRU to construct efficient two-sample tests aimed at comparing the response means was proposed in [9].

Although the MRRU model provides an asymptotic allocation proportion, ρ_1 or ρ_2, in $(0,1)$, it is common in clinical trials to target specific values in order to accomplish certain objectives (see [12]). Hence, ρ_1 and ρ_2 should depend on the unknown parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta$ that characterize the distributions of $D_{1,1}$ and $D_{2,1}$, respectively. To this end, in [10] an adaptive RRU (ARRU) model was introduced, in order to use accruing statistical information to skew the proportion towards arbitrary values. This model is defined by two continuous functions $f_1 : \Theta \to (0,1)$ and $f_2 : \Theta \to (0,1)$, such that $f_1(\boldsymbol{\theta}) \geq f_2(\boldsymbol{\theta})$ for any $\boldsymbol{\theta} \in \Theta$. The quantities $\rho_1 := f_1(\boldsymbol{\theta})$ and $\rho_2 := f_2(\boldsymbol{\theta})$ represent the desired limiting allocations for $N_{1,n}/n$, when the superior treatment is $\mathbb{T}_1$ $(m_1 > m_2)$ or $\mathbb{T}_2$ $(m_1 < m_2)$, respectively. The implementation of the ARRU model requires a sequence of random variables defined as $\hat{\rho}_{1,n} := f_1\left(\hat{\boldsymbol{\theta}}_{1,n}\right)$ and $\hat{\rho}_{2,n} := f_2\left(\hat{\boldsymbol{\theta}}_{2,n}\right)$ for any $n \geq 0$, where $\hat{\boldsymbol{\theta}}_{1,n}$ and $\hat{\boldsymbol{\theta}}_{2,n}$ indicate the corresponding adaptive estimators of the parameters.

The sequences $\{\hat{\rho}_{1,n}; n \geq 1\}$ and $\{\hat{\rho}_{2,n}; n \geq 1\}$ are random *thresholds* for the process of the urn proportion. Indeed, for any $n \geq 0$, at any time $n+1$ the updating rule is now changed as follows. If $X_{n+1} = 1$ and $Z_n \leq \hat{\rho}_{1,n}$, the extracted ball is returned on the urn together with $D_{1,n+1} = u(\xi_{1,n+1})$ new red balls. While, if $X_{n+1} = 0$ and $Z_n \geq \hat{\rho}_{2,n}$, the extracted ball is returned on the urn together with $D_{2,n+1} = u(\xi_{2,n+1})$ new white balls. If $X_{n+1} = 1$ and $Z_n > \hat{\rho}_{1,n}$, or if $X_{n+1} = 0$ and $Z_n < \hat{\rho}_{2,n}$, the urn composition is not modified. To ease notation, let denote $W_{1,n} = \mathbf{1}_{\{Z_n \leq \hat{\rho}_{1,n}\}}$ and $W_{2,n} = \mathbf{1}_{\{Z_n \geq \hat{\rho}_{2,n}\}}$. Formally, the sampled ball is returned on the urn together with $X_{n+1} D_{1,n+1} W_{1,n} + (1 - X_{n+1}) D_{2,n+1} W_{2,n}$ balls of the same color; hence, the urn composition is updated as follows

$$\begin{cases} Y_{1,n+1} = y_{1,0} + \sum_{i=1}^{n+1} X_i D_{1,i} W_{1,i-1}, \\ \\ Y_{2,n+1} = y_{2,0} + \sum_{i=1}^{n+1} (1 - X_i) D_{2,i} W_{2,i-1}. \end{cases}$$

Finally, set $Z_{n+1} = Y_{1,n+1}/Y_{n+1}$ and $Y_{n+1} = Y_{1,n+1} + Y_{2,n+1}$. We now present the asymptotic results for an ARRU.

3.1 Consistency for an ARRU Model

The following result (see Theorem 2.3 in [4]) shows the consistency of the urn proportion Z_n for any values of m_1 and m_2, when the random thresholds $\hat{\rho}_{1,n}$ and $\hat{\rho}_{2,n}$ converge with probability one.

Theorem 2 *Assume*

$$\hat{\rho}_{1,n} \xrightarrow{a.s.} \rho_1, \qquad \hat{\rho}_{2,n} \xrightarrow{a.s.} \rho_2. \tag{3}$$

Then,

$$Z_n \xrightarrow{a.s.} \begin{cases} \rho_1 & \text{if } m_1 > m_2, \\ Z_\infty & \text{if } m_1 = m_2, \\ \rho_2 & \text{if } m_1 < m_2, \end{cases} \tag{4}$$

where Z_∞ is a random variable such that $\mathbf{P}(Z_\infty \in [\rho_2, \rho_1]) = 1$, and, for any $x \in (\rho_2, \rho_1)$, we have $\mathbf{P}(Z_\infty = x) = 0$.

From (4), one can deduce the consistency of $N_{1,n}/n$ to the same limit as Z_n. Note that (3) is satisfied when the adaptive estimators $\hat{\boldsymbol{\theta}}_{1,n}$ and $\hat{\boldsymbol{\theta}}_{2,n}$ are consistent since the ARRU satisfies the assumptions of Theorem 3.1 in [15].

3.2 Asymptotic Distribution for an ARRU Model

We now focus on the asymptotic distribution for the proportion of sampled balls in an ARRU model. For the case $m_1 = m_2$, it was established in [4] that, conditionally on $Z_\infty \neq \{\rho_2, \rho_1\}$, the limiting distribution of the urn proportion is the same as an RRU (see Theorem 1). To this end, focus on those processes for which asymptotically, $\{Z_n \in A_n\}$, where $\{A_n; n \geq 1\}$ is a sequence of random sets such that $\cup_{n\geq 1} A_n = (\rho_2, \rho_1)$. For instance, in [4] the authors considered $A_n := (\rho_2 + CY_n^{-\alpha}, \rho_1 - CY_n^{-\alpha})$, with $0 < C < \infty$ and $0 < \alpha < 1/2$. Thus, the asymptotic distribution for the ARRU model is expressed in the following result (see Theorem 2.7 in [4]).

Theorem 3 *Assume (3) with $\rho_1 > \rho_2$ and $m_1 = m_2 = m$. Let $\sigma_1^2 := Var[D_{1,1}]$ and $\sigma_2^2 := Var[D_{2,1}]$. Then, on the sequence of sets $\{Z_n \in A_n\}$, $n \geq 1$, we have*

$$\sqrt{n}\left(\frac{N_{1n}}{n} - Z_\infty\right) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad (\textit{stably})$$

where, as in (2),

$$\Sigma = \left(1 + \frac{2\bar{\Sigma}}{m^2}\right) Z_\infty(1 - Z_\infty), \qquad \bar{\Sigma} = (1 - Z_\infty)\sigma_1^2 + Z_\infty \sigma_2^2.$$

It is worth noticing that the limiting distribution obtained in Theorems 1 and 3 is not Gaussian but a mixture distribution.

The asymptotic distribution for $\frac{N_{1n}}{n}$ in the case $m_1 \neq m_2$ has been investigated in [10]. Without loss of generality, let $m_1 > m_2$. Frequent changes to the thresholds $\hat{\rho}_{1,n}$ make the sequence $\frac{N_{1,n}}{n}$ too erratic to obtain a stable distribution, so consider thresholds that are updated "slowly" at exponential times $\{\lceil q^i \rceil, i \geq 0\}$, with $q > 1$.

Formally, for any $n \geq 0$, $\hat{\rho}_{j,n}$ is replaced by

$$\tilde{\rho}_{j,n} := \hat{\rho}_{j,\lfloor q^i \rfloor}, \quad as \ \lfloor q^i \rfloor \leq n < \lfloor q^{i+1} \rfloor, \ i \in \mathbb{N}. \tag{5}$$

Thus, the limiting distribution is established in the following result (see Theorem 2.2 in [10]) by assuming that the convergence of the thresholds is exponentially fast.

Theorem 4 *Consider an ARRU model with thresholds defined as in* (5)*. Assume that for any* $\epsilon > 0$*, there exists* $0 < c_1 < \infty$ *such that*

$$\mathbf{P}\left(|\hat{\rho}_{1,n} - \rho_1| > \epsilon\right) \ \leq \ c_1 \exp\left(-n\epsilon^2\right), \tag{6}$$

for large n. Then, letting $\bar{\rho}_{1,n} = \frac{\sum_{i=1}^{n} \tilde{\rho}_{1,i-1}}{n}$*, if* $m_1 > m_2$

$$\sqrt{n}\left(\frac{N_{1,n}}{n} - \bar{\rho}_{1,n}\right) \ \xrightarrow{d} \ \mathcal{N}\big(0, \rho_1(1-\rho_1)\big).$$

Note that Theorem 3 expresses the CLT for the MRRU with $\bar{\rho}_{1,n} = \rho_1$ for any $n \geq 1$.

3.3 Conclusions

This paper presents a collection of results that completely describes the asymptotic behavior of N_{1n}/n for a wide class of reinforced urn models used as response-adaptive designs for clinical trials. In particular, the urn models here considered are (i) the RRU design and (ii) the ARRU design.

Table 1 gathers the consistency results. Note that, when $m_1 > m_2$, the RRU asymptotically targets the best treatment, while the ARRU achieves any allocation defined as a function of the response distribution's parameters. For both these models, when $m_1 = m_2$ the urn proportion converges to a random variable Z_∞.

Table 2 gathers the second-order asymptotic results of N_{1n}/n. When $m_1 = m_2$, the asymptotic distribution is a mixture of Gaussian distributions determined by Z_∞. When $m_1 > m_2$, the rate of convergence is n^{1-m_2/m_1} for an RRU, while it is always $\sqrt{n}$ for an ARRU. In fact, in the RRU the convergence rate of $N_{1,n}/n$ reflects the rate of Z_n, which depends on m_1 and m_2 (see [14]), while in the ARRU, since some subjects are assigned when the urn composition does not change, the convergence

Table 1 Consistency results of N_{1n}/n for the RRU and the ARRU designs

Consistency	RRU	ARRU
$m_1 = m_2$	$\frac{N_{1n}}{n} \xrightarrow{a.s.} Z_\infty \in (0,1)$ (see [7, 16])	$\frac{N_{1n}}{n} \xrightarrow{a.s.} Z_\infty \in [\rho_2, \rho_1]$ (cf. Theorem 2)
$m_1 > m_2$	$\frac{N_{1n}}{n} \xrightarrow{a.s.} 1$ (see [7, 16])	$\frac{N_{1n}}{n} \xrightarrow{a.s.} \rho_1$ (cf. Theorem 2)

Table 2 Asymptotic distribution and rate of convergence of N_{1n}/n for the RRU and the ARRU design

Asymptotic distribution	RRU	ARRU
$m_1 = m_2$	$\sqrt{n}\left(\frac{N_{1n}}{n} - Z_\infty\right)$ $\xrightarrow{d} \mathcal{N}(0, \Sigma)$ (cf. Theorem 1)	On the sets $(\{Z_n \in A_n\}, n \geq 1)$ $\sqrt{n}\left(\frac{N_{1n}}{n} - Z_\infty\right) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ (cf. Theorem 3)
$m_1 > m_2$	$n^{1-\frac{m_2}{m_1}}\left(1 - \frac{N_{1n}}{n}\right) \xrightarrow{d} \eta^2$ (see [14])	$\sqrt{n}\left(\frac{N_{1,n}}{n} - \bar{\rho}_{1,n}\right) \xrightarrow{d} \mathcal{N}(0, \rho_1(1-\rho_1))$ (cf. Theorem 4)

rate of $N_{1,n}/n$ and Z_n may be different. In [10] it was discussed that the rate of Z_n in the ARRU is an open problem and it depends on the behavior of $\hat{\rho}_{1,n}$ and $\hat{\rho}_{2,n}$. However, when the threshold sequence is defined as in (5) and it converges as in (6), it is possible to show (Theorem 3) that for $N_{1,n}/n$ the convergence rate is $\sqrt{n}$ and the asymptotic distribution does not depend on m_1 and m_2. It is worth highlighting that a CLT for Z_n with convergence rate n was established for the case of fixed thresholds (MRRU) in [8]. However, since the fixed thresholds satisfy (5) and (6), the CLT for $N_{1,n}/n$ is the same in the MRRU and ARRU model.

References

1. Aletti, G., May, C., Secchi, P.: A central limit theorem, and related results, for two-color randomly reinforced urn. Ann. Appl. Probab. **41**, 829–844 (2009)
2. Aletti, G., May, C., Secchi, P.: A functional equation whose unknown is $P([0; 1])$ valued. J. Theor. Probab. **25**, 1207–1232 (2012)
3. Aletti, G., Ghiglietti, A., Paganoni, A.M.: A modified randomly reinforced urn design. J. Appl. Probab. **50**, 486–498 (2013)
4. Aletti, G., Ghiglietti, A., Vidyashankar, A.N.: Dynamics of an Adaptive Randomly Reinforced Urn. Technical report (2015). arXiv:1508.02629
5. Atkinson, A.C., Biswas, A.: Randomised Response-Adaptive Designs in Clinical Trials. Chapman and Hall/CRC, Boca Raton (2013)
6. Baldi Antognini, A., Giannerini, S.: Generalized Polya urn designs with null balance. J. Appl. Probab. **44**, 661–669 (2007)
7. Durham, S.D., Flournoy, N., Li, W.: A sequential design for maximazing the probability of a favourable response. Can. J. Stat. **26**(3), 479–495 (1998)
8. Ghiglietti, A., Paganoni, A.M.: Statistical properties of two-color randomly reinforced urn design targeting fixed allocations. Electron. J. Stat. **8**, 708–737 (2014)
9. Ghiglietti, A., Paganoni, A.M.: An urn model to construct an efficient test procedure for response adaptive designs. Stat. Methods Appl. (2014). doi:10.1007/s10260-015-0314-y
10. Ghiglietti, A., Vidyashankar, A.N., Rosenberger, W.F.: Central Limit Theorem for an Adaptive Ranldomly Reinforced Urn Model. Technical report (2015). arXiv:1502.06130
11. Hall, P., Heyde, C.C.: Martingale Limit Theory and its Application. Academic Press, New York (1980)
12. Hu, F., Rosenberger, W.F.: The Theory of Response-Adaptive Randomization in Clinical Trials. Wiley, New York (2006)

13. Lachin, J.M., Rosenberger, W.F.: Randomization in Clinical Trials: Theory and Practice. Wiley, New York (2002)
14. May, C., Flournoy, N.: Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. Ann. Stat. **37**(2), 1058–1078 (2009)
15. Melfi, F., Page, C.: Estimation after adaptive allocation. J. Plan. Stat. Inference **87**, 353–363 (2000)
16. Muliere, P., Paganoni, A.M., Secchi, P.: A randomly reinforced urn. J. Stat. Plan. Inference **136**, 1853–1874 (2006)
17. Zhang, L.-X., Hu, F., Cheung, S.H., Chan, W.S.: Immigrated urn models: theoretical properties and applications. Ann. Stat. **39**, 643–671 (2011)

Design of Computer Experiments Using Competing Distances Between Set-Valued Inputs

David Ginsbourger, Jean Baccou, Clément Chevalier, and Frédéric Perales

Abstract In many numerical simulation experiments from natural sciences and engineering, inputs depart from the classical moderate-dimensional vector set-up and include more complex objects such as parameter fields or maps. In this case, and when inputs are generated using stochastic methods or taken from a pre-existing large set of candidates, one often needs to choose a subset of "representative" elements because of practical restrictions. Here we tackle the design of experiments based on distances or dissimilarity measures between input maps, and more specifically between inputs of set-valued nature. We consider the problem of choosing experiments given dissimilarities such as the Hausdorff or Wasserstein distances but also of eliciting adequate dissimilarities not only based on practitioners' expertise but also on quantitative and graphical diagnostics including nearest neighbour cross-validation and non-Euclidean structural analysis. The proposed approaches are illustrated on an original uncertainty quantification case study from mechanical engineering, where using partitioning around medoids with ad hoc distances gives promising results in terms of stratified sampling.

D. Ginsbourger (✉)
Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, 1920 Martigny, Switzerland

IMSV, Department of Mathematics and Statistics, University of Bern, Alpeneggstrasse 22, 3012 Bern, Switzerland
e-mail: ginsbourger@idiap.ch; ginsbourger@stat.unibe.ch

J. Baccou • F. Perales
Institut de Radioprotection et de Sûreté Nucléaire, PSN-RES, SEMIA, Centre de Cadarache, France

Laboratoire de Micromécanique et d'Intégrité des Structures, IRSN-CNRS-UM, 13115 Saint-Paul-lès-Durance, France
e-mail: jean.baccou@irsn.fr; frederic.perales@irsn.fr

C. Chevalier
Institut de Statistique, Université de Neuchâtel, Av. de Bellevaux 51, 2000 Neuchâtel, Switzerland

Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
e-mail: clement.chevalier@unine.ch

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_15

1 Introduction

Here we consider a function $f : \mathbf{x} \in D \longrightarrow f(\mathbf{x}) \in \mathbb{R}$ stemming from some expensive deterministic computer experiment, where the input space D can possibly be a subset of $\mathbb{R}^q$ or some more complicated set of structured objects such as curves, maps, or trees. A main prerequisite is that D is endowed with a distance d, or more generally with a *dissimilarity* $\delta : (\mathbf{x}, \mathbf{y}) \in D \times D \longrightarrow \mathbb{R}$, a function satisfying $\forall \mathbf{x}, \mathbf{y} \in D, \delta(\mathbf{x}, \mathbf{y}) \geq 0, \delta(\mathbf{x}, \mathbf{x}) = 0$, and that reflects to some extent how different the outcome is supposed to be for any given couple of inputs. Here we also require the symmetry condition $\forall \mathbf{x}, \mathbf{y} \in D, \delta(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{y}, \mathbf{x})$ [11]. Depending on the context, δ may either be prescribed by expert knowledge, learnt based on data, or a combination of both. Here we are interested in the use of such dissimilarities for the design of experiments, with a mechanical engineering case study serving as motivating example, and a main focus on two interrelated questions. The first question, assuming that several candidate dissimilarity functions are available, is how to choose the most relevant one in order to model and predict the response of interest. Addressing it will lead us to consider some quantitative and graphical diagnostics. The second one, assuming δ is given, is how to extract a sub-sample from a sample of inputs, with the aim to suitably represent the distribution of responses. Our proposed approach consists in appealing to a clustering algorithm such as *Partitioning Around Medoids* (PAM) with respect to the chosen δ. In the considered test case, inputs are modelled as point sets, and several candidate distances between sets (Hausdorff, Wasserstein, and variants thereof) are considered. We appeal to non-Euclidean structural analysis concepts from spatial statistics [4] and also to nearest-neighbour approaches for studying the adequacy of dissimilarities and diagnosing what may reasonably be expected when appealing to "distance methods" based on a reference sample and candidate dissimilarities. The paper is organized as follows. In Sect. 2, we describe the motivating case study. In Sect. 3, we review a few selected distance methods and present the specific dissimilarities (distances between point sets) considered for the case study. In Sect. 4, we first consider a dissimilarity-based variography framework and display experimental results obtained on test case data. Then a nearest-neighbour approach serves to produce additional diagnostics. The candidate distances are used to perform stratified subsampling relying on PAM. Conclusions and perspectives as well as a list of references are presented in Sect. 5.

2 Motivating Case Study

In the framework of a research program at the French Institute for Radiological Protection and Nuclear Safety, mechanical simulations are performed with the CASTEM code [2] in order to calculate equivalent stresses on biphasic materials subjected to uni-axial traction. The system, an elastoplastic matrix containing elastic inclusions, is modelled as a unit square containing p circular inclusions, all with the

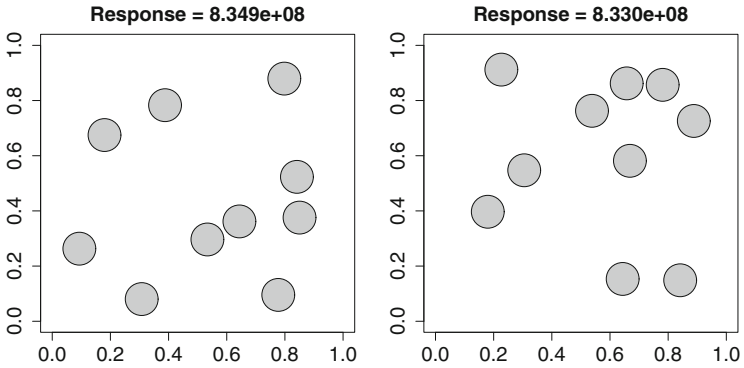

Fig. 1 Two randomly drawn CASTEM inputs with associated equivalent stress responses

same radius R. Macroscopic calculations are done by averaging over a large number of squares with random inclusion locations. Two examples of inclusion patterns with response values of equivalent von Mises stress are sketched on Fig. 1.

Simulations were all performed with input media containing $p = 10$ randomly distributed inclusions with constant radii ($R = 0.056419$). In such a context, an input configuration $\mathbf{x} = \{\mathbf{c}_1, \ldots, \mathbf{c}_p\}$ is parametrized by a set of 2-dimensional points $\mathbf{c}_i \in [0, 1]^2$ ($1 \leq i \leq p$). Given two arbitrary configurations $\mathbf{x}$ and $\mathbf{y}$, we aim at finding a relevant distance between them and making predictions of equivalent stress at inputs for which the simulation outcome is unknown, based on available simulation results. Of course, there are many ways of defining distances between such $\mathbf{x}$ and $\mathbf{y}$; and this richness can turn into a drawback when facing a choice between numerous candidates. In the next section, we principally focus on four candidate distances for the design of CASTEM simulations. The main application-driven objectives pursued here concern parsimoniously choosing sub-samples of inputs by relying on distance methods in order to get an empirical distribution of responses in the sub-sample close to the empirical distribution of responses in the larger sample.

3 Distance Methods: Basics and Considered Distances

Before focusing on the specifics of our test case, let us give a brief overview of distance methods from machine learning to spatial statistics. One of the most simple and yet efficient distance-methods for classification and prediction is the k-nearest neighbors algorithm (kNN). Assuming that the response of interest was observed for N instances $\mathbf{x}_1, \ldots, \mathbf{x}_N$ of the input, predicting the response at an arbitrary $\mathbf{x} \in D$ by kNN consists in averaging the responses of $\mathbf{x}$'s k nearest neighbors (among $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and in the sense of the chosen distance d), where $k \geq 1$ is a parameter of the algorithm. Beyond kNN, further distance methods have been

proposed for data analysis, including multi-dimensional scaling, phylogenetic trees, spectral clustering and more (see, e.g., [6] and references therein for an overview).

Distances are also key in spatial statistics [3, 5]. A spatially varying measurement of interest, say $z : \mathbf{x} \in D \longrightarrow \mathbb{R}$, is modelled as one realization of a random field $Z = \{Z_{\mathbf{x}}(\omega)\}_{\mathbf{x}\in D}$ ($\omega \in \Omega$, Ω standing for the underlying probability space). Assuming that Z's increments are squared integrable, $\gamma : (\mathbf{x}, \mathbf{y}) \in D^2 \longrightarrow \gamma(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\mathrm{Var}[Z_{\mathbf{x}} - Z_{\mathbf{y}}] \in \mathbb{R}$ exists and is called *semi-variogram* of Z. $\sqrt{2\gamma}$ then defines over D the so-called *canonical distance* associated with Z [1]. Given an arbitrary fixed distance d over D, when Z's mean is constant and $\gamma(\mathbf{x}, \mathbf{y})$ depends on $(\mathbf{x}, \mathbf{y})$ through $d(\mathbf{x}, \mathbf{y})$, Z and γ are called *isotropic*. Estimating how γ depends on d based on data is a difficult step, often referred to as *empirical variography*. In Sect. 4, we investigate distance methods and empirical variography in the case where $\mathbf{x}$ is point set ("configuration") and the underlying distances are chosen accordingly, as discussed now.

In our motivating case study we consider four candidate distances between input configurations, namely the Hausdorff [8] and Wasserstein [10] distances, together with ad hoc symmetrizations of them. In all cases, the configurations are represented by the locations of their respective centers. Denoting by $\mathbf{c}'_1, \ldots, \mathbf{c}'_p$ the centers of the configuration $\mathbf{y}$, the Hausdorff distance between $\mathbf{x}$ and $\mathbf{y}$ is defined in our case by

$$d_H(\mathbf{x}, \mathbf{y}) = \max\left(\max_{1\le i\le p}\min_{1\le j\le p} d_{\mathbb{R}^2}(\mathbf{c}_i, \mathbf{c}'_j), \max_{1\le j\le p}\min_{1\le i\le p} d_{\mathbb{R}^2}(\mathbf{c}_i, \mathbf{c}'_j)\right) \tag{1}$$

where $d_{\mathbb{R}^2}$ denotes the Euclidean distance over $\mathbb{R}^2$. The Hausdorff distance is quite popular in probability theory as it comes with a number of seminal mathematical results, but also in applied fields such as computer vision. Beyond d_H, another family of distances considered here, related to optimal transportation, are the so-called Wasserstein distances. Unlike Hausdorff distances, Wasserstein (or "earth mover's") distances do not only quantify the closeness of points of configurations to those of the other one, but also incorporate some more "physical" information on the cost to transform one configuration into the other one. Configurations (e.g., $\mathbf{x}$ and $\mathbf{y}$) are seen as "patterns" (i.e. finite support measures), and the distance between two patterns is calculated by pairing their respective points in such a way that each point of be in correspondence with a unique point of the other pattern. The distance is then defined based on the minimal value, over all possible pairings, of the average (be it in the L^2 sense or other) or maximal "ambient" (Euclidean, here) distances between paired points. We consider the following instance of this family of distances:

$$d_W(\mathbf{x}, \mathbf{y}) = \min_{\sigma\in\mathscr{S}_p} \sqrt{\frac{1}{p}\sum_{i=1}^{p} d^2_{\mathbb{R}^2}\left(\mathbf{c}_i, \mathbf{c}'_{\sigma(i)}\right)}, \tag{2}$$

where $\mathscr{S}_p$ is the set of permutations of $\{1, \ldots, p\}$. Additionally, some simple physical knowledge was used to design symmetrized distances dedicated to the test case. As the vertical force applied to the micro-structure is symmetrical with respect

to the $(x_1 = 0.5)$-axis, f is known to be invariant under the axial flip $v : (x_1, x_2) \in [0,1]^2 \longrightarrow (1 - x_1, x_2) \in [0,1]^2$. So we designed the corresponding symmetrized (or *quotient*) versions of d_L ($L \in \{W, H\}$): $d_{Lsym}(\mathbf{x}, \mathbf{y}) = \min\{d_L(\mathbf{x}, \mathbf{y}), d_L(\mathbf{x}, v(\mathbf{y}))\}$.

4 Application Results

Data Sets and Preliminary Results We consider two data sets, a preliminary 900-element one (*data set A*) and the main one (*data set B*), consisting of 404 instances of numerical simulation input/output tuples. Data set A was generated by sequentially drawing inclusion centers uniformly over $[0,1]^2$ with rejection in case of overlap between new and already included inclusions. While A was judged satisfactory in terms of space-fillingness, results (not displayed here by space limitation) from the targeted distance-based methods happened to be disappointing precisely because configurations turn out to be too "far" from each other to produce exploitable results, a consequence of the *curse of dimensionality* (see [9] and references therein). In order to get around this and produce a more informative data set (especially with respect to empirical variograms presented next), we conducted a new batch of simulations (data set B) with a more exploitable design. Data set B was generated by first sampling 116 random configurations following the same procedure as for data set A. Then the remaining configurations were generated by applying local perturbations (scrambling of randomly selected centers, with small to medium –i.e. 10^{-3}–10^{-2}– orders of magnitude). While there is some arbitrariness in the way B was designed, it has to be kept in mind that identifying modes of relevance (resp. of failure) of the proposed approaches fully falls in the objectives of this study. Besides this, we are interested in investigating which distance works best and how to uncover and/or quantify it based on empirical diagnostics. Diagnostics obtained on data set B via empirical variography and cross-validated kNN predictions are presented and discussed next. The section then presents some further results in stratification.

Diagnostics Based on Structural Analysis vs. kNN Cross-validated Predictions The first graphical diagnostic considered here is a set of four non-Euclidean empirical variograms based on the respective candidate distances. Denoting by d any arbitrary distance from $d_H, d_W, d_{Hsym}, d_{Wsym}$, we plot in Fig. 2 the corresponding empirical semi-variogram (*Matheron estimator*, see [3]) values:

$$\widehat{\gamma}(h_\ell) = \frac{1}{2\sharp N(h_\ell)} \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in N(h_\ell)} \big(f(\mathbf{x}_i) - f(\mathbf{x}_j)\big)^2, \tag{3}$$

where $N(h_\ell) = \big\{(\mathbf{x}_i, \mathbf{x}_j) : d(\mathbf{x}_i, \mathbf{x}_j) \in [\max(0, h_\ell - \Delta), h_\ell + \Delta]\big\}$ with a given tolerance parameter $\Delta > 0$, and $\sharp N(h_\ell)$ is the number of couples in $N(h_\ell)$. In Fig. 2, empirical semi-variograms are plotted with common values of h_ℓ (here $1 \le \ell \le 10$) for the four candidate distances; the coloured numbers near the points stand for the corresponding $\sharp N(h_\ell)$ values. A first comment regarding the left panel (Hausdorff

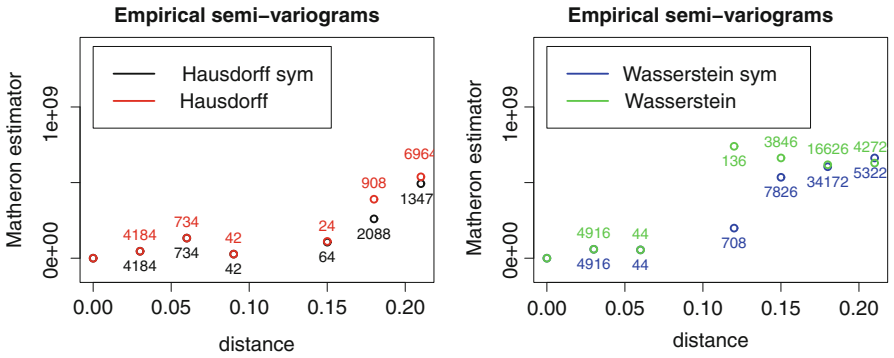

Fig. 2 Empirical variograms with respect to the four candidate distances (Hausdorff vs. Wasserstein, with possible symmetrization) of the equivalent constraint depending on inclusion centers

distances) is that taking the symmetry into account does not seem to affect things at short distances, while a slight departure is observed at the penultimate bin center. Focusing on the number of elements in the bins, the central regions of the empirical semi-variograms appear to be based on little information, and thus more faith should be put on their ends. As for Wasserstein distances (right panel), even if the central region creates the impression of a chaotic behaviour –especially for the d_W green curve–, it appears that the number of points at low distance is greater than in the Hausdorff case, a worthwhile piece of information for forthcoming considerations.

A second approach, investigated next for quantifying and graphically diagnosing how the candidate distances absolutely and comparatively perform, is based on cross-validated kNN predictions. For the four distances and a number of neighbours k ranging from 1 to 20, the following is implemented: for all $i \in 1, \ldots, N$ leave $\mathbf{x}_i$ out and predict it both by kNN and by taking the arithmetic average of the $N-1$ remaining responses. Finally, evaluate kNN's relative performance by taking either the ratio of the mean kNN error over the mean error when predicting by the mean, or the ratio of the median kNN error over the median error when predicting by the mean. Results are represented in Fig. 3. For the first criterion, symmetrizing the Hausdorff distance appears to lead to improved performances. In median, however, all distances perform similarly, and well compared to the prediction by the mean.

Distance-based Stratification We finally appeal to a clustering method, *Partitioning Around Medoids* (PAM), in order to select "representative" sub-samples of $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and approximate the cumulative distribution function of associated responses. This time, the letter k refers to the subsample size, which is in turn here the number of medoids in the PAM method. We rely on the Kolmogorov-Smirnov (K-S) statistic in order to assess and compare the performances of the different considered approaches. In addition to the PAM approach with the four candidate distances, we also consider uniform random sampling of k among N configurations. Values of the K-S statistic for the four candidate strategies versus 500 replications of the random strategy are plotted as a function of k in the left panel of Fig. 4. It

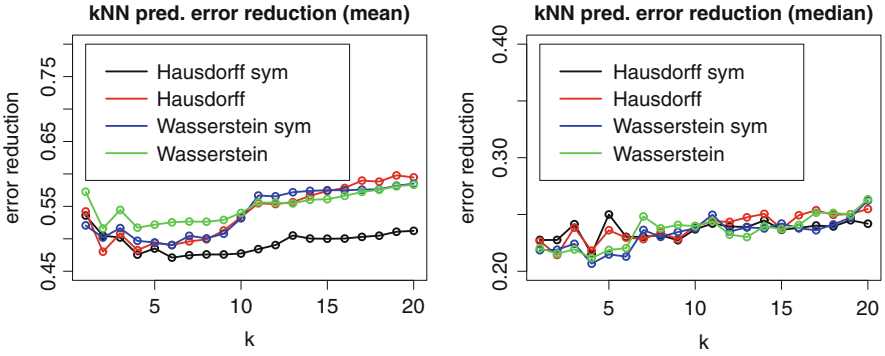

Fig. 3 Leave-one-out compared performances of kNN prediction versus naive prediction by the mean, both in terms of ratios of average errors and of median errors

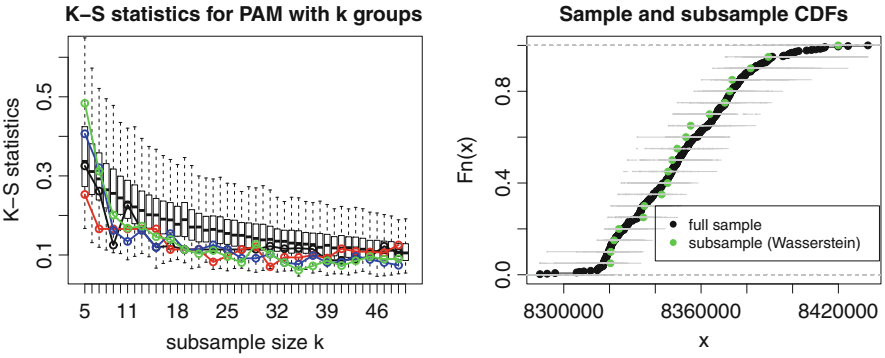

Fig. 4 *Left*: Performance assessment of random vs. stratified sampling using $d_H, d_{Hsym}, d_W, d_{Wsym}$ (*left*). *Right*: full sample CDF vs. subsample CDF obtained with the Wasserstein distance ($k = 20$). *Gray points* stand for quantiles obtained by drawing 50 subsamples uniformly at random

appears that for most values of k, the K-S statistic is smaller than the median K-S under the random scenario, whatever the candidate distance. Stratification with PAM hence appears as an appealing alternative to random sampling, as it avoids its variability. As an illustration, the empirical CDF obtained by PAM with the Wasserstein distance and $k = 20$ –approx. 5 % of the total budget– is represented (in green) against the empirical CDF of the full sample of responses. To finish with, a number of other approaches based on dissimilarity matrices may be beneficially applied to this problem; in particular, maximin and minimax distance designs [7] appear as credible alternatives for choosing representative sub-samples out of a reference sample.

5 Conclusion and Perspectives

We have investigated distance methods for the design of computer experiments with a view toward subsampling. A case study in mechanical engineering, where inputs are parametrized by point sets and four distances are in competition, has served as our basis. While non-Euclidean variography appeared appealing, results obtained on the test case did not shed much light on which distance was best suited for distance-based modelling. On the other hand, one of the two diagnostics based on kNN highlighted the potential interest of symmetrizing the Hausdorff distance. Nevertheless, all distances gave decent stratification results. While not suffering from the variability inherent in random subsampling, using the PAM algorithm with any of these distances and for most considered subsample sizes also delivered subsample distributions closer to the full sample distribution than obtained in median via random subsampling. However, these results have to be tempered for several reasons. First, all tests rely on a single data set with a specific structure, so that more studies with alternative data sets are necessary for validating and refining the conclusions by exploring their degree of generality. From the point of view of variogram design, it would be worthwhile to search for admissible models with respect to the Hausdorff and/or Wasserstein distances considered here (Gaussian and exponential models are not, according to numerical tests), but also to question the intrinsic stationarity assumption which was quietly made here. Besides this, the distances considered were almost off-the-shelf, and it would probably be beneficial to investigate further candidate distances for this or other test cases, taking available expertise into account [5]. Further perspectives include using this framework to adaptively refine the subsample, e.g., in order to foster the exploration of distributional tails.

Acknowledgements Part of this work has been conducted within the frame of the ReDice Consortium, which gathered industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Ecole des Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments.

References

1. Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Springer, New York (2007)
2. Cast3M software: http://www-cast3m.cea.fr. Accessed Dec 2015
3. Cressie, N.A.C.: Statistics for Spatial Data. Wiley, New York (1993)
4. Curriero, F.C.: On the use of non-Euclidean distance measures in geostatistics. Math. Geol. **38**(8), 907–926 (2006)
5. Ginsbourger, D., Rosspopoff, B., Pirot, G., Durrande, N., Renard, P.: Distance-based kriging relying on proxy simulations for inverse conditioning. Adv. Water Res. **52**, 275–291 (2013)
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2009)
7. Johnson, M.E., Moore, L.M., Ylvisaker, D.: Minimax and maximin distance designs. J. Stat. Plan. Inference **26**, 131–148 (1990)

8. Molchanov, I.: Theory of Random Sets. Springer, London (2005)
9. Radovanović, M., Nanopoulos, A., Ivanović, M.: Hubs in space: popular nearest neighbors in high-dimensional data. J. Mach. Learn. Res. **11**, 2487–2531 (2010)
10. Schuhmacher, D., Xia, A.: A new metric between distributions of point processes. Adv. Appl. Probab. **40**, 651–672 (2008)
11. von Luxburg, U.: Statistical learning with similarity and dissimilarity functions. Ph.D. thesis, Technische Universität Berlin (2004)

Optimal Design for the Rasch Poisson-Gamma Model

Ulrike Graßhoff, Heinz Holling, and Rainer Schwabe

Abstract Many tests, measuring human intelligence, yield count data. Often, these data can be analyzed by the Rasch Poisson counts model which incorporates parameters representing the ability of the respondents and the difficulty of the items. In a generalized version, the so-called Rasch Poisson-Gamma counts model, the ability parameter is specified as random with an underlying Gamma distribution. We will develop locally *D*-optimal calibration designs for an extended version of this model which includes two binary covariates in order to explain the difficulty of an item.

1 Introduction

Many intelligence tests allow for measuring memory, creativity and mental speed. The items of these tests consist of several stimuli. The number of correctly solved stimuli of such an item often follows a Poisson distribution. Especially for mental speed items this property has already been shown by Rasch [3] or later by e.g. Verhelst and Kamphuis [4]. Recently, Doebler and Holling [1] presented a rule-based mental speed test which stimulated the development of optimal designs to efficiently calibrate this test. Here, the Rasch Poisson-Gamma model, which incorporates random effects related to the test persons proved to be an adequate framework.

U. Graßhoff
School of Business and Economics, Humboldt University, Unter den Linden 6, D–10 099 Berlin, Germany
e-mail: grasshou@hu-berlin.de

H. Holling (✉)
Institute of Psychology, University of Münster, Fliednerstr. 21, D–48 149 Münster, Germany
e-mail: holling@uni-muenster.de

R. Schwabe
Institute for Mathematical Stochastics, Otto-von-Guericke University, PF 4120, D–39 016 Magdeburg, Germany
e-mail: rainer.schwabe@ovgu.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_16

Mental speed refers to the human ability to carry out mental processes, required for the solution of many cognitive tasks. Usually, mental speed is measured by elementary tasks with low cognitive demands in which speed of response is primary. These tasks represent items which may contain numerical, verbal or symbolic stimuli. Furthermore, these items can be differentiated by rules or conditions that determine the difficulty or easiness of an item.

The principle of a typical mental speed item will be explained by the following numerical item. This item consists of a set of 100 numbers between 1000 and 5000, e.g. 1245, 2246, 3673, 4345, ..., 2681. Here, the numbers represent stimuli. According to the instruction, a respondent has to mark those stimuli (numbers) that are divisible by 2 as fast as possible within a certain time, e.g. 20 s. Thus, this item is characterized by one rule, divisibility by 2. The stimuli are generated in such a way that only the surface of an item is changed but the difficulty is not affected or only by a negligible amount. Then stimuli are called incidentals. The score of such an item is represented by all correctly identified stimuli.

The difficulty of the item just outlined can be increased by adding further rules. From the following three rules, (a) the number has to be greater than 1621, (b) the number has to be less than 3456 or (c) the number has to contain two identical digits, a subset can be selected. When a subset of rules is added to the basic rule, i.e. divisibility by 2, a new item has been generated. Items defined in this way, i.e. by a set of rules and incidentals, establish a so-called item family. A lot of item families containing different numerical, verbal or symbolic stimuli have meanwhile been created for measuring mental speed (Doebler and Holling [1]).

Nowadays, such rule-based items are generated by computers on the fly, item banks are no longer necessary. Therefore, test security is increased as well as the economy of item generation. Furthermore, optimal design is an important means to alleviate the calibration of this kind of items. Therefore, the influence of the rules on the difficulty has to be determined.

2 The Poisson-Gamma Model for Count Data

In the Rasch Poisson counts model the number of solved tasks in a test is assumed to follow a Poisson distribution with mean (*intensity*) $\theta\sigma$, where the person parameter θ represents the ability of the respondent and the item parameter σ specifies the easiness of the test item. The ability and the easiness can be estimated in a two step procedure: first, in a calibration step, the item parameter is estimated conditional on known person parameters; and in a second step, the person parameter is estimated based on the item parameters previously obtained. In these two steps the items and the persons may be investigated separately. As in Graßhoff et al. [2] we will consider here the calibration step, but we will now allow the ability θ of a respondent to be random with known prior distribution. Usually in this context a Gamma distribution is assumed for the ability, which leads to the Poisson-Gamma model (see e.g. Verhelst and Kamphuis [4]).

To be more specific the conditional distribution of the number Y of correct answers given the ability $\Theta = \theta$ is Poisson with mean $\theta\sigma$, and the ability Θ is Gamma distributed with shape parameter $A > 0$ and inverse scale parameter $B > 0$.

Then the unconditional probabilities of Y can be obtained by integration of the joint density $\frac{(\theta\sigma)^y}{y!}e^{-\theta\sigma} \cdot \frac{B^A}{\Gamma(A)}\theta^{A-1}\exp(-B\theta)$ with respect to θ. It is well-known that the resulting distribution is (generalized) negative binomial with success probability $B/(\sigma + B)$ and (generalized) number of successes A which for integer A models the number of failures before the Ath success occurs The corresponding probability function is $P(Y = y) = \frac{\Gamma(y+A)}{y!\Gamma(A)}\left(\frac{B}{\sigma+B}\right)^A\left(\frac{\sigma}{\sigma+B}\right)^y$ with expectation $\mathrm{E}(Y) = \frac{A}{B}\sigma$ and variance $\mathrm{Var}(Y) = \frac{\sigma+B}{B}\mathrm{E}(Y)$. Thus the negative binomial is a common distribution to model overdispersion in count data. For the expectation $\frac{A}{B}\sigma$ fixed, the limiting distribution is again Poisson when B tends to infinity.

As in the Poisson count model, the easiness σ will be connected with the linear predictor based on the rules applied by the log link, $\sigma = \exp\{\mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta}\}$, where $\mathbf{x}$ is the experimental setting which may be chosen from a specific experimental region $\mathscr{X}$, $\mathbf{f} = (f_1, \ldots, f_p)^\top$ is a vector of known regression functions, and $\boldsymbol{\beta}$ the p-dimensional vector of unknown parameters to be estimated. Hence, the number of correct answers $Y(\mathbf{x})$ is Poisson-Gamma distributed with $E\{Y(\mathbf{x})\} = \frac{A}{B}\exp\{\mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta}\}$.

The items are generated in such a way that K different rules may be applied or not. Thus the explanatory variables x_k are binary, where $x_k = 1$, if the kth rule is applied, and $x_k = 0$ otherwise, $k = 1, \ldots, K$. The experimental settings $\mathbf{x} = (x_1, \ldots, x_K) \in \{0, 1\}^K$ constitute a binary K-way layout. In the particular case $\mathbf{x} = \mathbf{0}$ a basic item is presented. We specify the vector of regression functions by $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2, \ldots, x_K)^\top$. This means that there are no interactions of the form $x_k \cdot x_\ell$ or higher which would describe synergy or antagonist effects of the rules. For this model the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)^\top$ of dimension $p = K+1$ consists of a constant term β_0 and the K main effects β_k related to the application of the kth rule, respectively. In the present application it will be further assumed that $\beta_k \leq 0$ for the main effects because the application of a rule will typically increase the difficulty and, hence, decrease the easiness of an item.

Additionally we only consider the situation that each test person receives exactly one item in order to ensure independence of the observations.

3 Information and Design

The impact of an experimental setting on the quality of the maximum likelihood estimator of the parameter vector $\boldsymbol{\beta}$ is measured by the Fisher information matrix $\mathbf{M}(\mathbf{x}; \boldsymbol{\beta})$. Similarly to generalized linear models, the calculation of the Fisher information results in $\mathbf{M}(\mathbf{x}; \boldsymbol{\beta}) = \lambda(\mathbf{x}; \boldsymbol{\beta})\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top$, which depends on the particular

setting $\mathbf{x}$ and additionally on $\boldsymbol{\beta}$ through the intensity

$$\lambda(\mathbf{x};\boldsymbol{\beta}) = A\frac{\exp\{\mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta}\}}{\exp\{\mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta}\}+B}\,. \tag{1}$$

Consequently, for an exact design ξ consisting of N design points $\mathbf{x}_1,\ldots,\mathbf{x}_N$, the normalized information matrix equals $\mathbf{M}(\xi;\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i;\boldsymbol{\beta})$. For analytical ease we will use approximate designs ξ with mutually different design points $\mathbf{x}_1,\ldots,\mathbf{x}_n$, say, and corresponding (real valued) weights $w_i = \xi(\{\mathbf{x}_i\}) \geq 0$ with $\sum_{i=1}^{n} w_i = 1$. This approach is apparently appropriate, as typically the number N of items presented may be quite large. The information matrix is then more generally defined as

$$\mathbf{M}(\xi;\boldsymbol{\beta}) = \textstyle\sum_{i=1}^{n} w_i\lambda(\mathbf{x}_i;\boldsymbol{\beta})\mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)^\top. \tag{2}$$

The information matrix and, hence, optimal designs will depend on the parameter vector $\boldsymbol{\beta}$ through the intensity. For measuring the quality of a design we will use the popular D-criterion: a design ξ will be called locally D-optimal at $\boldsymbol{\beta}$ if it maximizes the determinant of the information matrix $\mathbf{M}(\xi;\boldsymbol{\beta})$.

We can factorize the information matrix $\mathbf{M}(\xi;\boldsymbol{\beta}) = A\mathbf{M}_0(\xi;\boldsymbol{\beta})$, where $\mathbf{M}_0(\xi;\boldsymbol{\beta})$ is the information in the standardized situation for which the original values of A, β_0 and B are replaced by $1, 0$ and $B\exp(-\beta_0)$, respectively. Hence, only $\det\{\mathbf{M}_0(\xi;\boldsymbol{\beta})\}$ has to be optimized, and we assume the standardized case ($A = 1$ and $\beta_0 = 0$) without loss of generality throughout the remainder of the paper.

4 Two Binary Predictors

To begin with we note that for the situation of only one rule ($K = 1$) there are only two possible settings $x = 1$ of application of the rule and $x = 0$ of the basic item. As there are two parameters in the model both settings have to be used and all such designs are saturated, i.e. the number of parameters and the number of settings coincide. For saturated designs it is well-known that the (locally) D-optimal design assigns equal weights ($w_i^* = 1/2$) to each of the settings whatever the value of $\boldsymbol{\beta}$ is because the information matrix factorizes into a diagonal matrix containing the weights and some square matrices related to the settings and the corresponding intensities.

When $K = 2$ rules are present, the four possible settings are $(1,1)$, where both rules are applied, $(1,0)$ and $(0,1)$, where either only the first or the second rule is used, respectively, and $(0,0)$ for the basic item. Hence, any design ξ is completely determined by the corresponding weights w_{11}, w_{10}, w_{01} and w_{00}, respectively. Further we denote by $\lambda_{x_1x_2} = \lambda\{(x_1,x_2);\boldsymbol{\beta}\}$ the related intensities for given parameter $B > 0$. It is worth-while noting that the information matrix (2) of any design depends on the parameters only through the intensities. Hence all

optimality results can be stated in terms of the intensities and can be transferred to any such model with binary predictors.

It looks natural to consider designs which use only the easiest settings $(0, 0)$, $(1, 0)$ and $(0, 1)$ in which at most one rule is applied. These designs are saturated, and among those the equireplicated design ξ_0 which assigns equal weights $1/3$ to the three settings is the best with respect to the D-criterion, independently of the value of $\boldsymbol{\beta}$. Based on the equivalence theorem Graßhoff et al. [2] derived a condition on the intensities under which the design ξ_0 is D-optimal for the pure Poisson model without a random effect. For non-positive values $\beta_k \leq 0$ this condition can be reformulated in terms of the inverse intensities as follows.

Theorem 1 *The design ξ_0 is locally D-optimal if and only if*

$$\lambda_{00}^{-1} + \lambda_{10}^{-1} + \lambda_{01}^{-1} \leq \lambda_{11}^{-1} . \tag{3}$$

The condition (3) is only based on the form of the information matrix (2) and the regression function $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2)^\top$ but does not depend on how the intensities are generated. Hence this condition carries over to the present situation of the Poisson-Gamma model and can be rephrased as

$$\beta_2 \leq \log\left[\frac{B\{1 - \exp(\beta_1)\}}{2\exp(\beta_1) + B\{1 + \exp(\beta_1)\}}\right] \tag{4}$$

in terms of the parameters. If the condition is not satisfied then items also have to be used in which both rules are applied. The parameter regions of β_1 and β_2, where the saturated design ξ_0 is locally D-optimal, are displayed in Fig. 1a for particular values of B. If the point (β_1, β_2) lies below the corresponding line, then ξ_0 is optimal. It is easy to see that the boundary condition (4) tends to the Poisson case, $\beta_2 \leq \log[\{1 - \exp(\beta_1)\}/\{1 + \exp(\beta_1)\}]$, when B tends to infinity.

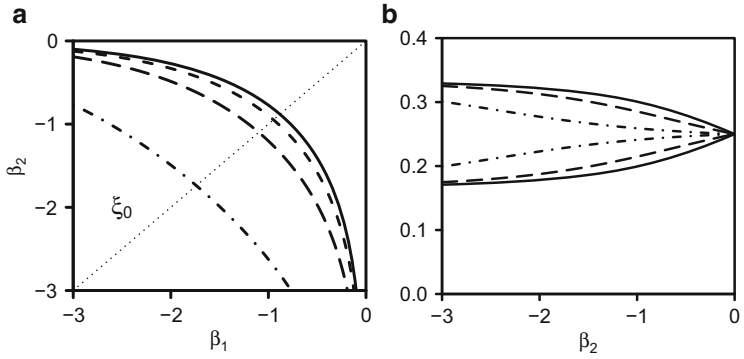

Fig. 1 (**a**): Dependence of regions for locally D-optimal designs on (β_1, β_2) for $B = 0.1$ (*dashes and dots*), $B = 1$ (*long dashes*), $B = 4$ (*short dashes*), and $B = \infty$ (Poisson; *solid line*); (**b**): optimal weights $w_{11}^* = w_{01}^* \leq 0.25$ and $w_{10}^* = w_{00}^* \geq 0.25$ for $B = 0.1$ (*dashes and dots*), $B = 1$ (*dashed line*) and $B = \infty$ (Poisson; *solid line*) in the case $\beta_1 = 0$

For illustrative purposes we will consider two particular parameter constellations, where optimal weights can be determined explicitly even when condition (3) is not satisfied. First we consider the situation of equal effect sizes, $\beta_1 = \beta_2 = \beta \leq 0$ indicated by the dotted line in Fig. 1a. Due to symmetry considerations with respect to interchanging the predictors, we can conclude that the optimal weights satisfy $w_{10}^* = w_{01}^*$. For given parameter $B > 0$, condition (3) yields that ξ_0 is optimal if $\beta \leq \beta_c = \log[\{\sqrt{2B(B+1)} - B\}/(B+2)]$. For $0 \geq \beta > \beta_c$ all four settings have to be presented in an optimal design. In that case the optimal weights can be calculated from

$$w_{10}^* = w_{01}^* = \frac{4\gamma + 2\sqrt{\gamma^2 + 12\rho_0\rho_1}}{3(4\rho_0\rho_1 - \gamma^2)}, \tag{5}$$

where $\rho_0 = \lambda_{10}/\lambda_{00}$ and $\rho_1 = \lambda_{10}/\lambda_{11}$ are intensity ratios, $\gamma = \rho_0 + \rho_1 - 4$,

$$w_{00}^* = \frac{1}{2} - w_{10}^* + \frac{1}{4}(\rho_1 - \rho_0)w_{10}^*$$
$$\text{and} \quad w_{11}^* = \frac{1}{2} - w_{10}^* - \frac{1}{4}(\rho_1 - \rho_0)w_{10}^*. \tag{6}$$

For the Poisson-Gamma model the dependence of the optimal weights on β is displayed in Fig. 2 for various values of B. It can be seen that the equally weighted four point design $\bar{\xi}$ (with $\bar{w}_{x_1,x_2} = 1/4$) is found to be optimal for the case of zero effects ($\beta = 0$).

Alternatively, when one of the effects vanishes, say $\beta_1 = 0$, which corresponds to the vertical axis in Fig. 1a, the intensities are constant in the first component. Then, because of symmetry, the optimal weights are also constant in x_1 and can be calculated to be

$$w_{11}^* = w_{01}^* = \frac{2\lambda_{00} - \lambda_{11} - \sqrt{\lambda_{00}^2 - \lambda_{00}\lambda_{11} + \lambda_{11}^2}}{6(\lambda_{00} - \lambda_{11})} \tag{7}$$

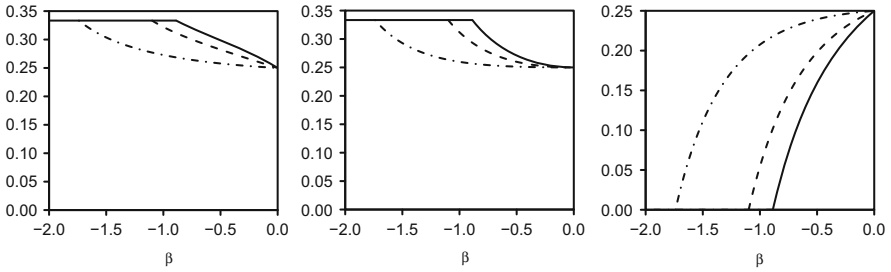

Fig. 2 Optimal weights w_{00}^* (*left panel*), $w_{10}^* = w_{01}^*$ (*middle panel*) and w_{11}^* (*right panel*) for $B = 0.1$ (*dashes and dots*), $B = 1$ (*dashed line*) and $B = \infty$ (Poisson; *solid line*) in the case $\beta_1 = \beta_2 = \beta$

and $w_{10}^* = w_{00}^* = 1/2 - w_{11}^*$. For the Poisson-Gamma model Fig. 1b exhibits these weights as functions of β_2 for various values of B. The weights $w_{11}^* = w_{01}^*$ decrease, when these items become more difficult, and approach $1/6$ for $\beta_2 \to -\infty$. Hence, more observations will be allocated to the items with lower difficulty.

5 Efficiency

Locally D-optimal designs may show poor performance, if erroneous initial values are specified for the parameters. As in Graßhoff et al. [2], we will perform a sensitivity analysis for the Poisson-Gamma model which is additionally influenced by the value of B.

We consider the parameter constellations of the preceding section for which we can determine the optimal weights explicitly. For the case of equal effect sizes ($\beta_1 = \beta_2 = \beta$) we display the D-efficiency of the saturated design ξ_0 in the left panel and the efficiency of the equally weighted four point design $\bar{\xi}$ that is optimal for $\beta_1 = \beta_2 = 0$ in the right panel of Fig. 3 for various values of B. If $\beta \leq \log[\{\sqrt{2B(B+1)} - B\}/(B+2)]$, the design ξ_0 is locally D-optimal and has, hence, efficiency 100 %. When β increases beyond this critical value, the efficiency of ξ_0 decreases to 83 % for $\beta = 0$ for all B. Moreover, for ξ_0 the efficiency decreases when B gets smaller. For $\bar{\xi}$ the efficiency obviously equals 100 % for $\beta = 0$ and decreases to 75 % when β tends to $-\infty$. Here the efficiency increases when B gets smaller which seems to be reasonable because for small values of B the intensities are nearly constant as in the corresponding linear model, where $\bar{\xi}$ is D-optimal.

Similarly, for the case $\beta_1 = 0$ we display the D-efficiency of the saturated design ξ_0 in the left panel and the efficiency of the equally weighted four point design $\bar{\xi}$ in the right panel of Fig. 4 for various values of B. For ξ_0 the efficiency again attains the minimal value of 83 % for the situation of no effects ($\beta_2 = 0$) and approaches 100 % when $\beta_2 \to -\infty$ for all B. For $\bar{\xi}$ the efficiency goes from 100 % for $\beta_2 = 0$

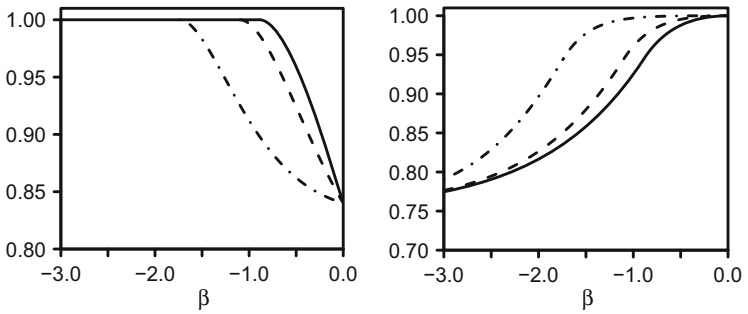

Fig. 3 Efficiency of ξ_0 (*left panel*) and $\bar{\xi}$ (*right panel*) for $B = 0.1$ (*dashes and dots*), $B = 1$ (*dashed line*) and $B = \infty$ (Poisson; *solid line*) in the case $\beta_1 = \beta_2 = \beta$

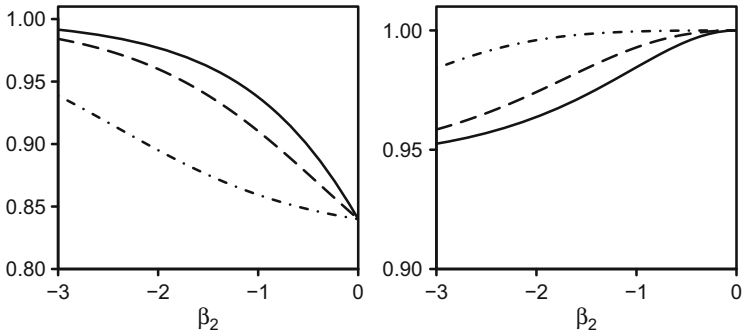

Fig. 4 Efficiencies of ξ_0 (*left panel*) and $\bar{\xi}$ (*right panel*) for $B = 0.1$ (*dashes and dots*), $B = 1$ (*dashed line*) and $B = \infty$ (Poisson; *solid line*) in the case $\beta_1 = 0$

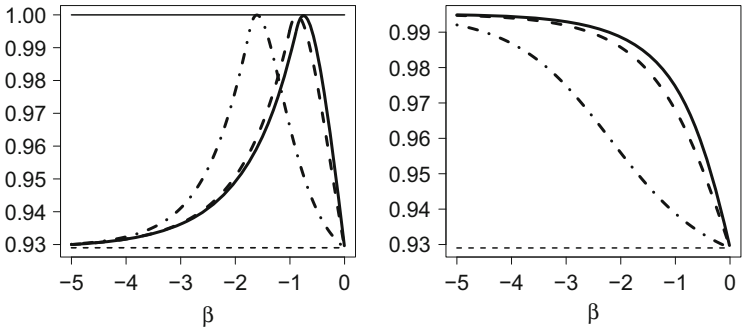

Fig. 5 Efficiency of ξ_m^* for $B = 0.1$ (*dashes and dots*), $B = 2$ (*dashed line*) and $B = \infty$ (Poisson; *solid line*) in the case $\beta_1 = \beta_2 = \beta$ (*left panel*) and $\beta_1 = 0$ (*right panel*)

to 94 % when $\beta \to -\infty$ for all B. It can again be noticed that for fixed β_2 the efficiency of ξ_0 decreases and of $\bar{\xi}$ increases when B gets smaller.

From these figures we may conclude that in the case of equal effects $\beta_1 = \beta_2 = \beta$ the minimal efficiency is 83 % for ξ_0 and 75 % for $\bar{\xi}$, where the minimum is attained at $\beta = 0$ for ξ_0 and at $\beta \to -\infty$ for $\bar{\xi}$, respectively.

The maximin efficient design ξ_m^* can be obtained by easy but tedious computations. It turns out that ξ_m^* is the same for all B and allocates equal weights $w_{00}^* = w_{10}^* = w_{01}^* = 0.31$ to the settings of the saturated design ξ_0 and the remaining weight $w_{11}^* = 0.07$ to the setting in which both rules are applied. More precisely, the maximin efficient design $\xi_m^* = (111\,\xi_0 + 44\,\bar{\xi})/155$ is a mixture of the two limiting locally optimal designs ξ_0 and $\bar{\xi}$, where the coefficients do not depend on B. Moreover, ξ_m^* is also maximin efficient in the Poisson model. As can be seen from Fig. 5 (left panel) the minimal efficiency of ξ_m^* equals 93 % for all B and is attained both for $\beta = 0$ and $\beta \to -\infty$. Additionally it appears that ξ_m^* is locally optimal for some β which varies with B. Moreover, as shown in the right panel of Fig. 5 the

efficiency of ξ_m^* is also at least 93 % when one of the effects vanishes. Hence, we conjecture that ξ_m^* is maximin efficient for all $\beta_1, \beta_2 \leq 0$.

6 Discussion

As pointed out, the performance of a design is completely determined in a model with binary predictors by the intensities for the single settings. Thus optimal designs can be obtained, in general, in terms of these intensities computed from the parameters using the model equations or, vice versa, conditions on the intensities can be transferred back to these parameters.

The designs outlined above can be used for calibrating mental speed items incorporating two rules. But, item families often contain more than two rules. Thus, extensions to higher dimensions are under consideration and require, in general, higher order conditions.

Acknowledgements This work was supported by grant Ho1286-6 of the Deutsche Forschungsgemeinschaft.

References

1. Doebler, A., Holling, H.: A processing speed test based on rule-based item generation: an analysis with the Rasch Poisson Counts model. Learn. Individ. Differ. (2015, in press)
2. Graßhoff, U., Holling, H., Schwabe, R.: Optimal design for count data with binary predictors in item response theory. In: Uciński, D., Atkinson, A.C., Patan, M. (eds.) mODa10 -Advances in Model-Oriented Design and Analysis, pp. 117–124. Springer, Cham (2013)
3. Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research, Copenhagen (1960)
4. Verhelst, N.D., Kamphuis, F.H.: A Poisson-Gamma model for speed tests. Measurement and Research Department Reports, 2009–2. Cito, Arnhem (2009)

Regular Fractions of Factorial Arrays

Ulrike Grömping and R.A. Bailey

Abstract For symmetric arrays of two-level factors, a regular fraction is a well-defined concept, which has been generalized in various ways to arrays of s-level factors with s a prime or prime power, and also to mixed-level arrays with arbitrary numbers of factor levels. This paper introduces three further related definitions of a regular fraction for a general array, based on squared canonical correlations or the commuting of projectors. All classical regularity definitions imply regularity under the new definitions, which also permit further arrays to be considered regular. As a particularly natural example, non-cyclic Latin squares, which are not regular under several classical regularity definitions, are regular fractions under the proposed definitions. This and further examples illustrate the different regularity concepts.

1 Introduction

An array is an $N \times n$ (N rows, n columns) table of symbols, for which the ith column contains s_i symbols. The columns are also called factors, the symbols are also called levels. If $s_1 = \cdots = s_n$, the array is called symmetric, otherwise mixed-level or asymmetric. A full factorial would have (a multiple of) $s_1 \times \cdots \times s_n$ rows; we consider a fraction with N rows. In a balanced array, each column contains each of its symbols equally often. If in addition each pair of columns contains each of its pairs of symbols equally often, the array is an orthogonal array (OA). We only consider balanced arrays, and in most cases OAs. Now we consider three established regularity definitions in more detail:

Cyclic group regularity or Abelian group regularity refers to arrays for which the s_i levels of the ith factor can be given in terms of the cyclic group $\mathbb{Z}/s_i\mathbb{Z}$ or,

U. Grömping (✉)
Beuth University of Applied Sciences Berlin, Dep. II, Luxemburger Str. 10, 13353 Berlin, Germany
e-mail: groemping@beuth-hochschule.de

R.A. Bailey
School of Mathematics & Statistics, University of St Andrews, St Andrews KY16 9SS, UK
e-mail: rab24@st-andrews.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_17

more generally, an Abelian group of order s_i, such that the fraction is a coset of a subgroup of the direct product of all these groups (i.e., of the full factorial).

Pseudo-factor regularity refers to arrays created using defining relations among prime-level pseudo-factors: any factor with non-prime number of levels is viewed as a full factorial in so-called pseudo-factors whose numbers of levels are primes. These arrays can also be characterized by the existence of a coding based on pseudo-factors such that the rows form a coset of a subgroup of an Abelian group, i.e., they are a special case of Abelian group regular arrays. Kobilinsky et al. [12] recently described algorithms for creating such arrays; these are implemented in the R package **planor** [11].

GF regularity refers to symmetric arrays with all factors at q levels, q a prime power, that can be created through defining equations for fractionating, also called generators, based on additive contrasts in the Galois field $GF(q)$. If q is prime, addition in $GF(q)$ coincides with addition modulo q, while it is different otherwise, see e.g. [5]. Equivalently, GF regularity can be characterized by labeling the q levels of each factor such that the rows form an affine subspace of the n-dimensional affine space over GF(q). This is the concept discussed by Bose [3] and implies all other types of regularity.

For symmetric 2-level or 3-level arrays, all these regularity types are equivalent. With more than three levels, they need no longer coincide, not even for symmetric arrays with number of levels a prime power. For example, the 4-level array, arranged as a Latin square in Table 1, is Abelian group regular but not GF or pseudo-factor regular. Latin squares also provide examples for which it appears anti-intuitive to consider them non-regular, but which are not regular according to the classic criteria: the 5-level array in Table 1 (the first array from Eendebak and Schoen [6]) is neither GF regular nor pseudo-factor regular nor Abelian group regular (but will be considered regular according to our proposed criteria).

Regularity is closely related to orthogonality. OAs are a widely known orthogonal structure; an even weaker one was introduced by Tjur [15] and termed "Tjur block structure" by Bailey [1]. Tjur block structures consist of factors orthogonal in the sense that projectors onto the corresponding subspaces commute, with the trivial 1-level factor and the supremum of any pair of factors always included (see Table 4 for an example of the supremum). All OAs are Tjur block structures, but the reverse

Table 1 Two Latin squares: a cyclic one for 4-level factors and a non-cyclic one for 5-level factors; entries show the level of factor C for each AB combination

4-level factors		B				5-level factors		B				
		0	**1**	**2**	**3**			**0**	**1**	**2**	**3**	**4**
A	**0**	0	1	2	3	A	**0**	0	1	2	3	4
	1	1	2	3	0		**1**	1	0	3	4	2
	2	2	3	0	1		**2**	2	3	4	0	1
	3	3	0	1	2		**3**	3	4	1	2	0
							4	4	2	0	1	3

is not true (see e.g. Table 4). If the supremum is not included, it can be added without destroying orthogonality, so that closure under suprema is a lesser issue, and we will call a structure that requires addition of suprema a *weak Tjur block structure*. All examples in this paper are at least weak Tjur block structures.

We propose a regularity definition based on commuting projectors, called "geometric regularity". The recent work by Grömping and Xu [9] offers the possibility for two further regularity definitions, based on model matrices but nevertheless coding-invariant. One definition, called "CC-regularity", is based on the individual so-called squared canonical correlations (SCCs) for the main effect of each factor, whereas a stricter definition ("R^2 regularity") requires that the average R^2 value for explaining each degree of freedom for a main effect be either 0 or 1. The building blocks of CC-regularity and R^2 regularity are introduced in the next section, before the three new regularity definitions and their interrelations are presented in Sect. 3. Section 4 summarizes the relations among the criteria and discusses practical issues regarding their assessment.

2 Generalized Word Length Patterns, SCCs, and R^2 Values

Definition 1 defines notation regarding model matrices and projectors. Definition 2 defines concepts in connection with the generalized word length pattern (GWLP), which was introduced by Xu and Wu [16] and generalizes the well-known word length pattern (WLP) for symmetric 2-level arrays.

Definition 1 Consider an $N \times n$ array with the ith column at s_i levels for $i = 1, \ldots, n$.

(i) $\mathbf{1}_N$ denotes a column of ones.
(ii) For $j = 0, \ldots, n$, a *j-factor set* is a subset of j columns of the array, and $\mathscr{S}_j = \{S \subseteq \{1, \ldots, n\} : |S| = j\}$ denotes the set of all j-factor sets.
(iii) For $S \in \mathscr{S}_j$, $\mathbf{X}_S$ denotes the $N \times c_{\text{full}}$ model matrix of a full model with all main effects and interactions up to degree j, where $c_{\text{full}} = \prod_{i \in S} s_i$; $\mathbf{X}_S$ consists of suitably assembled rows of the matrix that would be used in the full factorial. The first column of $\mathbf{X}_S$ is assumed to be $\mathbf{1}_N$, and $\mathbf{X}_{\{\}} = \mathbf{1}_N$.
(iv) $\mathbf{P}_S = \mathbf{X}_S(\mathbf{X}_S'\mathbf{X}_S)^{-}\mathbf{X}_S'$ is the orthogonal projector onto the column span of $\mathbf{X}_S$.
(v) For factor i, $\mathbf{X}_{\{i\}}$ denotes the $N \times s_i$ model matrix including $\mathbf{1}_N$, and $\mathbf{X}_i$ denotes the $N \times (s_i - 1)$ sub-matrix without $\mathbf{1}_N$. The columns of $\mathbf{X}_i$ are centered.
(vi) Factor i is said to be in *normalized orthogonal coding* if $\mathbf{X}_i'\mathbf{X}_i = N\mathbf{I}_{s_i-1}$.
(vii) The full model matrix $\mathbf{X}_S$ is said to be in *normalized orthogonal coding*, if all individual factors are in normalized orthogonal coding and interaction columns are constructed as products of main effects columns.
(viii) The matrix of the $\prod_{i \in S}(s_i - 1)$ highest order interaction columns from $\mathbf{X}_S$ in normalized orthogonal coding is called $\mathbf{X}_{\mathscr{I}(S)}$. Thus $\mathbf{X}_{\mathscr{I}(\{i\})} = \mathbf{X}_i$ in normalized orthogonal coding, and $\mathbf{X}_{\mathscr{I}(\{\})} = \mathbf{1}_N$.

Definition 2 Consider an $N \times n$ array with the ith column at s_i levels for $i = 1, \ldots, n$.

(i) For $S \in \mathscr{S}_j$, $a_j(S) = (\mathbf{1}_N' \mathbf{X}_{\mathscr{I}(S)})(\mathbf{1}_N' \mathbf{X}_{\mathscr{I}(S)})'/N2$ denotes the *projected a_j value*.
(ii) The *Generalized Word Length Pattern* (GWLP) $(A_0, A_1, A_2, \ldots, A_n)$ of the array is defined by $A_j = \sum_{S \in \mathscr{S}_j} a_j(S)$ for $j = 0, \ldots, n$.
(iii) The *resolution* R of the array is the smallest j such that $j > 0$ and $A_j > 0$.
(iv) A j-factor subset with resolution j is called a *full resolution subset*.

All OAs have resolution at least 3. Therefore, the GWLP for OAs is usually specified starting with A_3. For resolution 3 arrays, main effects can be estimated orthogonally to each other but may be confounded with interactions of two or more factors. For resolution $R > 3$, main effects are also orthogonal to interactions among up to $R-2$ factors. As one usually assumes lower order effects likely to be stronger than higher order effects, the entry A_R of the GWLP is the most important one. As it is the sum of the projected a_R values $a_R(S)$ over all R-factor sets S, these deserve attention. Grömping and Xu [9] provided two statistical interpretations for these, which are given below. Note that R without square always denotes the resolution, while R^2 denotes the coefficient of determination.

Lemma 1 *Let $S \in \mathscr{S}_R$. For a particular $i \in S$, $a_R(S)$ is the sum of*

(i) the $s_i - 1$ SCCs between the main effects model matrix $\mathbf{X}_i$ and the full model matrix $\mathbf{X}_{S-\{i\}}$;
(ii) the $s_i - 1$ R^2 values from regressing the $s_i - 1$ columns of $\mathbf{X}_i$ on $\mathbf{X}_{S-\{i\}}$, where $\mathbf{X}_i$ is in normalized orthogonal coding.

While the sums in Lemma 1 are identical, no matter which factor i in S is chosen, the individual summands can be different (see Table 2, where the sum is 1 for all factors). The individual R^2 values in (ii) also depend on the choice of normalized orthogonal coding and are therefore not further considered. Note that the SCCs are closely related to the *canonical efficiency factors* introduced by James and Wilkinson [10]; the latter are used in the literature on incomplete block designs and are implemented in the R package **dae** by Brien [4].

According to [9], a factor i in a resolution R array is fully confounded within an R-factor set S, if the respective $a_R(S)$ is equal to $s_i - 1$. This is the case exactly if all SCCs are 1 for this factor singled out as the main effects factor, or, equivalently, if all R^2 values from explaining the columns of $\mathbf{X}_i$ through $\mathbf{X}_{S-\{i\}}$ are 1. On the

Table 2 Non-regular 16×3 resolution 3 array from Eendebak and Schoen [6] (transposed)

	The transposed array																Squared canonical correlations		
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		for A and B	for C
A	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3	First	0.5	1
B	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	Second	0.5	
C	0	0	1	1	0	1	0	1	1	0	1	0	1	1	0	0	Third	0	

other hand, if all SCCs are zero, or, equivalently, if all R^2 values from explaining the columns of $\mathbf{X}_i$ through $\mathbf{X}_{S-\{i\}}$ are 0, the set S does not contribute anything to A_R.

Consider factor i in a j-factor set S. If all SCCs between $\mathbf{X}_i$ and $\mathbf{X}_{S-\{i\}}$ are in $\{0, 1\}$ then there is a coding for factor i such that each column of $\mathbf{X}_i$ is either fully explained by or uncorrelated with $\mathbf{X}_{S-\{i\}}$. Contrary to most results stated in [9], this insight applies to any j-factor set S regardless of its resolution. This makes the idea of using the SCCs for the assessment of array regularity worth pursuing. Also, the average R^2 value from regressing the columns of $\mathbf{X}_i$ (regardless of coding) onto those of $\mathbf{X}_{S-\{i\}}$ is 1 or 0 exactly if the entire factor is fully explained or not explained at all by the factors in $S - \{i\}$, i.e. this is a stricter variant of no partial confounding. These two versions of regularity are now formally defined, along with geometric regularity.

3 The New Regularity Definitions

Definition 3 A balanced $N \times n$ array is *CC regular*, if the following holds: for every subset S of at least two of the n factors, for every $i \in S$, all SCCs between $\mathbf{X}_i$ and $\mathbf{X}_{S-\{i\}}$ are 0 or 1 only.

Definition 4 A balanced $N \times n$ array is R^2 *regular*, if the following holds: for every subset S of at least two of the n factors, for every $i \in S$, the R^2 values obtained from regressing the columns of $\mathbf{X}_i$ on $\mathbf{X}_{S-\{i\}}$ are all 0 or all 1.

The expressions "CC regular" and "R^2 regular" have been inspired by their correspondence to 0/1 only SCC frequency tables and average R^2 frequency tables, as proposed in [7].

Theorem 1 *The following relations hold.*

(i) R^2 *regularity implies CC regularity.*
(ii) For symmetric 2-level arrays, R^2 *regularity and CC regularity are equivalent.*

Proof Consider a set $S \subseteq \{1, \ldots, n\}$, and a factor $i \in S$.

(i) R^2 regularity implies that the span of $\mathbf{X}_i$ is either contained in the span of $\mathbf{X}_{S-\{i\}}$ (which implies that all SCCs are 1), or is orthogonal to that space (which implies that all SCCs are 0).
(ii) The single R^2 value equals the single SCC for each pair $\mathbf{X}_{\{i\}}$ and $\mathbf{X}_{S-\{i\}}$. □

Definition 4 is equivalent to each factor's *average* R^2 value being 0 or 1, respectively. Note that R^2 regularity is possible only for symmetric arrays, as the sum of SCCs is the same for all factors in a set and coincides with the sum of R^2 values in the case of full resolution; this restriction does not hold for trivial cases, i.e. for mixed-level arrays that are obtained by crossing or nesting R^2 regular symmetric arrays with different numbers of levels. Furthermore, note that CC regularity and R^2 regularity request the 0/1 property for *all* sets of factors,

Table 3 24 × 13 array of resolution 3 from Kuhfeld [13] (transposed)

Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
B	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	2	2	2
C	1	1	1	2	2	2	1	1	1	2	2	2	1	1	1	2	2	2	1	1	1	2	2	2
D	1	2	2	1	1	2	1	1	2	1	2	2	1	1	2	1	2	2	1	2	2	1	1	2
E	2	1	1	1	2	2	1	2	2	1	1	2	1	2	2	1	1	2	1	1	2	2	2	1
F	1	1	2	2	2	1	2	1	2	1	1	2	1	2	2	1	2	1	2	1	1	1	2	2
G	1	2	2	1	2	1	1	2	1	2	1	2	1	2	1	2	1	2	2	1	2	1	1	2
H	2	1	2	2	1	1	1	2	1	1	2	2	1	1	2	2	1	2	2	2	1	1	2	1
J	1	2	1	2	1	2	2	2	1	1	1	2	1	2	2	2	1	1	1	2	1	2	1	2
K	2	2	1	2	1	1	1	1	2	2	1	2	1	2	1	1	2	2	2	2	1	2	1	1
L	2	2	1	1	2	1	2	1	1	1	2	2	1	1	2	2	2	1	2	1	2	2	1	1
M	2	1	2	1	1	2	2	1	1	2	1	2	1	2	1	2	2	1	1	2	2	1	2	1
N	7	6	3	5	2	11	8	4	12	9	10	1	1	10	9	12	4	8	11	2	5	3	6	7

Squared canonical correlations

for 3 factor sets		for 4 factor sets of resolution 4	
SCC	Frequency	SCC	Frequency
0	1320	1/9	1980
1	198		

which is much more than was investigated for generalized resolution (see [9]), where considerations for resolution R arrays were restricted to R-factor sets. Table 3 gives an example of a resolution 3 array that has regular 3-factor sets only, but contains partial confounding for 4-factor sets of resolution 4. Thus, it is not sufficient to restrict attention to R-factor sets when assessing CC regularity. We conjecture that it is sufficient to check all full resolution sets. For R^2 regularity, the proof is straightforward: for a full resolution set S and $i \in S$, all SCCs between $\mathbf{X}_i$ and $\mathbf{X}_{S-\{i\}}$ are 1; moreover, for $T \supset S$, all SCCs between $\mathbf{X}_i$ and $\mathbf{X}_{T-\{i\}}$ are 1; for $k \in T - S$, either all SCCs between $\mathbf{X}_k$ and $\mathbf{X}_{T-\{k\}}$ are 0, or k is part of another full resolution set which is independently assessed. For CC regularity, we have so far not found a proof for the conjecture.

We now define *geometric regularity*, using the fact that two projectors commute if and only if their column spans are geometrically orthogonal in the sense of [15]. It is a more lenient variant of the orthogonal block structures introduced by Bailey [1].

Definition 5 A balanced $N \times n$ array is *geometrically regular*, if the following holds: for any two subsets S and T of the n factors, $\mathbf{P}_S$ and $\mathbf{P}_T$ commute.

$S \subseteq T$ implies $\mathbf{P}_S\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_S = \mathbf{P}_S$, i.e., such pairs of sets need not be checked. The following theorem, which follows from equivalence of 0/1 canonical correlations to projector commuting, shows the close relationship between CC regularity and geometric regularity.

Theorem 2 *Let* $i \in \{1, \ldots, n\}$, *and* $\emptyset \neq S, T \subseteq \{1, \ldots, n\}$. *A balanced* $N \times n$ *array*

(i) is CC regular *iff* $\mathbf{P}_{\{i\}}$ *and* $\mathbf{P}_S$ *commute for all pairs i and S with* $i \notin S$;
(ii) is R^2 regular *iff* $\mathbf{P}_{\{i\}}\mathbf{P}_S \in \{\mathbf{0}, \mathbf{P}_{\{i\}}\}$ *for all pairs i and S with* $i \notin S$;
(iii) is geometrically regular *iff all SCCs between* $\mathbf{X}_S$ *and* $\mathbf{X}_T$ *are in* $\{0, 1\}$ *for all pairs S and T*.

Proof All parts of the theorem follow from two linear algebra results: projectors commute iff the eigenvalues of their product are all in $\{0, 1\}$ (follows from Fact 3.9.16 in [2] and Prop. 18.11 in [14]), and the non-zero SCCs of two column-centered matrices $\mathbf{M}$ and $\mathbf{N}$ coincide with the non-zero eigenvalues of the product of the corresponding orthogonal projectors $\mathbf{P_M P_N}$ (also in Prop. 18.11 in [14]). □

Theorem 2 shows that geometric regularity implies CC regularity. The reverse is not true: geometric regularity is stricter than CC regularity in that it requires orthogonality (in the geometric or 0/1 canonical correlation sense) between all effects, not only between main effects and other effects. For example, a CC regular (and even R^2 regular) but not geometrically regular array can be constructed as a Latin cube from the 5-level Latin square of Table 1 by crossing a 5-level height factor H (levels 0–4) with the Latin square and modifying the original Latin square factor C by taking its sum with H modulo 5; for this Latin cube, the projectors $\mathbf{P}_{\{H,A\}}$ and $\mathbf{P}_{\{B,C\}}$ do not commute, even though it is CC regular and R^2 regular. For symmetric 2-level arrays, however, geometric regularity is equivalent to the other regularity definitions.

The next example illustrates the generality of the proposed regularity concepts. Table 4 shows an array that has resolution 2 only (no OA) and is a weak Tjur block structure; the main effects of the two 4-level factors have one completely confounded df each (0/1 vs 2/3 for A is confounded with 0/1 vs 2/3 for B); these give rise to the two SCCs of 1. Thus, this array is CC regular and pseudo-factor regular. As the overlap between A and B main effects splits into a parallel portion and a portion orthogonal to it, projectors onto the spaces spanned by A and B commute. All other projectors commute as well so that the array is geometrically regular. It is also Abelian group regular, but not R^2 regular or GF regular.

Table 4 CC regular and geometrically regular resolution 2 array

The transposed array									Squared canonical correlations	
Run	1	2	3	4	5	6	7	8	SCC	Frequency
A	0	0	1	1	2	2	3	3	0	12
B	0	1	0	1	2	3	2	3	1	2
C	0	1	1	0	0	1	1	0		
The supremum of factors A and B										
H	0	0	0	0	1	1	1	1		

4 Final Remarks

We have extended GF regularity, pseudo-factor regularity and cyclic/Abelian group regularity with three new regularity definitions. The following hierarchy holds: GF regularity $\Rightarrow$ pseudo-factor regularity $\Rightarrow$ Abelian group regularity $\Rightarrow$ geometric regularity$\Rightarrow$ CC regularity. Furthermore, GF regularity $\Rightarrow$ R^2 regularity $\Rightarrow$ CC regularity. The reverse is not true: this is obvious for GF regularity and pseudo-factor regularity, or R^2 regularity and CC regularity, and follows from the examples discussed throughout the paper for the other implications. The new proposals thus provide weaker regularity requirements than the established ones. Their weakness is welcome, because it allows us to consider the 5-level Latin square of Table 1 as a regular array. Whether or not one also wants to consider as regular the Latin cube that can be constructed from it yields the distinction between CC or R^2 regularity on the one hand, and geometric regularity on the other hand.

All three new regularity definitions can be checked post-hoc without knowledge of the array construction principle. However, even with a moderate number of factors, the effort can be tremendous; nevertheless, checking remains more manageable than trying to verify construction-based regularity for unknown construction and level labeling. We mentioned that R^2 regularity can be established by considering only full resolution sets, and conjectured the same to hold for CC regularity. For feasibility reasons, the R package **DoE.base** [8] includes a check for R^2 or CC regularity based on full resolution sets only. Should the conjecture prove wrong, this check only provides for CC regularity within all lowest order factor sets. Checks for geometric regularity have so far not been implemented. Work is in progress on technical conditions, in addition to excluding pairs of subsets related by inclusion, that would reduce the computational burden of checking all pairs of subsets.

Acknowledgements Ulrike Grömping's initial work was supported by Deutsche Forschungsgemeinschaft (Grant GR 3843/1-1). Ulrike Grömping wishes to thank Hongquan Xu for fruitful discussions on an earlier version of this work. The collaboration with Rosemary Bailey was initiated at a workshop funded by Collaborative Research Center 823 at TU Dortmund University.

References

1. Bailey, R.A.: Orthogonal partitions in designed experiments. Des. Codes Cryptogr. **8**, 54–77 (1996)
2. Bernstein, D.S.: Matrix Mathematics. Princeton University Press, Princeton (2005)
3. Bose, R.C.: Mathematical theory of the symmetrical factorial design. Sankhya **8**, 107–166 (1947)
4. Brien, C.: **dae**: functions useful in the design and ANOVA of experiments. R package version 2.7-2. http://CRAN.R-project.org/package=dae (2015)
5. Dey, A., Mukerjee, R.: Fractional Factorial Plans. Wiley, New York (1999)
6. Eendebak, P., Schoen, E.: Complete series of non-isomorphic orthogonal arrays. http://pietereendebak.nl/oapage/ (2010)

7. Grömping, U.: Frequency tables for the coding invariant ranking of orthogonal arrays. http://www1.beuth-hochschule.de/FB_II/reports/ (2013)
8. Grömping, U.: **DoE.base**: Full factorials, orthogonal arrays and base utilities for DoE packages. R package version 0.27-1, http://CRAN.R-project.org/package=DoE.base (2015)
9. Grömping, U., Xu, H.: Generalized resolution for orthogonal arrays. Ann. Stat. **42**, 918–939 (2014)
10. James, A.T., Wilkinson, G.N.: Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. Biometrika **58**, 279–294 (1971)
11. Kobilinsky, A., Bouvier, A., Monod, H.: **planor**: Generation of regular factorial designs. R package version 0.2-3. http://CRAN.R-project.org/package=planor (2015)
12. Kobilinsky, A., Monod, H., Bailey, R.A.: Automatic generation of generalised regular factorial designs. Preprint Series 7, Isaac Newton Institute, Cambridge (2015)
13. Kuhfeld, W.: Orthogonal arrays. http://support.sas.com/techsup/technote/ts723.html (2010)
14. Puntanen, S., Styan, G.P.H., Isotalo, J.: Matrix Tricks for Linear Statistical Models. Springer, Heidelberg (2011)
15. Tjur, T.: Analysis of variance models in orthogonal designs. Int. Stat. Rev. **52**, 33–81 (1984)
16. Xu, H., Wu, C.F.J.: Generalized minimum aberration for asymmetrical fractional factorial designs. Ann. Stat. **29**, 1066–1077 (2001)

Likelihood-Free Extensions for Bayesian Sequentially Designed Experiments

Markus Hainy, Christopher C. Drovandi, and James M. McGree

Abstract When considering a Bayesian sequential design framework, sequential Monte Carlo (SMC) algorithms are a natural approach. However, these algorithms require the likelihood function to be evaluated. Therefore, they cannot be applied in cases where the likelihood is not available or is intractable. To overcome this limitation, we propose likelihood-free extensions of the standard SMC algorithm. We also investigate a specific simulation-based approximation of the likelihood known as the synthetic likelihood. The algorithms are applied and tested on a well-studied sequential design problem for estimating a non-linear function of linear regression parameters.

1 Introduction

Conducting statistical experiments sequentially allows the use of information provided by previously obtained data to improve the efficiency of future data collection. When considering a Bayesian design framework as in [2], sequential Monte Carlo (SMC) algorithms are a natural approach. However, these algorithms require evaluation of the likelihood function. Therefore, they cannot be applied in cases where the likelihood is not available or is intractable.

To overcome this limitation, we propose likelihood-free extensions of the standard SMC algorithm. These methods are simulation-intensive, and require large amounts of simulated data to be generated from the model of interest in a reasonable amount of time; see [7]. We also investigate a specific simulation-based approximation of the likelihood known as the synthetic likelihood [8].

M. Hainy (✉)
Department of Applied Statistics, Johannes Kepler University, Linz, Austria
e-mail: markus.hainy@jku.at

C.C. Drovandi • J.M. McGree
School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia
e-mail: c.drovandi@qut.edu.au; james.mcgree@qut.edu.au

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_18

To test the algorithms, we apply them to a well-studied sequential design problem for estimating a non-linear function of linear regression parameters introduced by [4], and compare the results between algorithms.

2 Estimation Procedure

2.1 Sequential Monte Carlo Sampling for Sequential Design

A total of T independent observations is to be collected over time. At time t ($t = 1, \ldots, T$), the target distribution is

$$\pi_t(\theta|y_{1:t}, d_{1:t}) \propto f(y_{1:t}|\theta, d_{1:t})\, \pi(\theta),$$

where $y_{1:t} \in \mathscr{Y}^t$ denotes the observations collected until t at design points $d_{1:t} \in \mathscr{D}^t$, $\theta \in \Theta$ are the model parameters, $f(y_{1:t}|\theta, d_{1:t}) = \prod_{j=1}^{t} f(y_j|\theta, d_j)$ denotes the likelihood function, $\pi(.)$ is the prior distribution, and $\pi_t(.|y_{1:t}, d_{1:t})$ is the posterior distribution at time t.

Assume that the posterior distribution at time $t-1$ is approximated by a weighted sample $\{\theta^i_{t-1}, W^i_{t-1}\}_{i=1}^N$, the so-called particle set, where the W^i_{t-1} are the normalised importance weights (i.e., $\sum_{i=1}^N W^i_{t-1} = 1$). The unnormalised posterior sample weights at time t can be obtained by computing

$$w^i_t = W^i_{t-1} f(y_t|\theta^i_{t-1}, d_t).$$

Over time, the effective sample size (ESS), see [6], of this posterior sample degenerates because of increasing imbalances in the weights. Therefore, it is necessary to resample the particles according to their weights when the ESS drops below a minimally acceptable threshold ESS_{min}. The resampled particles will have equal weights, but there will also be duplicated particles. To diversify the particle set, a move step is performed for each particle using a Markov chain Monte Carlo (MCMC) kernel. The MCMC kernel is applied R_t times to increase the chance that the particle is actually moved, where R_t might be determined dynamically by the acceptance rate of a previous move step (for details see [3]). Pseudo-code for the basic SMC sampler for Bayesian sequential design is given in Algorithm 1.

Line 3 of Algorithm 1, where the next design point is selected, demands a more detailed discussion. In myopic sequential design, the next optimal design point is found conditional on the previously collected observations. Therefore, to obtain d_{t+1}, we aim to optimise the expected utility function, which quantifies the amount of information expected to be provided by the experiment. It can be defined

Algorithm 1 SMC sampler for Bayesian sequential design

1. Draw $\theta_0^i \sim \pi(.)$ and set $W_0^i = 1/N$ for $i = 1, \ldots, N$
2. **for** t **in** 1 **to** T **do**
3. Select the next design point, d_t, which optimises the expected utility function (see text)
4. Collect a new observation, which is assumed to follow $y_t \sim f(.|\theta_{\text{true}}, d_t)$ given the "true" parameter θ_{true}
5. Re-weight: $w_t^i = W_{t-1}^i f(y_t|\theta_{t-1}^i, d_t)$ for $i = 1, \ldots, N$
6. Normalise $W_t^i = w_t^i / \sum_{j=1}^N w_t^j$ and set $\theta_t^i = \theta_{t-1}^i$ for $i = 1, \ldots, N$
7. Compute ESS $= 1/\sum_{i=1}^N (W_t^i)^2$
8. **if** ESS $<$ ESS$_{\text{min}}$ **or** $t = T$ **then**
9. Compute the tuning parameters of the MCMC proposal according to the current particle set
10. Resample according to the normalised weights to obtain $\{\theta_t^i, 1/N\}_{i=1}^N$
11. **for** i **in** 1 **to** N **do**
12. Move particle i with an MCMC kernel R_t times with invariant distribution $\pi_t(\theta|y_{1:t}, d_{1:t})$
13. **end for**
14. **end if**
15. **end for**

as follows:

$$U(d|y_{1:t}, d_{1:t}) = \int_{z \in \mathscr{Y}} u(z, d|y_{1:t}, d_{1:t}) f(z|y_{1:t}, d_{1:t}, d)\, \mathrm{d}z,$$

where $u(z, d|y_{1:t}, d_{1:t})$ is the utility function of the future observation z collected at design point d and $f(z|y_{1:t}, d_{1:t}, d)$ is the posterior predictive distribution of z given the previous observations. We will approximate the expected utility $U(d|y_{1:t}, d_{1:t})$ by Monte Carlo integration. Given a sample $\{z^k\}_{k=1}^K$ from the posterior predictive distribution, one can estimate $U(d|y_{1:t}, d_{1:t})$ by

$$\hat{U}(d|y_{1:t}, d_{1:t}) = \frac{1}{K} \sum_{k=1}^{K} u(z^k, d|y_{1:t}, d_{1:t}).$$

Sampling from $z^k \sim f(.|y_{1:t}, d_{1:t}, d)$ can be achieved by sampling θ^k from the current particle set $\{\theta_t^i, W_t^i\}_{i=1}^N$ and then $z^k \sim f(.|\theta^k, d)$.

The utility function $u(z, d|y_{1:t}, d_{1:t})$ depends on the objective of data collection, i.e., estimation, prediction, or model selection. In the example of this contribution, the inferential goal is to estimate the posterior distribution of a function of the parameters. Thus, the aim is to find the design point which leads to posterior distributions which are, on average, as narrow as possible. Suitable utility functions would be the Kullback-Leibler divergence between posterior and prior distribution or the posterior precision. These utility functions are functionals of the posterior $\pi(\theta|y_{1:t}, z, d_{1:t}, d)$. In order to estimate $u(z, d|y_{1:t}, d_{1:t})$, one therefore needs a sample

from $\pi(\theta|y_{1:t}, z, d_{1:t}, d)$. This is easily obtained by updating the weights from the particle set $\{\theta_t^i, W_t^i\}_{i=1}^N$:

$$w(z,d)^i = W_t^i f(z|\theta_t^i, d).$$

To find the optimal design d_{t+1}, we will simply discretise the design space $\mathscr{D}$ and compute $\hat{U}(d|y_{1:t}, d_{1:t})$ for all $d \in \mathscr{D}$. The algorithm for selecting the next design point is given by Algorithm 2.

2.2 Modifications for Intractable Likelihoods

If the likelihoods are intractable, the re-weighting in Line 5 of Algorithm 1 and in Line 5 of Algorithm 2 is not possible. Furthermore, standard MCMC in Line 12 of Algorithm 1 is also precluded. We therefore propose to resort to concepts known as approximate Bayesian computation (ABC) (see [7]) and to replace the likelihood $f(y|\theta, d)$ by the ABC likelihood

$$f_\varepsilon(y|\theta, d) = \int_{y^* \in \mathscr{Y}} K_\varepsilon\big[d\big(s(y), s(y^*)\big)\big] f(y^*|\theta, d)\, \mathrm{d}y^*,$$

where y^* are pseudo-data simulated from the likelihood $f(.|\theta, d)$, $K_\varepsilon[.]$ is a kernel function with bandwidth parameter ε, and $d\big(s(y), s(y^*)\big)$ gives the discrepancy between the summary statistics $s(.)$ of y and y^*. The ABC posterior resulting from this approximation of the likelihood is

$$\pi_\varepsilon(\theta|y, d) \propto f_\varepsilon(y|\theta, d)\pi(\theta).$$

Algorithm 2 Selecting the next design point

1. We have particles at current time t, $\{\theta_t^i, W_t^i\}_{i=1}^N$
2. **for** $d \in \mathscr{D}$ **do**
3. **for** k **in** 1 **to** K **do**
4. Simulate z^k from the posterior predictive distribution: $\theta^k \sim \{\theta_t^i, W_t^i\}_{i=1}^N$ and $z^k \sim f(.|\theta^k, d)$
5. Compute importance weights $w(z^k, d)^i = W_t^i f(z^k|\theta_t^i, d)$ for $i = 1, \ldots, N$
6. Normalise $W(z^k, d)^i = w(z^k, d)^i / \sum_{j=1}^N w(z^k, d)^j$ for $i = 1, \ldots, N$
7. Use $\{\theta_t^i, W(z^k, d)^i\}_{i=1}^N$ to estimate $u(z^k, d|y_{1:t}, d_{1:t})$
8. **end for**
9. Estimate the expected utility $\hat{U}(d|y_{1:t}, d_{1:t}) = (1/K) \sum_{k=1}^K u(z^k, d|y_{1:t}, d_{1:t})$
10. **end for**
11. Set d_{t+1} to the design point $d \in \mathscr{D}$ which optimises $\hat{U}(d|y_{1:t}, d_{1:t})$

It is an approximation of the true posterior distribution and depends on the choices of ϵ, $K_\varepsilon[.]$, and $s(.)$. If $s(.)$ is a sufficient summary statistic, then $\pi_\varepsilon(\theta|y,d) \to \pi(\theta|y,d)$ as $\varepsilon \to 0$.

If it is possible to simulate the sample $y_m^* \sim f(.|\theta,d)$, $m = 1,\ldots,M$, then one can compute an unbiased estimate of the ABC likelihood,

$$f_\varepsilon(y|\theta,d) \approx \frac{1}{M}\sum_{m=1}^{M} K_\varepsilon\big[d\big(s(y),s(y_m^*)\big)\big],$$

and use this estimate instead of the true likelihood in the re-weighting steps, as has been proposed by [5]. In particular, Line 5 of Algorithm 1 could be replaced by the two steps

- Simulate $y_m^{*i} \sim f(.|\theta_{t-1}^i, d_t)$ for $m = 1,\ldots,M$ and $i = 1,\ldots,N$
- Re-weight: $w_t^i = W_{t-1}^i \sum_{m=1}^{M} K_\varepsilon\big[d\big(y_t, y_m^{*i}\big)\big]$ for $i = 1,\ldots,N$

Line 5 of Algorithm 2 is correspondingly substituted by

- Simulate $z_m^{*i} \sim f(.|\theta_t^i, d)$ for $m = 1,\ldots,M$ and $i = 1,\ldots,N$
- Re-weight: $w(z^k,d)^i = W_t^i \sum_{m=1}^{M} K_\varepsilon\big[d\big(z^k, z_m^{*i}\big)\big]$ for $i = 1,\ldots,N$

Since we assume that observations are collected one at a time, we can employ the identity function as our summary statistic in the re-weighting steps. Following [5], in Algorithm 2 the simulation burden can be reduced by re-using the sample $\{\{z_m^{*i}\}_{m=1}^M\}_{i=1}^N$ to update all the particle weights for all $z^k \sim f(.|y_{1:t}, d_{1:t}, d)$.

For the MCMC steps in Line 12 of Algorithm 1, we suggest to employ the likelihood-free extension of the standard MCMC sampler presented in [7]. The ABC-MCMC kernel has the invariant distribution

$$\pi_\varepsilon(\theta, y_{1:t}^*|y_{1:t}, d_{1:t}) \propto K_\varepsilon\big[d\big(s(y_{1:t}), s(y_{1:t}^*)\big)\big] f(y_{1:t}^*|\theta, d_{1:t})\, \pi(\theta).$$

The marginal distribution with respect to θ is the ABC posterior given $y_{1:t}$ and $d_{1:t}$. The proposal distribution is chosen such that the likelihood terms cancel out in the acceptance ratio (for details see [7]). The acceptance probabilities are usually very small for likelihood-free MCMC, so it is expected that many MCMC repetitions are necessary to move a particle. Thus, R_t has to be set to a very high value compared to the standard SMC sampler.

To mitigate the curse of dimensionality when operating on the whole sample $y_{1:t}$ (see [1]), we have to choose a low-dimensional summary statistic $s(y_{1:t})$. In general, this summary statistic will not be sufficient, so the target ABC posterior in the resampling step will not converge to the true posterior distribution when $\varepsilon \to 0$.

2.3 Synthetic Likelihood

A special form of approximation to the likelihood was proposed by [8]. The distribution of the summary statistic $s(y)$ is approximated by a normal distribution known as the *synthetic likelihood*:

$$f\big(s(y)|\theta, d\big) \approx \mathcal{N}\big[s(y)|\,\hat{\mu}_\theta,\,\hat{\Sigma}_\theta\big].$$

The parameters $\hat{\mu}_\theta$ and $\hat{\Sigma}_\theta$ of the synthetic likelihood can be estimated by simulating pseudo-data $y_m^* \sim f(.|\theta, d)$ for $m = 1, \ldots, M$ and computing the sample mean vector and variance-covariance matrix of $\{s(y_m^*)\}_{m=1}^M$. The advantage of the SL ABC approach is that there is no need to choose an ABC tolerance, ε. The major disadvantage is that the true likelihood is approximated by a normal likelihood, which might not always be appropriate.

For the SL ABC sampler, we have to replace Line 5 of Algorithm 1 by

- Simulate $y_m^{*i} \sim f(.|\theta_{t-1}^i, d_t)$ for $m = 1, \ldots, M$ and $i = 1, \ldots, N$
- Compute $\hat{\mu}^i_{\theta_{t-1}}$ and $\hat{\sigma}^{2,i}_{\theta_{t-1}}$ from the sample $\{y_m^{*i}\}_{m=1}^M$ for $i = 1, \ldots, N$
- Re-weight: $w_t^i = W_{t-1}^i\, \mathcal{N}\big[y_t|\,\hat{\mu}^i_{\theta_{t-1}},\,\hat{\sigma}^{2,i}_{\theta_{t-1}}\big]$ for $i = 1, \ldots, N$

and Line 5 of Algorithm 2 by

- Simulate $z_m^{*i} \sim f(.|\theta_t^i, d)$ for $m = 1, \ldots, M$ and $i = 1, \ldots, N$
- Compute $\hat{\mu}^i_{\theta_t}$ and $\hat{\sigma}^{2,i}_{\theta_t}$ from the sample $\{z_m^{*i}\}_{m=1}^M$ for $i = 1, \ldots, N$
- Re-weight: $w(z^k, d)^i = W_t^i\, \mathcal{N}\big[z^k|\,\hat{\mu}^i_{\theta_t},\,\hat{\sigma}^{2,i}_{\theta_t}\big]$ for $i = 1, \ldots, N$

As for the standard ABC algorithm, we can take $s(y) = y$ for the re-weighting steps and re-use the sample $\{\{z_m^{*i}\}_{m=1}^M\}_{i=1}^N$ for all z^k.

Finally, the invariant distribution for the MCMC steps in Line 12 of Algorithm 1 is proportional to

$$\mathcal{N}\big[s(y_{1:t})|\,\hat{\mu}_\theta,\,\hat{\Sigma}_\theta\big]\,\pi(\theta),$$

where $\hat{\mu}_\theta$ and $\hat{\Sigma}_\theta$ are the mean vector and the variance-covariance matrix of the simulated sample $\{s(y^*_{m,1:t})\}_{m=1}^M$ and $y^*_{m,1:t} \sim f(.|\theta, d_{1:t})$.

3 Example

3.1 Example Setup

The example we use to demonstrate our algorithms is a sequential design problem first investigated by [4] in a classical design setting.

The observations y_t follow a normal distribution:

$$y_t \sim \mathcal{N}\big[\,.\,|\,\mu = \theta_1 d_t + \theta_2 d_t^2,\, \sigma^2 = 1\big], \quad t = 1, \ldots, T.$$

This is a simple linear regression model with one quadratic factor d_t, no intercept, and a fixed variance of 1. The design space is $\mathscr{D} = [-1, 1]$, and the goal is to efficiently estimate the maximum or minimum of the response curve, given by the function $g(\theta) = -0.5\theta_1/\theta_2$. Therefore, it is a nonlinear design problem. Ford and Silvey [4] prove that the limiting sequential design for this problem converges to the optimal design measure. If $|g| \leq 0.5$, the optimal design measure puts probability $0.5 + g$ at $d = +1$ and $0.5 - g$ at $d = -1$, while if $|g| > 0.5$, it puts probability $0.5 + 0.25 \cdot g^{-1}$ at $d = +1$ and $0.5 - 0.25 \cdot g^{-1}$ at $d = -1$.

The likelihood function is available for this example, so we can assess the accuracy and the efficiency of the likelihood-free extensions compared to the likelihood-based sequential design algorithm of [2]. Moreover, we can easily check the plausibility of the results by contrasting them to the analytic solution for the classical design problem.

3.2 Settings

We will consider three settings for the true parameters: $(\theta_1, \theta_2)_{\text{true}} = (1, 1)$, $(1, 2)$, and $(1, 4)$. These were also analyzed by [4]. Furthermore, we will assume uniform prior distributions: $\theta_1 \sim U[0, 5]$ and $\theta_2 \sim U[0.001, 5]$.

In accordance with the original design criterion used by [4], we take the precision of the posterior distribution $\pi(g(\theta)|y_{1:t}, z, d_{1:t}, d)$ to be the utility function $u(z, d|y_{1:t}, d_{1:t})$. The objective at time t is to select the design point that *maximises* the expected posterior precision. This design problem is comparable to the classical design problem studied by [4], so we expect to obtain similar results.

For the MCMC move steps, we use random walk proposal distributions with normal increments. The proposal's variance-covariance matrix is given by the empirical variance-covariance matrix of the current particle set.

As summary statistics for the MCMC move steps, we take the ordinary least squares (OLS) estimates of the parameters if the design matrix is invertible; that is, if observations have been collected at more than one design point. The discrepancy function is the Mahalanobis distance between $s(y_{1:t})$ and $s(y^*_{1:t})$ using the variance-covariance matrix of the OLS estimator. When all observations have been collected at only one design point, we take the mean of these observations as the single summary statistic.

We choose a uniform kernel for the standard ABC sampler, i.e., $K_\varepsilon\big[d\big(s(y), s(y^*)\big)\big] \propto 1$ if $d\big(s(y), s(y^*)\big) \leq \varepsilon$ and $K_\varepsilon\big[d\big(s(y), s(y^*)\big)\big] = 0$ if $d\big(s(y), s(y^*)\big) > \varepsilon$.

The tolerance level ε is set to 0.1 for the re-weighting steps and for the MCMC steps using the Mahalanobis distance of the OLS estimates, while ε is set to $0.1/\sqrt{t}$ for the MCMC steps using the sample means.

In all examples, we have a particle set of size $N = 1000$. When searching for the next optimal design point, we obtain a sample of size $K = 1000$ from the posterior predictive distribution for each $d \in \mathscr{D}$ to perform Monte Carlo integration. Furthermore, for the standard ABC and SL ABC algorithms, we set the size of the simulated observation samples to $M = 1000$. Data are collected for $T = 100$ time points as in [4]. The design space is discretised between -1 and 1 with a spacing of 0.1, so at each t there are 21 candidate design points to be evaluated for determining the next design point.

3.3 Results

For each setting for the true parameter θ_{true} $((\theta_1, \theta_2)_{\text{true}} = (1,1), (1,2), (1,4))$ and for each method, we conducted a simulation study comprising 30 simulation runs. Table 1 reports the average posterior means of the final posterior samples, $\{g(\theta^i_{t=100})\}_{i=1}^N$, over the 30 runs for all settings and methods. The standard deviations of the posterior means are given in parentheses. One can see from Table 1 that the distributions of the posterior means are almost identical for all methods and they are all centered around the true values $g(\theta_{\text{true}})$. The standard deviations of the final posterior distributions for $g(\theta_{t=100})$ are also very similar across all methods (not reported). Therefore, for this example, all three methods attain the same level of quality of statistical inference after $T = 100$ observations.

Next, for each method and setting, we compare the distribution of design points visited during data collection to the theoretical asymptotic results derived by [4] for the classical design problem. Almost all design points which were visited during the simulation runs were either equal or close to -1 or 1. Hence, for each method and setting, we consider the average proportion of negative design points over all 30 runs and compare it to the theoretical fraction of observations taken at $d = -1$. The results are given in Table 2. It can be seen that the fractions obtained by our simulation-based methods for a finite sample closely correspond to the theoretical fractions. Furthermore, the differences between the standard SMC sampler and the likelihood-free extensions are very small. The SMC sampler with the standard ABC extensions exhibits a slightly higher variation of the proportion of negative design points than the other samplers.

Table 1 Average posterior means of $g(\theta_{t=100})$ over the 30 simulation runs for different settings and methods. Standard deviations in parentheses

True θ_2	$g(\theta_{\text{true}})$	Standard SMC	SMC with ABC	SMC with SL ABC
1	− 0.5	− 0.500 (0.052)	− 0.512 (0.057)	− 0.526 (0.070)
2	− 0.25	− 0.241 (0.022)	− 0.252 (0.026)	− 0.251 (0.022)
4	− 0.125	− 0.125 (0.013)	− 0.124 (0.013)	− 0.126 (0.013)

Table 2 Theoretical fraction at $d = -1$ and average fractions of design points <0 visited during data collection over the 30 simulation runs for different settings and methods. Standard deviations in parentheses

True θ_2	Theoretical	Standard SMC	SMC with ABC	SMC with SL ABC
1	1	0.863 (0.035)	0.829 (0.046)	0.857 (0.036)
2	0.75	0.733 (0.028)	0.720 (0.037)	0.744 (0.028)
4	0.625	0.627 (0.018)	0.631 (0.028)	0.622 (0.022)

4 Conclusion

In this paper, we have presented approaches for performing sequential design for models with intractable likelihoods by extending the sequential Monte Carlo framework of [2]. A well-known linear regression example was discussed as a proof-of-concept. Future work will consider motivating applications and more sophisticated methods for optimising the utility at each iteration.

Acknowledgements We thank Werner G. Müller for suggesting the illustrative example and three anonymous reviewers for their helpful comments. Markus Hainy has been supported by the French Science Fund (ANR) and Austrian Science Fund (FWF) bilateral grant I-833-N18. Christopher Drovandi was supported by an Australian Research Council's Discovery Early Career Researcher Award funding scheme (DE160100741).

References

1. Blum, M.G.B.: Approximate Bayesian computation: a nonparametric perspective. J. Am. Stat. Assoc. **105**(491), 1178–1187 (2010). doi:10.1198/jasa.2010.tm09448
2. Drovandi, C.C., McGree, J.M., Pettitt, A.N.: Sequential Monte Carlo for Bayesian sequentially designed experiments for discrete data. Comput. Stat. Data Anal. **57**, 320–335 (2013). doi:10.1016/j.csda.2012.05.014
3. Drovandi, C.C., Pettitt, A.N.: Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics **67**(1), 225–233 (2011). doi:10.1111/j.1541-0420.2010.01410.x
4. Ford, I., Silvey, S.D.: A sequentially constructed design for estimating a nonlinear parametric function. Biometrika **67**, 381–388 (1980)
5. Hainy, M., Müller, W.G., Wagner, H.: Likelihood-free simulation-based optimal design with an application to spatial extremes. Stoch. Environ. Res. Risk A. Stoch. Environ. Res. Risk A. **30**(2), 481–492 (2016). doi:10.1007/s00477-015-1067-8
6. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer, New York (2001)
7. Sisson, S.A., Fan, Y.: Likelihood-free Markov chain Monte Carlo. In: Brooks, S.P., Gelman, A., Jones, G., Meng, X.L. (eds.) Handbook of Markov Chain Monte Carlo, Handbooks of Modern Statistical Methods, pp. 319–341. Chapman & Hall/CRC Press, Boca Raton (2011)
8. Wood, S.N.: Statistical inference for noisy nonlinear ecological dynamic systems. Nature **466**, 1102–1104 (2010). doi:10.1038/nature09319

A Confidence Interval Approach in Self-Designing Clinical Trials

Guido Knapp

Abstract In self-designing clinical trials, a special case of adaptive group sequential experiments, a confidence interval for the difference of normal means is derived, in which the results of the respective study stages are combined using the weighted inverse normal method. During the course of the self-designing trial, the sample sizes as well as the number of study stages can be simultaneously determined in a completely adaptive way. Practical rules for updating sample sizes and assigning weights to the study stages are presented. The implementation of the self-designing trial and the resulting confidence interval are demonstrated using real data.

1 Introduction

Let the difference of two normal means be the effect size of interest in a clinical examination for a new agent to a standard agent with regard to (at least) non-inferiority. In the analysis, the confidence interval approach is of particular attractiveness, see [3], since the position of the final confidence interval determines the kind of result of the study, independently of the question whether originally the study was planned as a non-inferiority or a superiority trial.

In the following, we consider flexible adaptive group sequential trials in the sense that, besides the adaptive choice of the sample sizes for the different stages, the number of stages can be determined in an adaptive way, known as self-designing as introduced in [4, 9].

In the self-designing approach, one decides adaptively after each interim-analysis whether exactly one or at least two further study stages will be performed using unblinded results of all already conducted interim-analyses. The self-designing trial ends when the (finite) variance of an a priori fixed final test statistic has been 'used up'. Self-designing rules are derived in [5] where the weighted inverse normal method is used for combining p-values of the study stages with independent samples. Simultaneously, the weights and the sample sizes can be chosen adaptively.

G. Knapp (✉)
Department of Statistics, TU Dortmund University, D-44221 Dortmund, Germany
e-mail: guido.knapp@tu-dortmund.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_19

In a self-designing trial, a confidence interval for difference of two normal means is constructed in [2], where the variance parameter is assumed to be known. As in [9], the sequence of possible sample sizes is fixed in advance and just the weights assigned to the stages of the trial are really chosen adaptively. For unknown variance, an approximate solution is given in [2]. Extending the methodology in [5, 6] to the combination of parameterized p-values, we will derive an exact confidence interval for the difference of normal means with unknown variance. Suitably combined learning rules provide an effective chance to choose both sample sizes and weights simultaneously in an adaptive way. In our approach, we consider parameterized pivotal t-statistics involving the unknown parameter and combine them using the weighted inverse normal method from meta-analysis, see [7].

The outline of the paper is as follows. In Sect. 2, the basics for a self-designing study of comparing normal outcomes are summarized. The construction of a confidence interval for the mean difference is described in Sect. 3. Section 4 contains the adaptive planning for sample sizes and weights when the difference of means is the parameter of interest. In Sect. 5, an example is considered in which the methods presented so far are illustrated.

2 Basic Principles of a Self-Designing Study

Let X_E and X_C be independent normally distributed random variables with mean μ_E in an experimental group E, mean μ_C in an (active) control group C, and common variance $\sigma^2 > 0$, that is, succinctly

$$X_E \sim \mathcal{N}(\mu_E, \sigma^2) \quad \text{and} \quad X_C \sim \mathcal{N}(\mu_C, \sigma^2)\,. \tag{1}$$

A comparative trial is consecutively carried out in a number of, say k, stages with independent samples, denoted by $stg(1), \ldots, stg(k)$. In the i-th stage, we observe responses of the random quantities X_{Eij}, $j = 1, \ldots, n_{Ei} > 1$, and X_{Cij}, $j = 1, \ldots, n_{Ci} > 1$, where n_{Ei} and n_{Ci} are the sample sizes in the respective groups. The mean difference in $stg(i)$ is then

$$Y_i = \frac{1}{n_{Ei}} \sum_{j=1}^{n_{Ei}} X_{Eij} - \frac{1}{n_{Ci}} \sum_{j=1}^{n_{Ci}} X_{Cij} = \bar{X}_{Ei} - \bar{X}_{Ci}, \quad i = 1, \ldots, k. \tag{2}$$

The variance parameter σ^2 is estimated in the i-th stage by the pooled estimator

$$S_i^2 = \frac{1}{n_{Ei} + n_{Ci} - 2} \left\{ \sum_{j=1}^{n_{Ei}} (X_{Eij} - \bar{X}_{Ei})^2 + \sum_{j=1}^{n_{Ci}} (X_{Cij} - \bar{X}_{Ci})^2 \right\}, \quad i = 1, \ldots, k, \tag{3}$$

which follows a scaled χ^2-distribution with $n_{Ei} + n_{Ci} - 2$ degrees of freedom, that is,

$$(n_{Ei} + n_{Ci} - 2)\,\frac{S_i^2}{\sigma^2} \;\sim\; \chi^2_{n_{Ei}+n_{Ci}-2}. \tag{4}$$

An estimator of the variance of Y_i in the i-th stage is given by

$$\widehat{\mathrm{Var}}(Y_i) = \left(\frac{1}{n_{Ei}} + \frac{1}{n_{Ci}}\right) S_i^2, \tag{5}$$

and Y_i and S_i^2 are stochastically independent, $i = 1, \ldots, k$. Let y_i and s_i^2 denote the realizations of the random variables.

Let us assign a positive normed weight, say $w_i > 0$, to each stage i, $i = 1, \ldots, k$, with $\sum_{i=1}^{k} w_i = 1$. Based on considerations in [2, 4–6, 9], the sample sizes as well as the weights may be chosen in a completely adaptive way. All the information of the unblinded data of previous stages can be used to choose simultaneously the sample size and the weight for the next stage. Let $stg(0)$ denote a priori information and/or external restrictions, we express the adaptive choice of sample sizes and weights as

$$n_i = \hat{n}(i-1) = \hat{n}\{stg(0), stg(1), \ldots, stg(i-1)\},\ n_i = n_{Ei} + n_{Ci}, \tag{6}$$

and

$$w_i = \hat{w}(i-1) = \hat{w}\{stg(0), stg(1), \ldots, stg(i-1)\}, \tag{7}$$

where $w_i \le 1 - w_\Sigma(i-1)$, $w_\Sigma(i) = \sum_{j=1}^{i} w_i$, $w_\Sigma(0) = 0$, $w_\Sigma(k) = 1$, $w_i > 0$, $i = 1, \ldots, k$.

Note that the number k of performed stages is random and will be realized during the course of the sequential trial subject to the choice of weights. Of course, k has to be finite (almost surely), and for practical reasons, k should be bounded by some reasonable constant. Introducing a minimum weight for a stage, say $w_{\min}$, $0 < w_{\min} < 1$, we obtain the boundary as $k \le 1/w_{\min}$. A minimum sample size, say $n_{\min}$, may also be introduced, so that

$$n_i \ge n_{\min} \ge 4 \qquad \text{and} \qquad w_i \ge w_{\min} > 0, \quad i = 1, \ldots, k. \tag{8}$$

The use of minimum weight and minimum sample size leads to useful conditions to stop the trial and can adjust some non-practicable suggestions of the (automatic) learning rules for choosing n_i and w_i discussed in Sect. 4.

3 A Confidence Interval for the Mean Difference

Let $\Delta_0 \geq 0$ be an a priori defined non-inferiority bound. We are interested in testing

$$\mathrm{H}_{0,\Delta}: \ \mu_E \leq \mu_C - \Delta \quad \text{versus} \quad \mathrm{H}_{1,\Delta}: \ \mu_E > \mu_C - \Delta \ , \quad 0 \leq \Delta \leq \Delta_0, \tag{9}$$

at a given level α, $0 < \alpha < 1/2$. The alternative hypothesis $\mathrm{H}_{1,\Delta}$ means ($\Delta-$)non-inferiority for $0 < \Delta \leq \Delta_0$ and superiority of E with regard to C for $\Delta = 0$.

Let $\vartheta = \mu_E - \mu_C$ denote the difference of means unbiasedly estimated by Y_i in $stg(i)$, $i = 1, \ldots, k$, see (2). In the i-th stage, let us define the parameterized pivotal t-statistic

$$T_i(\vartheta) = \frac{Y_i - \vartheta}{\sqrt{(1/n_{Ei} + 1/n_{Ci})\, S_i^2}} \sim t_{n_{Ei}+n_{Ci}-2} \ , \tag{10}$$

that is, $T_i(\vartheta)$ follows a t-distribution with $n_{Ei} + n_{Ci} - 2$ degrees of freedom given ϑ.

Let $F_{t(\nu)}$ denote the cumulative distribution function of a t-variate with ν degrees of freedom and Φ^{-1} the inverse of the standard normal distribution function Φ we obtain

$$Z_i(\vartheta) = \Phi^{-1}[F_{t(n_{Ei}+n_{Ci}-2)}\{T_i(\vartheta)\}] \sim \mathcal{N}(0,1) \ , \ i = 1, \ldots, k. \tag{11}$$

Although, the $Z_i(\vartheta)$ from (11) are not independent, k is random, and sample sizes and weights may be chosen adaptively as described in (6) and (7), the final combining statistic follows a specified test distribution, that is,

$$Z_{k,\mathrm{fin}}(\vartheta) = \sum_{i=1}^{k} \sqrt{w_i}\, Z_i(\vartheta) \sim \mathcal{N}(0,1) \quad \text{with} \quad w_\Sigma(k) = \sum_{i=1}^{k} w_i = 1, \tag{12}$$

see [4, 5, 9] for various proofs of (12).

The continuous distribution functions $F_{t(\nu_i)}(\cdot)$ as well as the inverse distribution function $\Phi^{-1}(\cdot)$ are (strictly) monotone increasing functions in their arguments. The pivotal statistic $T_i(\vartheta)$ from (10) is monotone decreasing in ϑ, implying that $\Phi^{-1}\{F_{t(\nu_i)}[T_i(\vartheta)]\}$ is monotone decreasing in ϑ. Hence, the function $Z_{k,\mathrm{fin}}(\vartheta)$ is monotone decreasing in ϑ.

So we can define the following confidence interval on ϑ,

$$\mathrm{CI}(\vartheta) = \left\{ d \in \mathbb{R} \mid \Phi^{-1}(\alpha) \leq Z_{k,\mathrm{fin}}(d) \leq \Phi^{-1}(1-\alpha) \right\} = [\ \vartheta_L \ , \ \vartheta_U \ ], \tag{13}$$

where ϑ_L and ϑ_U are the unique solutions of the equations:

$$Z_{k,\mathrm{fin}}(\vartheta_L) = \Phi^{-1}(1-\alpha) \quad \text{and} \quad Z_{k,\mathrm{fin}}(\vartheta_U) = -\Phi^{-1}(1-\alpha).$$

The confidence coefficient of $\mathrm{CI}(\vartheta)$ is $1-2\alpha$, $0<\alpha<1/2$. Since the solutions in (13) are unique, they can easily be found iteratively using standard statistical software packages.

Let us now apply the confidence interval to the test problem (9). We reject the null hypothesis at level α for the alternative $\mathrm{H}_{1,\Delta}$, $\Delta\in[0\,,\ \Delta_0]$, if $-\Delta$ lies below $\mathrm{CI}(\vartheta)$. However, we do not reject H_{0,Δ_0}, if $\mathrm{CI}(\vartheta)$ covers $-\Delta_0$, more succinctly, with ϑ_L from (13),

$$\begin{array}{l}\text{if } -\Delta<\vartheta_L\,,\ \text{ then reject } \mathrm{H}_{0,\Delta},\\ \text{if } -\Delta_0\geq\vartheta_L\,,\ \text{then stay with } \mathrm{H}_{0,\Delta_0}.\end{array}\tag{14}$$

4 Adaptive Planning for Sample Sizes and Weights

The confidence interval $\mathrm{CI}(\vartheta)$ in (13) is computed after $k-1$ interim analyses based on unblinded data. In cases when an unexpectedly favourable parameter constellation has been observed up to stage j and provided that $w_\Sigma(j)<1$, this may lead to considerations to switch from showing non-inferiority to showing superiority. The trial is then continued by further planning with $\Delta=0$. On the other hand, originally planned as a superiority trial, a first interim analysis may reveal that an unexpectedly large number of subjects would be required. In case of an active control, one may then decide to switch from showing superiority to showing non-inferiority, and to reduce the sample size of the rest of the trial by choosing some $\Delta>0$ in the further planning. Note that in this situation, a non-inferiority bound Δ_0 should have been defined at the beginning of the study, see also the discussion in [3]. In the following, we present some learning rules for choosing sample sizes and weights adaptively with the possibility of switching in the planning between non-inferiority and superiority.

For given type I and II error rates α, $0<\alpha<1$, and β, $0<\beta<1$, respectively, let us consider, for ease of presentation, the approximate normal sample size spending function. Two steering parameters u_j and v_j will be introduced for each stage j in order to cover a wide range of reasonable updating possibilities, whose realizations would then depend on a given practical situation. We plan with equal sample sizes for both groups at each stage. Based on information up to stage j, an estimate $A_j(\Delta)>0$ of the standardized mean difference $(\vartheta+\Delta)/\sigma$ may be given, where $A_j(\Delta)$ is defined below. The power is considered at the point $\vartheta+\Delta=\sigma A_j(\Delta)$ in the alternative $H_{1,\Delta}$. For testing $H_{0,\Delta}$ from (9) by use of a t-test of level α at stage $j+1$, a power of $1-\beta$ is approximately attained when the total sample size for both groups at stage $j+1$ is chosen as

$$f_j(\alpha,\beta,\Delta)=\frac{4\,[\max\{0\,,\ \Phi^{-1}(1-\alpha)+\Phi^{-1}(1-\beta)\}]^2}{A_j(\Delta)^2},\qquad j=0,1,\ldots,k,\tag{15}$$

with estimate

$$A_j(\Delta) = u_j \sum_{i=1}^{j} \frac{\tilde{n}_i}{\sum_{h=1}^{j} \tilde{n}_h} \left(\frac{y_i + \Delta}{s_i} \right) + (1 - u_j) \frac{\mu_{E0} - \mu_{C0} + \Delta}{v_j\, s(j) + (1 - v_j)\, s_0} > 0, \quad \Delta \geq 0,$$

$$s(j) = \left(\sum_{i=1}^{j} \frac{n_i - 2}{\sum_{h=1}^{j} n_h - 2j}\, s_i^2 \right)^{1/2}, \quad \tilde{n}_i = \frac{2}{1/n_{E_i} + 1/n_{C_i}},$$

$$n_i = n_{E_i} + n_{C_i}, \qquad 0 \leq u_j \leq 1, u_0 = 0, \quad \text{and} \qquad 0 \leq v_j \leq 1, v_0 = 0,$$

where $\mu_{E0} - \mu_{C0} + \Delta > 0$ denotes a predefined value from the alternative $H_{1,\Delta}$ at $stg(0)$ and $s_0^2 > 0$ a supposed value for σ^2. An unrealistically small value in (15) may be replaced by some reasonable sample size, for instance, by $n_{\min}$ from (8).

Let us briefly comment on the role of the two steering parameters u_j and v_j, $0 \leq u_j \leq 1$ and $0 \leq v_j \leq 1$. By choosing $u_j = 0$ and $v_j = 0$, we get a purely prior information based sample size plan with respect to the parameters. The choice $u_j = 0$ and $v_j > 0$ leads to adaptive plans that only use updated variances, where $s(j)^2$ is the pooled estimate of σ^2 up to $stg(j)$. For $u_j = 1$, involving $\tilde{n}_i$, the harmonic mean of realized sample sizes, the term $A_j(\Delta)$ is a short-cut version of the meta-analytical combination of standardized mean differences.

Let us assume that up to $stg(j-1)$ we have determined sample sizes and weights where $w_\Sigma(j-1) < 1$, by planning with $\Delta_1, \ldots, \Delta_{j-1} \in [0, \Delta_0]$ and at $stg(j)$ we want to plan with Δ_j, that is, we have in mind to reject H_{0,Δ_j}, $\Delta_j \in [0, \Delta_0]$, see (9). With the realized sample sizes n_{E_i} and n_{C_i}, $i = 1, \ldots, j-1, j \geq 2$, and defining $Z_0(-\Delta_j) = 0$, we compute the combination statistic up to $stg(j-1)$, see (11),

$$Z_{j-1}^{\Sigma}(-\Delta_j) = \sum_{i=1}^{j-1} \sqrt{w_i}\, Z_i(-\Delta_j)\,, \quad j \geq 1. \tag{16}$$

Suppose we want to obtain a significant result after the next stage by assigning the full remaining weight $1 - w_\Sigma(j-1)$ to this stage. Then, by use of the projected p-value, say $\hat{p}_{j,m}$, the following combination statistic

$$Z_{j,m}(-\Delta_j) = Z_{j-1}^{\Sigma}(-\Delta_j) + \sqrt{1 - w_\Sigma(j-1)}\, \Phi^{-1}(1 - \hat{p}_{j,m}), \quad j \geq 1, \tag{17}$$

should attain the critical value $\Phi^{-1}(1-\alpha)$, that is,

$$\hat{p}_{j,m} = 1 - \Phi\left[\left\{ \Phi^{-1}(1-\alpha) - Z_{j-1}^{\Sigma}(-\Delta_j) \right\} \Big/ \sqrt{1 - w_\Sigma(j-1)} \right], \qquad j \geq 1. \tag{18}$$

This projected p-value is gained with the (conditional) power $1 - \beta$ at $\vartheta + \Delta_j = \sigma A_{j-1}(\Delta_j) > 0$ by choosing the sample size for the next stage j according to (15) as

$$m_j = m_j(\beta) = f_{j-1}(\hat{p}_{j,m}, \beta, \Delta_j), \qquad j \geq 1. \tag{19}$$

In the above procedure, the full weight is used up and stage j is the last one. In those cases when estimates of the parameters are not yet stable, only a part of $m_j(\beta)$ should be used as sample size n_j, that is $n_j = \varepsilon_j\, m_j(\beta)$, with $0 < \varepsilon_j \leq 1$. The remaining weight after stage $(j-1)$ is also divided proportionally to assign the weight $w_j = \varepsilon_j \{1 - w_\Sigma(j-1)\}$ to stage j, that is, summarized,

$$w_j = \varepsilon_j \{1 - w_\Sigma(j-1)\}, \quad n_j = \varepsilon_j\, m_j(\beta), \quad n_{Ej} = n_{Cj} \approx n_j/2, \quad j \geq 1. \tag{20}$$

The choice of w_j means a proportional partition of the remaining variance of the final $\mathcal{N}(0,1)$-test distribution.

Choosing a smaller power $(1-\beta_j)$, a possible choice of ε_j is provided by

$$\varepsilon_j = \varepsilon_j(\beta_j) = \frac{m_j(\beta_j)}{m_j(\beta)}, \quad m_j(\beta_j) = f_{j-1}(\hat{p}_{j,m}, \beta_j, \Delta_j), \quad \beta \leq \beta_j < 1,\ j \geq 1. \tag{21}$$

Note that β_j is only a lower bound of the type II error rate in stage j as long as $w_j < 1 - w_\Sigma(j-1)$. A similar basic idea is discussed in [5] and applied in a 3-stage self-designing clinical trial with normal outcomes in [6].

5 An Example Showing Switching from Non-inferiority to Superiority

In [8], a controlled clinical trial concerning patients with acne papulopustulosa was discussed assuming the distributional properties from Sect. 2. An adaptive 3-stage group sequential test of Pocock type was used, which led to early stopping for superiority of E with respect to C after the second stage at the one-sided overall significance level of $\alpha = 0.005$. The response variable was the reduction of bacteria (after 6 weeks of treatment) from baseline, examined on agar plates and measured as log CFU/cm^2, CFU: colony forming units. We have used the parameter estimates as presented in Table 1. The non-inferiority margin may be predefined as $\Delta_0 = 0.1$.

The test level is chosen as $\alpha = 0.005$ and the power as $1 - \beta = 0.80$. Each stage is planned with equal sample sizes in both groups. Planning with $\Delta_1 = 0.1$

Table 1 Self-designing two-stage clinical trial concerning patients with acne papulopustulosa: data and confidence intervals on the treatment difference $\vartheta = \mu_E - \mu_C$ and on the standard deviation σ

Stage	Adaptive sample size	Adaptive weight	Treatment difference	Standard deviation	p-value $p_i(-\Delta)$	
0	—	—	0.8	1.0	$p_i(-0.1)$	$p_i(0)$
1	24	$\sqrt{0.4} = 0.63$	1.549	1.316	0.0028	0.0043
2	12	$\sqrt{0.6} = 0.77$	1.580	1.472	0.0381	0.0463

for showing non-inferiority, we get the prior guess $A_0(\Delta_1) = 0.9$ using the prior guesses of ϑ and σ from Table 1. With the critical value $\Phi^{-1}(0.995) = 2.576$, we obtain the total sample size for a one-stage trial using (15)

$$m_1 = f_0(0.005,\ 0.2,\ 0.1) = 57.6.$$

Note that, for the superiority test with $\Delta = 0$, the total sample size would be 73.

It was intended to start with a $(1/3)m_1$. But, by randomizing medications in blocks of size 6, the first sample was chosen as $n_1 = 24$, that is, $\varepsilon_1 = n_1/m_1 = 0.4 = w_1$, see (20). The trial started and we obtained $y_1 = 1.549$ and $s_1 = 1.316$, leading to the small p-value $p_1(-0.1) = 0.0028$. Consequently, we decided to switch to showing superiority. That means, we chose $\Delta_2 = 0$.

First, we computed $Z_1(-\Delta_2) = \sqrt{0.4}\ \Phi^{-1}(1 - 0.0043) = 1.66$ and then the projected p-value, see (18),

$$\hat{p}_{2,m} = 1 - \Phi\{(2.576 - 1.66)/\sqrt{0.6}\} = 1 - \Phi(1.18),$$

leading to, see (19), with $\Delta_2 = 0$,

$$m_2 = \frac{4\,(1.18 + 0.84)^2}{(1.549\ /\ 1.316)^2} = 11.7.$$

We put $u_j = 1$ in (15) because the prior guesses turned out as too cautious. So it was decided to finish the trial by assigning the full remaining weight to the second stage, $w_2 = 0.6$, and to choose the sample size $n_2 = 12$.

By the results of the second stage, see Table 1, we obtain

$$Z_{2,\text{fin}}(0) = 0.63 \cdot 2.63 + 0.77 \cdot 1.68 = 2.95 > 2.576,$$

and equating $Z_{2,\text{fin}}(\vartheta)$ to 2.576 and to -2.576 we obtained the 99 %-confidence interval $\text{CI}(\vartheta) = [0.231, 2.894]$. This led to rejection of the one-sided null hypothesis in favour of superiority. Note that the final approximative repeated confidence interval was $[0.21, 2.92]$ in [8] leading to the same test decision as in our case.

6 Discussion

In Sect. 4, we proposed learning rules for updating sample sizes and study weights in a self-designing trial. Incorporating the minimal weight from (8) in the rules leads to a maximum number of study stages to be undertaken.

In the adaptive planning of sample sizes, we used a normal approximation in (15) for a t-variate with $(n-2)$ degrees of freedom. Nearly exact values can be achieved by replacing n by $n_{corr} = n(n-2)/(n-4), n \geq 5$. Moreover, we assume equal sample

sizes for both groups in (19). This assumption may be relaxed when applying ideas of response-adaptive randomized trials, see [10], in this framework. Note that the adaptive weights in (20) mainly depend on the total sample sizes in the study stages and are not affected by the allocation scheme.

We used the inverse normal method for combining parameterized p-values of test statistics following a continuous distribution. The recursive combination tests based on the inverse normal method exhaust the type I error rate in this case, see [1], and consequently, the confidence interval in (13) is an exact one. For non-continuous distributions, the conditional distribution of the parameterized p-value in each stage given all p-values of previous stages must be stochastically larger than or equal to the uniform distribution on [0, 1], briefly p-clud, so that the recursive combination tests keep the predefined type I error. Following the strategy in Sect. 3 for a parameterized pivotal statistic, fulfilling the p-clud condition, instead of $T_i(\vartheta)$ in (10) would yield a confidence interval for the parameter of interest with confidence coefficient of at least $(1 - 2\alpha)$. Explicit learning rules for updating samples size and study weights are then context-dependent and may still have to be developed.

Acknowledgements This paper is based on some earlier joint work with Joachim Hartung.

References

1. Brannath, W., Posch, M., Bauer, P.: Recursive combination tests. J. Am. Stat. Assoc. **97**, 236–244 (2002)
2. Cheng, Y., Shen, Y.: Estimation of a parameter and its exact confidence interval following sequential sample size reestimation trials. Biometrics **60**, 910–918 (2004)
3. EMEA (The European Agency for the Evaluation of Medicinal Products) Points to Consider on Switching between Superiority and Non-inferiority. London, CPMP/EWP/482/99 (2000)
4. Fisher, L.: Self-designing clinical trials. Stat. Med. **17**, 1551–1562 (1998)
5. Hartung, J.: A self-designing rule for clinical trials with arbitrary response variables. Control. Clin. Trials **22**, 111–116 (2001)
6. Hartung, J.: Flexible designs by adaptive plans of generalized Pocock- and O'Brien-Fleming-type and by self-designing clinical trials. Biom. J. **48**, 521–536 (2006)
7. Hartung, J., Knapp, G., Sinha, B.K.: Statistical Meta-Analysis with Applications. Wiley, Hoboken (2008)
8. Lehmacher, W., Wassmer, G.: Adaptive sample size calculations in group sequential trials. Biometrics **55**, 1286–1290 (1999)
9. Shen, Y., Fisher, L.: Statistical inference for self-designing clinical trials with a one-sided hypothesis. Biometrics **55**, 190–197 (1999)
10. Zhu, H., Hu, F.: Sequential monitoring of response-adaptive randomized clinical trials. Ann. Stat. **38**, 2218–2241 (2010)

Conditional Inference in Two-Stage Adaptive Experiments via the Bootstrap

Adam Lane, HaiYing Wang, and Nancy Flournoy

Abstract We study two-stage adaptive designs in which data accumulated in the first stage are used to select the design for a second stage. Inference following such experiments is often conducted by ignoring the adaptive procedure and treating the design as fixed. Alternative inference procedures approximate the variance of the parameter estimates by approximating the inverse of expected Fisher information. Both of these inferential methods often rely on normal distribution assumptions in order to create confidence intervals. In an effort to improve inference, we develop bootstrap methods that condition on a non-ancillary statistic that defines the second stage design.

1 Introduction

In many experiments, the evaluation of designs with respect to a specific objective requires knowledge of the true model parameters. Examples include optimal designs for the estimation of nonlinear functions of the parameters in linear models; optimal designs for the estimation of linear functions of parameters in nonlinear models; and dose-finding designs where it is desired to treat patients at a pre-specified quantile of the response function. In the absence of perfect knowledge of the model parameters, it is appealing to update initial parameter estimates using data accumulated from all previous stages and to allocate observations in the current stage by an assessment of designs evaluated at these estimates. Such procedures result in designs that are functions of random variables whose distributions depend on the model parameters, i.e., the designs are not ancillary statistics.

A. Lane (✉)
Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA
e-mail: Adam.Lane@cchmc.org

H. Wang
University of New Hampshire, 33 Academic Way, Durham, NH 03824, USA
e-mail: HaiYing.Wang@unh.edu

N. Flournoy
University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA
e-mail: flournoyn@missouri.edu

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_20

In practice it is common to conduct inference ignoring the adaptive nature of the experiment and treating the design as if it were an ancillary statistic. The potential issues resulting from such procedures are well known; see [1]. Briefly, when a design is ancillary, it is non-informative with respect to model parameters. Analysing adaptive experiments with non-ancillary designs as if the design were ancillary does not account for all the information contained in the design. An alternative often suggested is to approximate the variance of the maximum likelihood estimate (MLE) with an approximation of the inverse of the expected Fisher information (defined below). Regardless of the method, the distribution of the MLE resulting from adaptive experiments are often assumed to follow a normal distribution when used to create confidence intervals.

We develop three bootstrap procedures that approximate the distribution of the MLE conditional on a non-ancillary statistic that defines the second stage design. The bootstrap reduces the reliance on the assumption of normality. A bootstrap procedure in the context of adaptive experiments with binary outcomes, primarily in the context of urn models, has been developed previously [7].

2 The Model and an Illustration of the Problem

Consider an experiment conducted to measure a constant θ. Suppose it is possible to obtain unbiased observations y of θ from two sources; each source $k = r, s$ produces errors from a $\mathcal{N}(0, \sigma_k^2)$, where $\sigma_r^2 \neq \sigma_s^2$ are known constants. Throughout this paper Y and y denote the random variable and its observed realization, respectively. A function $\psi(\theta)$ is defined to determine which of the two sources should be used based on maximizing efficiency, ethics, cost or other considerations. For illustration, we use $\psi(\theta) = r$ if $\theta < c$ and s otherwise, given some constant c.

The adaptive procedure is as follows: obtain an initial cohort of n_1 independent observations, $y_1 = (y_{11}, \ldots, y_{1n_1})$, from source r. Evaluate ψ at the MLE of θ based on first stage observations; $\hat{\theta}_1 = \bar{y}_1 = \sum_{j=1}^{n_1} y_{1j}/n_1$. The function $\psi(\bar{y}_1)$ determines the source from which a second cohort of n_2 independent observations, $y_2 = (y_{21}, \ldots, y_{2n_2})$, will be obtained and induces a correlation between Y_1 and Y_2. This model has been used to illustrate conditional inference in experiments with ancillary designs; see [2] and [4]. Here we use the model to illustrate conditional inference with non-ancillary designs in adaptive experiments.

Let $N = n_1 + n_2$ represent the total sample size; and let $S_r = (-\infty, c)$ and $S_s = (c, \infty)$ be the regions of the parameter space that selects the second stage source. The second stage design is completely determined by the non-ancillary variable $I(\overline{Y}_1 \in S_k)$, $k = r, s$, where $I(\cdot)$ is the indicator function. Therefore, $Y_2 | \overline{Y}_1 \in S_k$ is independent of $\overline{Y}_1$, $k = r, s$. The log likelihood is $l_\theta = -n_1 (\bar{y}_1 - \theta)^2 / 2\sigma_r^2 - n_2 (\bar{y}_2 - \theta)^2 / 2\sigma_{\psi(\bar{y}_1)}^2 +$ constant. Defining $w(\bar{y}_1) = (n_1/\sigma_r^2 + n_2/\sigma_{\psi(\bar{y}_1)}^2)^{-1}$, the MLE based on both stages is $\hat{\theta} = w(\bar{y}_1)(n_1\bar{y}_1/\sigma_r^2 + n_2\bar{y}_2/\sigma_{\psi(\bar{y}_1)}^2)$.

To approximate the distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k$, $k = r, s$, three bootstrap methods are developed. Resulting parameter estimates and confidence intervals are compared to analyses where the variance of the MLE is approximated using the inverse of the observed and expected Fisher information: $\mathscr{I}(\hat{\theta}) = -[\partial^2 l_\theta / \partial \theta^2]_{\theta=\hat{\theta}} = w^{-1}(\overline{y}_1)$ and $\mathscr{F}_\theta = \mathrm{E}[\mathscr{I}(\hat{\theta})]$, respectively. Confidence intervals for comparative methods are then constructed using normal quantiles. If the design were ancillary, then the variance of the MLE would be $w(\overline{y}_1)$. Using $\mathscr{F}_\theta^{-1}$ to approximate the variance, does not treat the design as ancillary. But it is a function of θ and must be estimated; $\mathscr{F}_{\hat{\theta}}^{-1}$ is used here. In contrast, since $Y_2|\overline{Y}_1 \in S_k$ is independent of $\overline{Y}_1$, $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k$ has variance

$$\mathrm{Var}\left[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k\right] = w_k^2\left(\frac{n_1^2}{\sigma_r^4}\mathrm{Var}\left[\overline{Y}_1|\overline{Y}_1 \in S_k\right] + \frac{n_2}{\sigma_k^2}\right),$$

where $w_k = (n_1/\sigma_r^2 + n_2/\sigma_k^2)^{-1}$. The design of the second stage is completely determined by $I(\overline{Y}_1 \in S_k)$. Therefore $\{\psi(\overline{Y}_1)|\overline{Y}_1 \in S_k\} = k$. The function $w(\overline{Y}_1)$ depends on $\overline{Y}_1$ only through $\psi(\overline{Y}_1)$. Hence $\{w(\overline{Y}_1)|\overline{Y}_1 \in S_k\} = w_k$, for $k = r, s$.

Let $p_k = P(\overline{Y}_1 \in S_k)$, $E_k = \mathrm{E}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k]$, $V_k = \mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_k]$ and $\mathscr{I}_k = \mathscr{I}(\hat{\theta})$, $k = r, s$ for short.

From a series of 10,000 simulations, Table 1 presents, for $k = r, s$, p_k, E_k, $N\widehat{se}^2$ and tail probabilities, $\mathrm{P}[\theta < C_l]$ and $\mathrm{P}[\theta > C_u]$, where $\widehat{se}^2$ is either $\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$, V_k, $\mathscr{I}_k^{-1}$ or $\mathscr{F}_\theta^{-1}$; $C_l = \hat{\theta} - Z_{1-\alpha/2}\widehat{se}$, $C_u = \hat{\theta} + Z_{1-\alpha/2}\widehat{se}$ and Z_α is the α quantile of the standard normal distribution. Values of $\theta = 1$, $\sigma_r = 1$, $\sigma_s = 3$, $n_1 = 50$, $n_2 = 150$, $c = \theta + 1/\sqrt{n_1}$ and $\alpha = 0.05$ are used throughout. Both $\mathscr{I}_k^{-1}$ and $\mathscr{F}_\theta^{-1}$ are significantly greater than V_k, $k = r, s$. Both V_k, $k = r, s$, are significantly less than $\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$. If $\overline{Y}_1 \in S_r$, then E_r is nearly unbiased, and despite slightly unbalanced tail probabilities, overall coverage would be adequate if V_r were known and could be used. When $\overline{Y}_1 \in S_s$, the bias is considerable and the coverage is unacceptable. We focus throughout on improvements when $\overline{Y}_1 \in S_s$.

Table 1 Conditional probability, conditional expectation, variance approximation method along with its estimate and tail probabilities by the source of the second stage observations

Source	p_k	E_k	Variance approximation	$N\widehat{se}^2$	$\mathrm{P}[\theta < C_l]$	$\mathrm{P}[\theta > C_u]$
r	0.77	0.99	V_r	0.88	0.02	0.04
			$\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$	1.77	0.00	0.01
			$\mathscr{I}_r^{-1}$	1.00	0.01	0.03
			$\mathscr{F}_\theta^{-1}$	1.47	0.00	0.01
s	0.23	1.14	V_s	1.24	0.39	0.00
			$\mathrm{Var}[\hat{\theta}(\overline{Y}_1, \overline{Y}_2)]$	1.77	0.26	0.00
			$\mathscr{I}_s^{-1}$	3.00	0.10	0.00
			$\mathscr{F}_\theta^{-1}$	1.47	0.33	0.00

3 Conditional Bootstrap Inference

In an effort to improve inference following adaptive experiments, three bootstrap methods are developed. We begin with a straightforward and intuitive conditional bootstrap procedure. Unfortunately, as will be discussed, the conditions required for this procedure to give accurate inference are extremely restrictive.

Conditional Bootstrap Method 1 (BM1)

1. Construct a probability distribution, $\hat{F}_i$, by putting mass $1/n_i$ at each point $y_{i1}, \ldots, y_{in_i}$, $i = 1, 2$. With fixed $\hat{F}_1$ and $\hat{F}_2$, draw random samples of size n_1 and n_2 from $\hat{F}_1$ and $\hat{F}_2$, respectively. Denote the vector of bootstrap samples as $y_i^*, i = 1, 2$.
2. Repeat step 1 B times. For the bootstrap samples satisfying $\bar{y}_1^* \in S_{\psi(\bar{y}_1)}$, use $\bar{y}_1^*$ and $\bar{y}_2^*$ to find $\hat{\theta}^*$. Use $(C_l, C_u) = (Q^*_{\alpha/2}, Q^*_{1-\alpha/2})$ as the $(1-\alpha)$ confidence interval, where Q^*_α is the α quantile of the bootstrapped sample distribution $\hat{\theta}^*(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in S_{\psi(\bar{y}_1)}$.

Let P^* be the empirical probability measure given Y_1 and let $[\cdot]_b$ represent the bth sampled bootstrap. Then $P^*(\overline{Y}_1^* \in S_{\psi(\bar{y}_1)}) \approx \sum_{b=1}^{B} I([\bar{y}_1^*]_b \in S_{\psi(\bar{y}_1)})/B$ is the probability the mean of a bootstrap sample is in $S_{\psi(\bar{y}_1)}$.

Table 2 presents the simulation results p_s, E_s, V_s, C_l, C_u, $\mathrm{P}[\theta < C_l]$ and $\mathrm{P}[\theta > C_u]$. The first row results are for the distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$. The second and third row results use $\mathscr{I}_s^{-1}$ and $\mathscr{F}_{\hat{\theta}}^{-1}$, and tail probabilities found using

Table 2 Results are conditional on the region $\overline{Y}_1 \in S_s$. Method describes the procedure by which the approximate distribution was obtained; p_s is approximated by $\sum_{b=1}^{B} I([\bar{y}_1^*]_b \in S_s)/B$ for the bootstrap and by $P(\overline{Y}_1 \in S_s|\theta = \hat{\theta})$ for the expected information. E_s, V_s, the confidence limits and tail probabilities are also provided. All values are averages from 10,000 simulations except the column p_s which uses the median. "True" refers to either the empirical distribution of $\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ or $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ obtained from simulation

Estimate	Method	p_s	E_s	N*V$_s$	C_l	C_u	P[$\theta < C_l$]	P[$\theta > C_u$]
$\hat{\theta}$	True	0.23	1.14	1.24	0.99	1.30	–	–
$\hat{\theta}$	$\mathscr{F}_{\hat{\theta}}^{-1}$	0.59	1.14	2.17	0.93	1.34	0.14	0.00
$\hat{\theta}$	$\mathscr{I}_s^{-1}$	–	1.14	3.00	0.90	1.38	0.10	0.00
$\hat{\theta}^*$	BM1	0.67	1.19	1.87	1.02	1.39	0.63	0.00
$\hat{\theta}^+$	BM2	0.25	1.13	1.36	0.99	1.30	0.41	0.00
$\tilde{\theta}$	True	0.23	1.00	5.25	0.66	1.28	–	–
$\tilde{\theta}$	$\Delta_{\tilde{\theta}}$	0.74	1.00	6.47	0.65	1.35	0.02	0.01
$\tilde{\theta}^+$	BM3	0.25	0.98	5.60	0.65	1.29	0.02	0.02

quantiles of the standard normal distribution. The fourth row results are from BM1 using $B = 1000$ bootstrap samples. All values are averages except p_s whose values are skewed and for which the median provides a better summary of the data. Both $P^*(\overline{Y}_1^* \in S_s)$ and $P(\overline{Y}_1 \in S_s|\theta = \hat{\theta})$ are significantly greater than p_s. The methods BM1 and $\mathscr{F}_{\hat{\theta}}^{-1}$ produce similar variances, both of which are less than $\mathscr{I}_s^{-1}$. All three approximations are significantly greater than V_s. However, the mean of the bootstrap distribution has greater bias than $\hat{\theta}$. Thus BM1 does not improve inference in comparison to using $\mathscr{F}_{\hat{\theta}}^{-1}$.

3.1 Technical Details for the Conditional Bootstrap

In this section we consider the MLE conditional on an indicator function of the first stage sample mean. This illustrates the implications of conditioning in two-stage adaptive experiments with non-ancillary designs and why the inference produced by BM1 was poor. Let $T_k = (a_k, d_k)$, where $a_k = \theta + a_k'/\sqrt{n_1}$, $d_k = \theta + d_k'/\sqrt{n_1}$, a_k' and d_k' are constants. This is a slightly more general division of the parameter space than S_k which has only one of a_k' and d_k' finite. Note, we consider the parameter space defined in a local neighborhood of θ. This is done so that $P(\overline{Y}_1 \in S_k), k = r, s$ has positive probability not equal to 1 for large n_1; otherwise the design would be deterministic for large first stage sample sizes and not adaptive. Consider

$$
\begin{aligned}
&P^*\left[\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x | \sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] \\
&\quad = \frac{P^*\left[\{\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x\} \cap \{\sqrt{n_1}(\overline{Y}_1^* - \overline{Y}_1) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\}\right]}{P^*\left[\sqrt{n_1}(\overline{Y}_1^* - \overline{Y}_1) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right]} \\
&\quad = \frac{P\left[\{\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x\} \cap \{\sqrt{n_1}(\overline{Y}_1 - \theta) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\}\right] + o(1)}{P\left[\sqrt{n_1}(\overline{Y}_1 - \theta) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o(1)} \\
&\quad = P\left[\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x | \sqrt{n_1}\left(\overline{Y}_1 - \theta\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o(1),
\end{aligned}
\tag{1}
$$

where $x \in \mathbb{R}$, and the equalities hold almost surely under P. Note, from the left hand side of equation (1), that a conditional bootstrap should include bootstrap samples satisfying $\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)$, i.e., $\overline{Y}_1^* \in (a_k - \overline{\varepsilon}_1, d_k - \overline{\varepsilon}_1)$, where $\overline{\varepsilon}_1 = \overline{Y}_1 - \theta$. BM1 considers bootstrap samples satisfying $\overline{Y}_1^* \in (a_k, d_k)$ and hence

its poor performance. Unfortunately, θ is unknown and must be estimated. In (1), if one can replace θ with an estimator (say $\tilde{\theta}$) that converges to θ at a rate faster than $\sqrt{n_1}$, then

$$P^*\left[\sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) < x | \sqrt{n_1}\left(\overline{Y}_1^* - \overline{Y}_1\right) \in \sqrt{n_1}(a_k - \tilde{\theta}, d_k - \tilde{\theta})\right]$$
$$= P\left[\sqrt{n_1}\left(\overline{Y}_1 - \theta\right) < x | \sqrt{n_1}\left(\overline{Y}_1 - \theta\right) \in \sqrt{n_1}(a_k - \theta, d_k - \theta)\right] + o_P(1). \tag{2}$$

In a single stage experiment this would not be possible. However, in a two-stage adaptive experiment in which $n_1 = o(n_2)$, such an estimator may exist. Theoretically this is somewhat unsatisfactory, but practically this is interesting especially since it has been shown that two-stage experiments are optimized when $n_1 = O(\sqrt{n_2})$; see [5] and [6]. It is also common, for practical or logistical reasons, that a small pilot study precedes a much larger follow-up.

Note $\overline{Y}_1 | \overline{Y}_1 \in T_k \sim \mathcal{TN}(\theta, \sigma_k^2/n_1; T_k)$, where $\mathcal{TN}$ denotes a truncated normal distribution. Suppose an estimator, $\tilde{\theta}$, exists as described and let $\tilde{T}_k = (a_k - \tilde{\varepsilon}_1, d_k - \tilde{\varepsilon}_1)$, where $\tilde{\varepsilon}_1 = \overline{Y}_1 - \tilde{\theta}$. Then (2) implies $\{\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\}$ can be approximated by a $\mathcal{TN}(\overline{y}_1, \sigma_k^2/n_1; \tilde{T}_k)$ and therefore

$$\mathrm{E}^*_{\overline{Y}_1}\left[\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\right] \approx \overline{Y}_1 + b_{1k}(\tilde{\theta}); \tag{3}$$

$$\mathrm{Var}_{\overline{Y}_1}\left[\overline{Y}_1^* | \overline{Y}_1^* \in \tilde{T}_k\right] \approx \frac{\sigma_k^2}{n_1}\left[1 + \frac{\gamma_{\tilde{\theta}}(a_k)\phi\{\gamma_{\tilde{\theta}}(a_k)\} - \gamma_{\tilde{\theta}}(d_k)\phi\{\gamma_{\tilde{\theta}}(d_k)\}}{U}\right] - \left\{b_{1k}(\tilde{\theta})\right\}^2.$$

where $b_{1k}(\theta) = \sigma_k \left[\phi\{\gamma_\theta(a_k)\} - \phi\{\gamma_\theta(d_k)\}\right] / (\sqrt{n_1} U)$, $\gamma_\theta(a_k) = \sqrt{n_1}(a_k - \theta)/\sigma_k$, $U = \Phi\{\gamma_\theta(b)\} - \Phi\{\gamma_\theta(a)\}$; $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability density function (PDF) and cumulative distribution function (CDF) of the standard normal distribution, respectively. Note a conditional bootstrap leads to an additional bias term in the expectation in equation (3). This must be accounted for in any conditional bootstrap procedure.

4 Adjusted Conditional Bootstrap Methods

The bootstrap methods developed in this section are predicated on the existence of an estimator that converges to θ at a rate faster than $\sqrt{n_1}$. Provided standard regularity conditions hold and $n_1 = o(n_2)$, such an estimator will always exists in the form of the second stage MLE conditional on the second stage design. However, in cases where the bias has an explicit form, it is possible to obtain $\tilde{\theta}$ by adapting

the bias reduction method suggested in [8]. Note

$$\begin{aligned}\mathrm{E}\left[\hat{\theta}(\overline{Y}_1,\overline{Y}_2)|\overline{Y}_1\in S_k\right] &= w_k\mathrm{E}\left[n_1\overline{Y}_1/\sigma_k^2+n_2\overline{Y}_2/\sigma_k^2|\overline{Y}_1\in S_k\right]\\ &= w_k\left(n_1/\sigma_r^2+n_2/\sigma_k^2\right)\theta+w_kn_1b_{1k}(\theta)/\sigma_r^2\\ &= \theta+w_kn_1b_{1k}(\theta)/\sigma_r^2.\end{aligned}$$

Let $b_k(\theta)=w_kn_1b_{1k}(\theta)/\sigma_r^2$. Then a bias corrected estimate of θ is $\tilde{\theta}=\hat{\theta}-b_k(\tilde{\theta})$. The Newton-Raphson method was used to solve this equation. One iteration was sufficient, that is, $\tilde{\theta}=\hat{\theta}-\lambda_k(\hat{\theta})$, where $\lambda_k(\hat{\theta})=b_k(\hat{\theta})/[1+(\partial b_k(\theta)/\partial\theta)_{\theta=\hat{\theta}}]$ was used for analytic expressions and numeric calculations.

Now we develop a bootstrap method that adjusts the conditioning region per the discussion in Sect. 3.1. Let $\tilde{S}_r=\{-\infty,c-\tilde{\varepsilon}_1\}$ and $\tilde{S}_s=\{c-\tilde{\varepsilon}_1,\infty\}$, where $\tilde{\varepsilon}_1=\overline{Y}_1-\tilde{\theta}$.

Adjusted Conditional Bootstrap Method (BM2): Repeat BM1 keeping only bootstrap samples satisfying $\overline{y}_1^*\in\tilde{S}_{\psi(\overline{y}_1)}$. Let $\hat{\theta}^+=\hat{\theta}^*-b_{\psi(\overline{y}_1)}(\tilde{\theta})$ and $(C_l^+,C_u^+)=(Q_{\alpha/2}^+,Q_{1-\alpha/2}^+)$ as the $(1-\alpha)$ confidence interval, where Q_α^+ is the α quantile of the bootstrap sample distribution of $\hat{\theta}^+(\overline{Y}_1^*,\overline{Y}_2^*)|\overline{Y}_1^*\in\tilde{S}_{\psi(\overline{y}_1)}$.

Note $\hat{\theta}^+$ is used in place of $\hat{\theta}^*$ to correct for the additional bias term previously discussed.

A concern in inference when the variance of the MLE depends on the parameter estimates is how sensitive the approximate distribution is to this estimate. Figure 1 (top left) plots the histogram of the simulated distribution of $\{\hat{\theta}(\overline{Y}_1,\overline{Y}_2)|\overline{Y}_1\in S_s\}-E_s$ (solid line). In the same figure, for simulations that correspond to the 0.025, 0.50 and 0.975 quantiles of $\hat{\theta}$, histograms of the bootstrap sample distributions of $\{\hat{\theta}^+(\overline{Y}_1^*,\overline{Y}_2^*)|\overline{Y}_1^*\in\tilde{S}_s\}-\hat{\theta}$ (dotted, dashed and dot-dashed line) are plotted. This figure illustrates how well BM2 works across the domain of θ at approximating the distribution $\hat{\theta}(\overline{Y}_1,\overline{Y}_2)|\overline{Y}_1\in S_s$. Compare this to Fig. 1 (top right) which plots the PDF of $\mathcal{N}(0,\mathscr{F}_{\hat{\theta}}^{-1})$ for $\mathscr{F}_{\hat{\theta}}^{-1}$ evaluated at the same quantiles of $\hat{\theta}$. The bootstrap distribution is less sensitive to the value of $\hat{\theta}$ than $N(0,\mathscr{F}_{\hat{\theta}}^{-1})$ and it provides a better approximation to the shape of the target distribution.

The fifth row of Table 2 shows that with BM2 $P^*(\overline{Y}_1^*\in\tilde{S}_s)$ is nearly equal to p_s; $\mathrm{E}^*[\hat{\theta}^+(\overline{Y}_1^*,\overline{Y}_2^*)|\overline{Y}_1^*\in\tilde{S}_s]$ is approximately equal to E_s; $\mathrm{Var}^*[\hat{\theta}^+(\overline{Y}_1^*,\overline{Y}_2^*)|\overline{Y}_1^*\in\tilde{S}_s]$ is only slightly greater than V_s; and the confidence interval endpoints correspond to the quantiles of $\hat{\theta}$. It is clear that BM2 provides a more accurate representation of the conditional distribution of $\hat{\theta}$ than the alternatives. However, BM2 still fails to provide adequate coverage. This is not a problem with the bootstrap but rather a problem with the distribution of $\hat{\theta}$ being tightly centered around the wrong value.

Considering the poor coverage of BM2, we develop a bootstrap procedure for the bias corrected estimator $\tilde{\theta}$. Recall $\tilde{\theta} = \hat{\theta} - \lambda_k(\hat{\theta})$, $k = r, s$, is simply a function of $\hat{\theta}$. Thus BM2 can be adapted to estimate its distribution. Because the distribution of the bias is skewed, bias corrected bootstrap confidence intervals were used; see [3].

Bias Adjusted Conditional Bootstrap Method (BM3): Replace $\hat{\theta}^+$ in BM2 with $\tilde{\theta}^+ = \hat{\theta}^+ - \lambda_{\psi(\bar{y}_1)}(\hat{\theta}^+)$. Use $(C_l, C_u) = [\widehat{CDF}^{-1}\{\Phi(2v_0 + Z_{\alpha/2})\}, \widehat{CDF}^{-1}\{\Phi(2v_0 - Z_{\alpha/2})\}]$, where $\widehat{CDF}(t) = P^*[\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_{\psi(\bar{y}_1)} < t]$ and $v_0 = \Phi^{-1}\widehat{CDF}(\tilde{\theta})$.

Since $\tilde{\theta}$ is a function of $\hat{\theta}$, for comparison we approximate the variance of $\tilde{\theta}$ with $\Delta_{\tilde{\theta}} = [\{\partial(\theta - \lambda_{\psi(\bar{y}_1)}(\theta))/\partial\theta\}^2 \mathscr{I}_\theta^{-1}]_{\theta=\hat{\theta}}$. Figure 1 (bottom left) plots the histogram of the simulated distribution of $\{\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - \mathrm{E}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$ (solid line). In the same figure, for simulations that correspond to the 0.025, 0.50 and 0.975 quantiles of $\tilde{\theta}$, histograms of the bootstrap sample distributions of $\{\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \tilde{\theta}$ (dotted, dashed and dot-dashed line) are plotted. For comparison, Fig. 1 (bottom right) plots the probability density functions of a $\mathscr{N}(0, \Delta_{\tilde{\theta}})$ with $\Delta_{\tilde{\theta}}$ evaluated at the same quantiles of $\tilde{\theta}$. Once again we see that

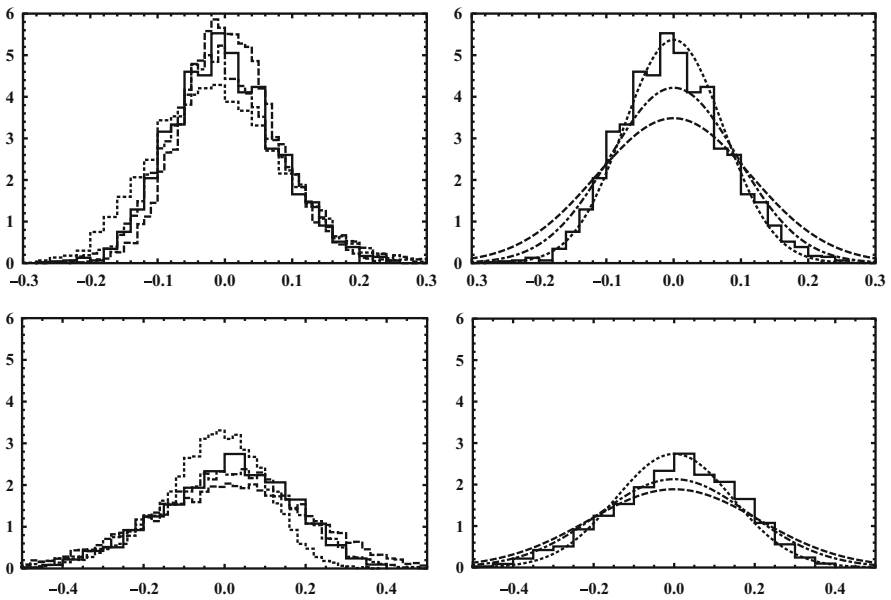

Fig. 1 Histograms of bootstrap distributions $\{\hat{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \hat{\theta}$ (*top left*) and $\{\tilde{\theta}^+(\overline{Y}_1^*, \overline{Y}_2^*)|\overline{Y}_1^* \in \tilde{S}_s\} - \tilde{\theta}$ (*bottom left*). Probability density functions of $N(0, \mathscr{I}_{\hat{\theta}}^{-1})$ (*top right*) and $N(0, \Delta_{\tilde{\theta}})$ (*bottom right*). The *dotted*, *dot dashed* and *dashed lines* correspond to the bootstrap distribution or expected information for the 0.025, 0.50, and 0.975 of quantiles of $\hat{\theta}$ or $\tilde{\theta}$. In each case the *solid line* is the histogram of the distribution of $\{\hat{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - E_s$ or $\{\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s\} - \mathrm{E}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$

Table 3 Mean square error (MSE) for the true distribution of $\hat{\theta}$ and $\tilde{\theta}$ along with the bootstrap MSE of $\hat{\theta}^+$ or $\tilde{\theta}^+$ for the case when $\overline{Y}_1 \in s$

Estimate	Method	MSE
$\hat{\theta}$	True	4.92
$\hat{\theta}^+$	BM2	4.98
$\tilde{\theta}$	True	5.25
$\tilde{\theta}^+$	BM3	5.71

the BM3 well approximates the distribution $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$ across most of the domain of $\tilde{\theta}$, perhaps with the exception of the 0.025 quantile.

The sixth row of Table 2 shows the simulation results the distribution of $\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s$. The seventh and eighth rows are results for $\Delta_{\tilde{\theta}}$ and BM3, respectively. BM3 provides an unbiased estimate of the mean; slightly overestimates the $\text{Var}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$; and provides confidence limits which coincide with the correct quantiles of $\tilde{\theta}$. Coverage is slightly less than the nominal level, but is a significant improvement over inference methods for $\hat{\theta}$. Note using $\Delta_{\tilde{\theta}}$ significantly overestimates $\text{Var}[\tilde{\theta}(\overline{Y}_1, \overline{Y}_2)|\overline{Y}_1 \in S_s]$ and slightly skews the confidence limits.

The performance of $\hat{\theta}^+$ and $\tilde{\theta}^+$ reflects the bias versus variance tradeoff. Table 3 compares the mean square error (MSE) for the simulated distribution of $\hat{\theta}$ and $\tilde{\theta}$ along with the bootstrap MSE of $\hat{\theta}^+$ and $\tilde{\theta}^+$ for the case when $\psi(\overline{y}_1) = s$. Despite the shortcomings of the procedure for $\hat{\theta}^+$, it still provides a lower MSE than $\tilde{\theta}^+$ in this example.

References

1. Baldi Antognini, A., Giovagnoli, A.: On the asympotic inference for response-adaptive experiments. METRON – Int. J. Stat. **49**, 29–46 (2006)
2. Dette, H., Bornkamp, B., Bretz, F.: On the efficiency of two-stage response-adaptive designs. Stat. Med. **32**, 1646–1660 (2013)
3. Efron, B.: Nonparametric standard errors and confidence intervals. Can. J. Stat. **9**, 139–158 (1981)
4. Efron, B., Hinkley, D.V.: Assessing the accuracy of the maximum likelihood estimate: observed versus expected Fisher information. Biometrika **65**, 457–483 (1978)
5. Lane, A., Yao, P., Flournoy, N.: Information in a two-stage adaptive optimal design. J. Stat. Plan. Inference **144**, 173–187 (2014)
6. Pronzato, L., Pazman, A.: Design of Experiments in Nonlinear Models. Springer, New York (2013)
7. Rosenberger, W.F., Hu, F.: Bootstrap methods for adaptive designs. Stat. Med. **18**, 1757–1767 (1999)
8. Whitehead, J.: On the bias of maximum likelihood estimation following a sequential test. Biometrika **73**, 573–581 (1986)

Study Designs for the Estimation of the Hill Parameter in Sigmoidal Response Models

Tobias Mielke

Abstract Sigmoidal models are frequently considered for the description of the dose-response relationship in dose-finding studies. Designs for these models depend on unknown parameters, which enter the model in a nonlinear way. A particular problem in frequentistic analysis using the sigmoid EMax model is the estimation of the Hill parameter, which may be problematic to estimate. The estimation problem for the Hill parameter and a model-based design approach to limit this problem will be examined in this paper.

1 Introduction

The sparse adoption of innovative design and analysis methods into dose finding practice served as a motivation for the EMA qualification opinion on the MCPMod approach for the model based design and analysis of dose-finding studies (CHMP [2]). Optimal design theory improves the information content of the study and may be used to evaluate the efficiency of study designs under model misspecification (Bretz et al. [1]). Dragalin et al. [4] describe the sigmoid EMax model as a flexible model in dose-finding, which might be used instead of examing multiple models with the full MCPMod approach. The sigmoid EMax model is considered as a reasonable model for dose-finding, having a number of different motivations and parameterizations (Goutelle et al. [6]). Thomas et al. [9] describe a meta analysis of dose-finding trials from a pharmaceutical company, which leads to the conclusion that EMax models describe the data of most dose-finding problems very well. Unfortunately, maximum likelihood estimators for sigmoidal models have similar limitations as in the binomial response model considered by Silvapulle [8]. Separated low-response and high-response dose-ranges may not allow the description of the increasing part of the dose response curve. The maximum likelihood estimator for the Hill parameter will then tend to high values, which in turn leads to computationally singular information matrices. Regulaziation techniques, Bayesian modelling and bootstrapping limit this problem related to sigmoidal models from

T. Mielke (✉)
ICON Clinical Research, Cologne, Germany
e-mail: tobias.mielke@iconplc.com

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_21

an analysis perspective. Alternatively, study designs may be optimized to improve the distribution of the Hill parameter estimator. The target of this publication is the examination of study designs for sigmoidal models, with an application to the sigmoid EMax model. In the second section, the statistical model will be introduced with the definition of the sigmoid EMax model and the definition of study designs. The estimation problem for the Hill parameter will be displayed using an example. A similar estimation problem will be examined in section three for piecewise linear models. These considerations will lead to design recommendations for sigmoidal models, which will be compared to standard design criteria using a simulation study for a sigmoid EMax model. Section five will close the article with a discussion of the results.

2 Response Model and Problem Setting

The average response of n_i subjects on the i-th dose group d_i is considered to be normally distributed

$$Y_i = \eta(\mathrm{d}_i, \theta) + \epsilon_i \sim \mathcal{N}\left(\eta(\mathrm{d}_i, \theta), \frac{1}{n_i}\sigma^2\right),\ i = 1, \ldots, G.$$

The error ϵ_i in the considered dose response model might include uncertainty on individual effects, which may be modeled via patient-wise varying intercepts. The dose-response relation is assumed to follow a sigmoid EMax model

$$\eta(\mathrm{d}_i, \theta) := \theta_1 + \theta_2 \times \frac{\mathrm{d}_i^{\theta_4}}{\theta_3^{\theta_4} + \mathrm{d}_i^{\theta_4}}. \tag{1}$$

Note, that the sigmoid EMax and the Logistic model coincide up to the scale of the dose and parameters

$$\eta(\exp(\mathrm{d}), \theta) = \theta_1 + \theta_2 \times \exp\left(\frac{\mathrm{d} - \log(\theta_3)}{\theta_4^{-1}}\right)\left\{1 + \exp\left(\frac{\mathrm{d} - \log(\theta_3)}{\theta_4^{-1}}\right)\right\}^{-1}.$$

The sigmoid EMax model is popular among practitioners and the derivation of optimum experimental designs for the sigmoid EMax model is as simple as for the logistic model, such that this paper will focus the sigmoid EMax model in the following.

The parameter θ_1 in the sigmoid EMax model describes the placebo response, whereas θ_2 describes the maximum asymptotic effect above placebo. The parameter θ_3 specifies the location of the dose with 50 % of the maximum asymptotic effect above placebo and is referred to as the ED_{50}-parameter. The Hill parameter is given by $\theta_4 > 0$ and defines the steepness of the dose-response curve. Values of θ_4 below

1 lead to a high increase for low doses, whereas values of θ_4 above 1 lead to the characteristic sigmoidal shape, which gets steeper with increasing values of θ_4. The standard EMax model is a special case of the sigmoid EMax model with $\theta_4 = 1$.

Approximate study designs on a discrete dose range are considered, targeting the relative allocation of patients to G admissible dose groups

$$\mathbf{x} := \begin{pmatrix} \mathrm{d}_1 & \dots & \mathrm{d}_G \\ \omega_1 & \dots & \omega_G \end{pmatrix}, \ \sum_{i=1}^{G} \omega_i = 1.$$

A generalization of the results to a continuous dose range $\mathscr{X}$ using available optimum design algorithms is straight forward (e.g. Fedorov and Leonov [5]). Given an appropriate study design $\mathbf{x}$, the asymptotic normality of the ML-estimator is generally considered

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, \mathbf{M}_\theta(\mathbf{x})^{-1}\right), \ N \to \infty,$$

where N is the total sample size and $\mathbf{M}_\theta(\mathbf{x})$ denotes the Fisher Information matrix

$$\mathbf{M}_\theta(\mathbf{x}) := \frac{1}{\sigma^2} \sum_{i=1}^{G} \omega_i \frac{\partial \eta(\mathrm{d}_i, \theta)}{\partial \theta} \frac{\partial \eta(\mathrm{d}_i, \theta)}{\partial \theta^\top}.$$

Due to the nonlinear dependence of the sigmoid EMax model on the shape parameters, the information matrix will depend on the unknown parameters.

We consider as an example a dose-finding study with equal randomization into 4 active dose groups and a placebo group. A simulated dataset is displayed on the left-hand side of Fig. 1. Observations were simulated with a standard deviation of $\sigma = 100$ and a maximum effect of 100 in the maximum dose group. A high standard deviation in comparison to the effect size is not uncommon in early phases of drug development.

The situation displayed in Fig. 1 demonstrates the typical problem with sigmoidal model fits. The best fitting curve is close to a step function, where the increase crosses the observation at dose 50. This common problem may lead for an unbounded parameter space to missing likelihood estimates due to computationally singular matrices. Note that the value of the likelihood remains almost unchanged when increasing the value of the Hill parameter from $\theta_4 = 20$ to $\theta_4 = 30$, given a value of $\theta_3 = 45$. The likelihood-function is here very flat around the maximum.

The simulated data were based on a dose finding study with $N = 200$ patients. Data were simulated from a sigmoid EMax model with true ED_{50} at 50 and Hill parameter $\theta_4 = 4$ using the R package DoseFinding. Parameters were constrained to be located within the range of 0.1 and 150 for the ED_{50} and 0.1 and 60 for the Hill parameter. The joint distribution of the ED_{50} and Hill parameter estimator is displayed on the right-hand side of Fig. 1. Given the scatter plot, the estimates of the ED_{50} are clustered around the values 50 and 25. These clusters are generated by the

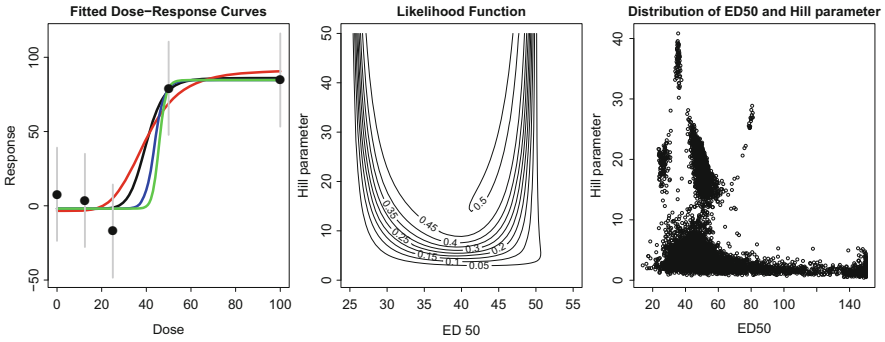

Fig. 1 *Left-hand side*: Simulated data and fitted models with different constraints on the Hill parameter ($\theta_4 \leq 5,\ 10,\ 20,\ 30$) and corresponding likelihood function. *Right-hand side*: Joint distribution of the ED_{50} and Hill parameter estimates for a design with 200 subjects

design and the resulting underlying distribution and are not related to the nonlinear least squares algorithm. Each of the points in the two clusters represents a situation as displayed on the left-hand side of Fig. 1. Moving the examined doses higher, e.g. 25–35 and 50–60 will also move the location of the clusters to 35 and 60.

The probability of running into the encountered problem depends on the sample size. The estimator of the Hill parameter will exceed the value 10 with simulated 39 % probability, when using the considered study design with N = 200. Increasing the sample size does little to improve this probability (33 % with N = 500 and 25 % with N = 1000). These simulated probabilities differ from the values resulting with the normal approximation (12 %, 3 %, 0.5 %, for N = 200, 500, 1000). The problem is not specific to the sigmoid EMax model. The same behaviour may be observed for the estimator corresponding to θ_4 in the Logistic model. Given the situation described in Fig. 1, it is obvious that the selection of study doses is of high importance when using the sigmoidal models in dose finding.

3 Design Considerations

The Hill parameter defines together with the ED_{50} the increasing part of the dose-response curve. Miller et al. [7] describe designs for the examination of an interesting part of the dose response curve in a similar model. We will here focus on another range of interest, namely the doses which provide 25–75 % of the maximum response.

The estimation problem in the sigmoid EMax model is related to a flat likelihood function around the maximum. Design criteria target, via the Fisher information matrix $\mathbf{M}_\theta$, the expected curvature of the likelihood function around the true parameters and are hence reasonable for the problem at hand. However, optimal designs are generally constructed under the assumption of asymptotic normality.

As Fig. 1 displays, the distribution of the parameter estimator might not be close to a normal distribution for limited sample sizes. In the following, a piecewise linear model will be examined as an approximation to general sigmoidal shapes. The estimation problem may be well described using this piecewise linear model. This problem description in the simple model will support the definition of study designs improving the probability of fitting success for nonlinear sigmoidal models.

3.1 Designs for the Piecewise Linear Function

Consider as an approximation to the sigmoid EMax model the piecewise linear function, which is defined by

$$\eta_1(x,\theta) := \begin{cases} \theta_1 & x \le \theta_3 \\ \theta_1 + \theta_2(x-\theta_3) & \theta_3 < x < \theta_4 \\ \theta_1 + \theta_2(\theta_4-\theta_3) & x \ge \theta_4 \end{cases}.$$

Unknown parameters for η_1 are the changepoints θ_3 and θ_4, the mean response on the low range θ_1 and the slope θ_2 of the increasing part. Fitting the piecewise linear model to the data is as problematic as fitting a sigmoidal model to the data, with the slight difference that the ML-estimator might not be well-defined. See Fig. 2 for an example on a variety of shapes described by model η_1, which will all fit the data equally well. The situation in Fig. 2 leads also for the sigmoid EMax model to a problematic model fit. Any curve with high increase between the doses will lead to almost the same prediction value at the examined doses. For simplicity, we will consider balanced designs on 4 doses in the following.

Remark 1 Consider a balanced allocation to 4 support points $x_1 = a, x_4 = b$ and $x_2, x_3 \in (a,b)$ with average responses $y_1, \ldots, y_4$ and a common variance. The least squares-estimator for the changepoints θ_3 and θ_4 of $\eta_1(x;\theta)$ is well-defined on the interval (a,b) if, and only if,

(i) Responses are strictly monotone increasing (decreasing) and
(ii) $y_2(x_3-x_1) - y_3(x_2-x_1) - y_1(x_3-x_2) < 0\ (>0)$ and
(iii) $y_3(x_4-x_2) - y_4(x_3-x_2) - y_2(x_4-x_3) > 0\ (<0)$.

This remark may be proven easily for the direction from the conditions to the well-defined estimator. The other direction is more tedious. The main part of the proof from the uniqueness to conditions (i), (ii) and (iii) is based on the unique representation of the increasing part. This leads to the existence of the two changepoints, which need to be located within the intervals (x_1, x_2) and (x_3,x_4). Given this result and an assumed direction for the slope, condition (i) on monotone responses follows. The results on conditions (ii) and (iii) are proven using the well-defined changepoints and considerations on the best fit for the monotone data. The

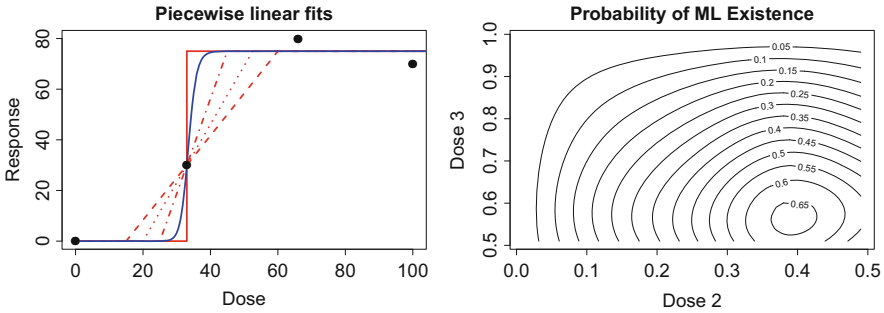

Fig. 2 *Left-hand side*: Multiple piecewise model fit with the same likelihood and sigmoid EMax fit. *Right-hand side*: Probability for a well-defined MLE for different support points of the design

unique estimator in the remark is given by

$$\widehat{\theta}_1 := y_1,\ \widehat{\theta}_2 := \frac{y_3 - y_2}{x_3 - x_2},\ \widehat{\theta}_3 := x_2 + \frac{y_1 - y_2}{\hat{\theta}_2},\ \widehat{\theta}_4 := x_2 + \frac{y_4 - y_2}{\hat{\theta}_2}.$$

The increasing part of η_1 is defined by the two points (x_2,y_2) and (x_3,y_3). The constant parts are on the levels y_1 and y_4, such that the sum of squares is zero.

Given the equivalence in Remark 1, support points of the study design may be optimized to target via conditions (i), (ii) and (iii) a maximized probability for a unique least squares estimator. An example is given in the right-hand side of Fig. 2. The underlying shape assumption is a sigmoid EMax model with ED_{50} at 50 and Hill parameter 4, as we target the design optimization for the sigmoid EMax model. The dose with 33 % of the maximum effect is given for this model at 41, whereas the dose with 66 % of the maximum effect is given at 57. The maximum probability (66 %) for a unique least squares estimator is attained when allocating patients to doses 40 and 57. The influence of the design on the uniqueness probability is relatively high. A balanced allocation to 0, 33, 66 and 100 will attain only a probability of 54 %.

3.2 Design Criteria for the Sigmoid EMax Model

We will compare the changepoint design (**CP**) with standard design criteria Φ, defined as

$$\Phi_{D;S;\theta}(\mathbf{x}) := \log |S\mathbf{M}_\theta(\mathbf{x})^{-1}S^\top| \text{ and } \Phi_{L;\theta}(\mathbf{x}) := \text{tr}\, L\mathbf{M}_\theta(\mathbf{x})^{-1}.$$

D-optimality ($S = I_4$, **D**) is of interest in dose-finding given the implications on the MSE due to the equivalence with G-optimality. Given the problem of the Hill parameter estimation, a focus on the increasing part of the curve using D_S optimality for θ_3 and θ_4 will be considered ($\mathbf{D_S}$). Linear criteria for the minimization of the

variance of a certain parameter (e.g. θ_4, ($\mathbf{L}_{,,4}$)) or for the optimization of the dose estimation will be studied. Asymptotic normality assumptions together with the Δ-method lead to a matrix L, which defines the criterion of interest. The ED_p-optimality is derived with the use of the Δ-method (Dette et al. [3]). The ED_p is defined as the dose leading to $p\%$ of the maximum effect on the examined dose range. An ED_{50} optimal design shall hence minimize the variance on the ED_{50} estimation ($\mathbf{ED_{50}}$). Only locally optimal designs will be considered. We will examine two versions of the criteria to verify, whether designs should be optimized under the true parameter or using the expected parameter values for the problematic Hill parameter case.

4 Design Evaluation

The simulation and design model is a sigmoid EMax model as given in formula 1, where the ED_{50} parameter is given by 50 and the Hill parameter is set to 4. We consider a placebo effect of 0 and a maximum effect of 100. The standard deviation is given by $\sigma = 100$. The design will be evaluated for a total sample size of $N = 200$ subjects. Designs were optimized on the discrete dose-range of integers in [0, 100]. The average MSE on the full dose range (*MSE*), the average MSE on the range from the true ED_{25} to the true ED_{75} ($MSE_{25,75}$), the probability that the ED_{50}-estimator is located between the true ED_{25} and the true ED_{75} ($P(ED_{50})$) and the probability that the estimator of the Hill parameter is below 10 ($P(\hat{\theta}_4 < 10)$) were the studied operating characteristics. Given the simulation results in Table 1, the design considerations using the approach via the piecewise linear **CP** are promising. The design for the piecewise linear model targets doses the 33 % and 66 % effect quantiles. This will increase robustness in the estimation of the Hill parameter as compared to the locally optimal designs with the true Hill parameter of $\theta_4 = 4$. The $\mathbf{D_S}$-optimal design is the most promising optimality-criterion, although the performance of **CP** for the Hill parameter is not reached. Note that the ED_{50}-optimal

Table 1 Design characteristics for locally optimal designs with $\theta_4 = 4$ and $\theta_4 = 10$

$Design(\boldsymbol{\theta}_{4=4})$	$L_{\boldsymbol{\theta}_4}$	ED_{50}	D_S	D	CP
$P(\hat{\theta}_4 < 10)$	0.697	0.526	0.787	0.755	0.825
$P(ED_{50})$	0.779	0.795	0.892	0.857	0.893
MSE	245.52	332.20	198.82	212.31	185.81
$MSE_{25,75}$	339.17	534.16	253.38	371.45	215.25
$Design(\boldsymbol{\theta}_{4=10})$	$L_{\boldsymbol{\theta}_4}$	ED_{50}	D_S	D	CP
$P(\hat{\theta}_4 < 10)$	0.829	0.508	0.806	0.800	0.832
$P(ED_{50})$	0.889	0.515	0.830	0.849	0.910
MSE	214.27	340.09	212.65	207.78	190.58
$MSE_{25,75}$	158.40	509.09	175.69	194.02	224.02

design were very weak in getting the ED_{50} right. Allocation for the ED_{50}-optimal design is concentrated on the doses around the true ED_{50}, which will limit the accuracy on the whole dose range.

5 Discussion

Problems in the analysis of dose-finding studies with the sigmoidal models were examined from the design perspective. The estimator for the Hill parameter in the sigmoid EMax model will generally not follow a normal distribution. Design considerations using piecewise linear models may help to construct robust designs for Hill parameter estimation. Alternatively, locally optimal designs on extreme values for the model parameters may add value to the estimation of the Hill parameter, as studied in Sect. 4. Additional considerations will help to examine the robustness of the proposed approach under model-uncertainty.

Acknowledgements The author is grateful for very valuable discussions and input on this topic from Bjoern Bornkamp, Rainer Schwabe, Vladimir Dragalin and Valerii Fedorov.

References

1. Bretz, F., Dette, H., Pinheiro, J.: Practical considerations for optimal designs in clinical dose finding studies. Stat. Med. **29**, 731–742 (2010)
2. Committee for Medicinal Products for Human Use (CHMP): Qualification opinion of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty. European Medicines Agency, EMA/CHMP/SAWP/757052/2013 (2014)
3. Dette, H., Kiss, C., Bevanda, M.: Optimal designs for the Emax, log-linear and exponential models. Biometrika **97**, 513–518 (2009)
4. Dragalin, V., Hsuan, F., Padmanabhan, S.K.: Adaptive designs for dose finding studies based on sigmoid EMax model. J. Biopharm. Stat. **17**, 1051–1070 (2007)
5. Fedorov, V.V., Leonov, S.L.: Optimal Design for Nonlinear Response Models. Chapman & Hall/CRC Biostatistics Series. CRC Press/Taylor & Francis Group, Boca Raton (2013)
6. Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M., Maire, P.: The Hill equation: a review of its capabilities in pharmacological modelling. Fundam. Clin. Pharmacol. **22**, 633–648 (2008)
7. Miller, F., Guilbaud, O., Dette, H.: Optimal designs for estimating the interesting part of a dose-effect curve. J. Biopharm. Stat. **17**, 1097–1115 (2007)
8. Silvapulle, M.: On the existence of maximum likelihood estimators for the binomial response models. J. R. Stat. Soc. **43**, 310–313 (1981)
9. Thomas, N., Sweeney, K., Somayaji, V.: Meta-analysis of clinical dose-response in a large drug development portfolio. Stat. Biopharm. Res. **6**, 302–317 (2014)

Controlled Versus "Random" Experiments: A Principle

Werner G. Müller and Henry P. Wynn

Abstract The contrast and tension between controlled experiment and passive observation is an old area of debate to which philosophers of science have made contributions. This paper is a discussion of the issue in the context of modern Bayesian optimal experimental design. It is shown with simple examples that a mixture of controlled and less controlled experiments can be optimal, and this is stated as a general principle. There is a short discussion of a wider theory in the last section.

1 The Principle

The distinction between controlled experiment and passive observation is a long-standing theme in the history of experimentation and the philosophy of science. Francis Bacon [1] and John Stuart Mill [8] emphasize controlled experimentation. Bacon classifies experiments into different types such as the experimentum cruxis (experiment at the cross roads) and Mill distinguishes "artificial experiments" and "spontaneous" experiments. Claude Bernard [2] writes at length on the distinction.

This distinction has become important in a type of product design, promoted by the Japanese engineer, Genichi Taguchi and sometimes now given the generic term robust engineering design (RED), cf. [11]. The idea is that experiments on product prototypes need to set different levels not just of "design factors" but also "noise factors", which may affect the product in manufacturing or actual use.

In medicine, economics and the social sciences the randomized trial attempts to guard against external factors, which may bias the inference. In the literature on the methodology in (experimental) economics there is real concern about the distinction (cf. e.g. [6]), probably greater than in the natural sciences, because of the difficulty of describing and controlling the populations and subpopulations about

W.G. Müller (✉)
Department of Applied Statistics, Johannes Kepler University, Linz, Austria
e-mail: werner.mueller@jku.at

H.P. Wynn
Department of Statistics, London School of Economics, London, UK
e-mail: h.wynn@lse.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_22

which inference is made. This may affect the portability of models to unobserved populations. Note that despite its relevance the topic of experimental design has gained little attention in the field of behavioral or experimental economics, see [7] and [4] for exceptions.

The above issue has become important in the study of causation, for example when a controllable (intervention) variable X affects a variable Y, but where there is an intermediate (mediation) variable M. In that case one may simply investigate the effect of X on Y, in some sense ignoring M. For example one may endeavour to randomize in such a way to obtain a sample in which there is a representative population of M-values, without actually measuring M. Alternatively, or in addition, one may bring M into the experiment, keep the X values fixed but control the M values to an experimental range. It may even be the case that in a laboratory setting a controlled experiment may use a wider range of values of the independent variable than is found in a less controlled environment; accelerated life-testing is an example. In modern discussions of causation the distinction colours debate about whether elucidation of causation requires intervention (cf. e.g. [10]).

The aim of this note is simply stated in terms of establishing a principle. We will not describe fully where the principle is to be applied, but we will certainly show in some basic examples that it holds. The principle is that

a mixture of a controlled and a passive experiment may be optimal.

The argument for the principle can be given in a heuristic way, and the examples above may be useful in this regard. There may be some key parameters in a system which are well estimated with a controlled experiment, but if the model is to operate in, be portable to, a wider environment then part of the experiment should favour the estimation of parameters relevant to that environment. If the principle is accepted then discussions of whether controlled experiments are better than more passive observation should include assessment of the *extent* to which experiments should reflect the operating environment. The issue is well recognized in mainstream statistics modelling via the distinction between fixed, random effect and mixed models (see e.g. [9]). However, the distinction is less well analyzed in optimal experimental design. We will adopt a Bayesian optimal experimental design framework, see [3] for fundamentals.

2 A Simple Example

We consider an example in which $Y = (Y_1, \dots Y_{n_1})$ is a vector of observations from what we will call the controlled experiment and $Z = (Z_1, \dots, Z_{n_2})$ is a vector of observations from what we will call the random experiment. We will then investigate the relative size of n_1 and n_2 satisfying some optimum allocation problem. There are two parameters θ and ϕ and, following a Bayesian approach, these will be taken as random, with independent prior distributions. There will also be random errors of observation $\{\epsilon_i\}$.

The model is

$$Y_i = \theta + \epsilon_i, \quad i = 1, \ldots, n_1, \tag{1}$$

$$Z_j = \theta + \phi + \delta_j, \quad j = n_1 + 1, \ldots, n_1 + n_2. \tag{2}$$

We assume that $\{\epsilon_1, \ldots, \epsilon_{n_1}\}$ are iid $N(0, \sigma_1^2)$, $\{\delta_{n_1+1}, \ldots, \delta_{n_1+n_2}\}$ are iid $N(0, \sigma_2^2)$ and that $\theta \sim N(\mu_1, \tau_1^2)$ and $\phi \sim N(\mu_2, \tau_2^2)$, with $\sigma_1, \sigma_2, \mu_1, \mu_2, \tau_1, \tau_2$ all known. We also assume that the $\{\epsilon_i, \delta_j\}$, θ and ϕ are all independent.

The interpretation of this model is that equation (1) is a system which applies in a laboratory setting where we can simply try to measure θ. Equation (2) represents some kind of less controllable operational environment with an additional random component ϕ.

It is assumed that our inference concerns θ and ϕ. The posterior analysis is straightforward. We first write down the covariance matrix between the two sample means, $\overline{Y}$ and $\overline{Z}$, which are jointly sufficient statistics, and the parameters:

$$\operatorname{cov}\left(\overline{Y}, \overline{Z}, \theta, \phi\right) = \begin{pmatrix} \frac{\sigma_1^2}{n_1} + \tau_1^2 & \tau_1^2 & \tau_1^2 & 0 \\ \tau_1^2 & \frac{\sigma_2^2}{n_2} + \tau_1^2 + \tau_2^2 & \tau_1^2 & \tau_2^2 \\ \tau_1^2 & \tau_1^2 & \tau_1^2 & 0 \\ 0 & \tau_2^2 & 0 & \tau_2^2 \end{pmatrix}. \tag{3}$$

The posterior (conditional) covariance matrix of θ, ϕ is

$$\begin{pmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{pmatrix} - \begin{pmatrix} \tau_1^2 & \tau_1^2 \\ 0 & \tau_2^2 \end{pmatrix} \begin{pmatrix} \frac{\sigma_1^2}{n_1} + \tau_1^2 & \tau_1^2 \\ \tau_1^2 & \frac{\sigma_2^2}{n_2} + \tau_1^2 + \tau_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \tau_1^2 & 0 \\ \tau_1^2 & \tau_2^2 \end{pmatrix}. \tag{4}$$

We consider the Bayesian D-optimum case or, equivalently in this Gaussian example, the Shannon information case. This means we need to minimize, with respect to the choice of n_1, n_2, the following quantity:

$$\det\{\operatorname{cov}(\theta, \phi | \overline{Y}, \overline{Z})\} = \frac{\tau_1^2 \tau_2^2 \sigma_1^2 \sigma_2^2}{n_1 n_2 \tau_1^2 \tau_2^2 + n_1 \sigma_2^2 \tau_1^2 + n_2 \sigma_1^2 \tau_1^2 + n_2 \sigma_1^2 \tau_2^2 + \sigma_1^2 \sigma_2^2}. \tag{5}$$

To set this up as an optimisation problem we assume that every observation has the same cost and the total sample size to be fixed at $n = n_1 + n_2$. We rewrite this, ignoring the discreteness as $n_1 = pn$, $n_2 = (1 - p)n$, where $0 \leq p \leq 1$, and p represents the proportion of experimental effort on experiment (1).

The inverse of the criterion $\det\{\operatorname{cov}(\theta, \phi | \overline{Y}, \overline{Z})\}$ is quadratic in p, with a maximum at

$$p^* = \frac{1}{2}(1 - R),$$

where

$$R = \frac{1}{n}\left(\frac{\sigma_1^2}{\tau_1^2} + \frac{\sigma_1^2}{\tau_2^2} - \frac{\sigma_2^2}{\tau_2^2}\right).$$

There are various situations, dependent on the values of p^*:

1. $R \geq 1$ implies all controlled : $n_1 = 0,\ n_2 = n$
2. $R \leq -1$ implies all "random" : $n_1 = n,\ n_2 = 0$
3. $-1 \leq R \leq 1$ implies the mixture principle : $0 < n_1, n_2 < 1$
4. $R \to 0$ implies a mixture with $p \to \frac{n}{2}$

The distinction between random and controlled experiment is seen to be dependent on a single parameter R involving the balance between (subjective) prior knowledge and sample noise expressed via $\frac{\sigma_1^2}{\tau_1^2}$, $\frac{\sigma_2^2}{\tau_2^2}$ and $\frac{\sigma_1^2}{\tau_2^2}$ which could be called signal to noise ratios. In terms of optimal design this is a very restricted problem because once the base experiments have been set the optimization is only in terms of a single mixing parameter. This mixing is familiar in the optimal design of experiments as will, perhaps, be clearer from the next section.

3 The General Regression Case

We extend the material in Sect. 2 to the general linear model case. Let θ and ϕ be r- and s-dimensional vectors, respectively. Let $f_i^{(1)},\ i = 1, \ldots, n_1, f_i^{(2)},\ i = 1, \ldots, n_2$, be r-vectors and $f_i^{(3)},\ i = 1, \ldots, n_2$, s-vectors of independent variables. The model is

$$Y_i = \theta^T f_i^{(1)} + \epsilon_i,\ \ i = 1, \ldots, n_1 \tag{6}$$

$$Z_j = \theta^T f_j^{(2)} + \phi^T f_j^{(3)} + \delta_j,\ \ j = n_1 + 1, \ldots, n_1 + n_2, \tag{7}$$

where the assumptions on the ϵ_i and δ_j are as before but now $\theta \sim N(0, \Sigma_1)$ and $\phi \sim N(0, \Sigma_2)$, and θ, ϕ and the $\{\epsilon_i, \delta_j\}$ are all independent. With obvious notation the matrix version is

$$Y = F_1\theta + \epsilon \tag{8}$$

$$Z = F_2\theta + F_3\phi + \delta. \tag{9}$$

With the notation $C(U)$ and $C(U|V)$ for covariance matrices we have the determinental result

$$\det\{C(U, V)\} = \det\{C(U)\}\det(V|U). \tag{10}$$

If we take logarithms we have, except for constants the familiar addition property of Shannon information. Expanding now in two ways we have:

$$\det\{C(Y,Z,\theta,\phi)\} = \det\{C(\theta,\phi)\}\det\{C(Y,Z|\theta,\phi)\},$$
$$\det\{C(Y,Z,\theta,\phi)\} = \det\{C(Y,Z)\}\det\{C(\theta,\phi|Y,Z)\},$$

giving

$$\det\{C(\theta,\phi|Y,Z)\} = \frac{\det\{C(\theta,\phi)\}\det\{C(Y,Z|\theta,\phi)\}}{\det\{C(Y,Z)\}}. \tag{11}$$

The marginal covariance matrix of Y,Z is

$$C(Y,Z) = \begin{pmatrix} F_1\Sigma_1F_1^T + \sigma_1^2 I_{n_1} & F_1\Sigma_1F_2^T \\ F_2\Sigma_1F_1^T & F_2\Sigma_1F_2^T + F_3\Sigma_2F_3^T + \sigma_2^2 I_{n_2} \end{pmatrix}.$$

As we saw in the last section it is easier to work with the inverse determinant. Using a standard determinental identity we have

$$\det\{C(\theta,\phi|Y,Z)\}^{-1} = \det\left\{ \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & \Sigma_2^{-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{\sigma_1^2}F_1^TF_1 + \frac{1}{\sigma_2^2}F_2^TF_2 & \frac{1}{\sigma_2^2}F_2^TF_3 \\ \frac{1}{\sigma_2^2}F_3^TF_2 & \frac{1}{\sigma_2^2}F_3^TF_3 \end{pmatrix} \right\}.$$

We can appeal to the familiar principle of continuous optimal design by considering the normalized moment matrices $M_{ij} = \frac{1}{n}F_i^TF_j$. Then we introduce the proportion p as in the last section and seek to maximize, with respect to p:

$$\det\{C(\theta,\phi|Y,Z)\}^{-1} = \det\left\{ \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & \Sigma_2^{-1} \end{pmatrix} + \begin{pmatrix} \frac{pn}{\sigma_1^2}M_{11} + \frac{(1-p)n}{\sigma_2^2}M_{22} & \frac{(1-p)n}{\sigma_2^2}M_{23} \\ \frac{(1-p)n}{\sigma_2^2}M_{32} & \frac{(1-p)n}{\sigma_2^2}M_{33} \end{pmatrix} \right\}.$$

We see that $\psi(p) = \log[\det\{C(\theta,\phi|Y,Z)\}^{-1}]$ is concave in p. Maximising over p is, as explained, a simplified optimal design problem. But we are first interested in establishing conditions for mixing. Using concavity we have three cases which depend on the derivative, $\psi'(p)$, of $\psi(p)$:

1. $\psi'(0) \le 0,\ \psi'(1) \le 0 \Rightarrow n_1 = n, n_2 = 0$
2. $\psi'(0) \ge 0,\ \psi'(1) \ge 0 \Rightarrow n_1 = 0, n_2 = n$
3. $\psi'(0) \ge 0,\ \psi'(1) \le 0 \Rightarrow 0 < n_1, n_2 < n,$ (mixing)

Using the standard formula for $\frac{\partial}{\partial p}\log(\det)$ we can compute

$$\psi'(0) = \mathrm{tr}\left\{ \left(\Sigma^{-1} + \frac{n}{\sigma_2^2}M_2 \right)^{-1} \left(\frac{n}{\sigma_1^2}M_1 - \frac{n}{\sigma_2^2}M_2 \right) \right\},$$

$$\psi'(1) = \mathrm{tr}\left\{\left(\Sigma^{-1} + \frac{n}{\sigma_1^2}M_1\right)^{-1}\left(\frac{n}{\sigma_1^2}M_1 - \frac{n}{\sigma_2^2}M_2\right)\right\},$$

where $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}, M_1 = \begin{pmatrix} M_{11} & 0 \\ 0 & 0 \end{pmatrix}, M_2 = \begin{pmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{pmatrix}$.

The quantity

$$p^* = \arg\max_p \psi(p)$$

can also be used to establish the conditions, as in the last section, but there is in general no closed form expression for p^*. However, with considerable additional assumptions we can give a closed form. This is similar to classical optimal design where we can find explicit solutions to optimum design problems in highly structured cases.

Thus, let (i) $\Sigma_1 = \tau_1^2 I_r,\ \Sigma_2 = \tau_2^2 I_s,\ (ii)\ M_{11} = I_r, M_{22} = I_s, M_{23} = M_{32} = 0$. In terms of experimental design conditions (ii) can be thought of as orthogonality conditions. Then

$$\det\{C(\theta, \phi|Y, Z)\}^{-1} = \left(\frac{1}{\tau_1^2} + \frac{pn}{\sigma_1^2}\right)^r \left(\frac{1}{\tau_2^2} + \frac{(1-p)n}{\sigma_2^2}\right)^s.$$

Taking the logarithm, differentiating with respect to p and solving we obtain:

$$p^* = \frac{r}{r+s}\left(1 - \tilde{R}\right),$$

where

$$\tilde{R} = \frac{1}{n}\left(\frac{\sigma_1^2}{\tau_1^2} - \frac{\sigma_2^2}{\tau_2^2}\right).$$

We have conditions rather similar to the last case, but not quite a generalisation because $M_{23} = M_{32} = 0$ did not hold there. These conditions are:

1. $\tilde{R} \geq 1$ implies all controlled : $n_1 = 0,\ n_2 = n$
2. $\tilde{R} \leq -\frac{s}{r}$ implies all "random" : $n_1 = n,\ n_2 = 0$
3. $-\frac{s}{r} \leq \tilde{R} \leq 1$ implies the mixture principle : $0 < n_1, n_2 < 1$
4. $\tilde{R} \to 0$ implies a mixture with $p \to \frac{r}{r+s}$

Although this is a considerable simplification it reveals the role played by the model dimensions r and s (which can be thought of as model degrees of freedom), the sample size and two simple signal to noise ratios. We trust this may help with intuition in more sophisticated problems.

4 Future Research

In [5] the general problem of learning in the context of design of experiments is discussed. It is seen that the notion of a convex information functional on prior distributions provides a background to the choice of the objective function. It is pointed out that an experiment can be considered as a choice of *sampling distribution*. The convexity discussed here is convexity over this choice.

Thus a generic version of the current work would consider a parameter θ with prior distribution $\pi(\theta)$ and two fixed base experiments with multivariate sampling distributions $f_1(y_1|\theta)$ and $f_2(y_2|\theta)$, respectively. The base experiments would be replicated independently with sample sizes n_1 and n_2 respectively.

If $\mathbf{y_1}$ and $\mathbf{y_2}$ represent the fully replicated experiments and $\phi(\cdot)$ a suitable information functional then:

$$\psi(n_1, n_2) = \mathrm{E}_{\mathbf{Y_1},\mathbf{Y_2}} \mathrm{E}_{\theta|\mathbf{Y_1},\mathbf{Y_2}} \phi\{\pi(\theta|\mathbf{Y_1}, \mathbf{Y_2})\},$$

would be the preposterior expected information. Then with our substitution $n_1 = pn, n_2 = (1-p)n$ we seek to show concavity in the proportion p. Some progress has been made with Shannon information when $\phi(\cdot)$ is $\log(\cdot)$.

With a general result of this kind one could cover the current case when one experiment is more controlled than another. But by induction one could also cover mixtures of many different experiments, for example when there are more complex intervention strategies in the medical and social sciences. In such cases, we conjecture the derivation of special formulae based on sums and differences of signal to noise ratios, as in our examples, provided the individual component experiments are simple.

References

1. Bacon, F.: Novum organum scientiarum (1620). In: Jardine, L., Silverthorne, M. (eds.) Cambridge Philosophical Texts (2000)
2. Bernard, C.: Introduction à l'étude de la médecine expérimentale, Paris (1865). English edition, Dover, New York (1957)
3. Chaloner, K., Verdinelli, I.: Bayesian experimental design: a review. Stat. Sci. **10**, 273–304 (1995)
4. El-Gamal, M.A., Palfrey, T.R.: Economical experiments: Bayesian efficient experimental design. Int. J. Game Theory **25**, 495–517 (1996)
5. Hainy, M., Müller, W.G., Wynn, H.P.: Learning functions and approximate Bayesian computation design: ABCD. Entropy **16**, 4353–4374 (2014)
6. Kasy, M.: Why experimenters should not randomize, and what they should do instead. Technical report 36154, Harvard University OpenScholar Working Paper (2013)
7. List, J.A., Sadoff, S., Wagner, M.: So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. Exp. Econ. **14**, 439–457 (2011)
8. Mill, J.S.: A System of Logic. John W. Parker, West Strand, London (1843)

9. McLean, R.A., Sanders, W.L., Stroup, W.W.: A unified approach to mixed linear models. Am. Stat. **45**, 54–64 (1991)
10. Pearl, J.: Causality: Models, Reasoning and Inference, 2nd edn. Cambridge University Press, Cambridge (2009)
11. Taguchi, G.: Taguchi on Robust Technology Development: Bringing Quality Engineering Upstream. ASME Press (American Society of Mechanical Engineers), New York (1992)

Adaptive Designs for Optimizing Online Advertisement Campaigns

Andrey Pepelyshev, Yuri Staroselskiy, and Anatoly Zhigljavsky

Abstract We investigate the problem of adaptive targeting for real-time bidding in online advertisement using independent advertisement exchanges. This is a problem of making decisions based on information extracted from large data sets related to previous experience. We describe an adaptive strategy for optimizing the click through rate which is a key criterion used by advertising platforms to measure the efficiency of an advertisement campaign. We also provide some results of statistical analysis of real data.

1 Introduction

Online advertisement is a growing area of marketing where advertisements can be personalized depending on user's behaviour. To determine user preferences, advertising platforms record data with visited webpages, previous impressions (i.e. ads shown), clicks, conversions, geographical information derived from IP address and then use these data to design strategies when, where and to whom to show some advertisements. Online advertisement has two main forms: one is related to leading technology companies like Google and another is processed by independent ad exchanges [13].

Ad exchanges use auctions with Real-Time Bidding (RTB), which is a magnificent way of delivering online advertising. As mentioned in [3], spending on RTB in the US during 2014 reached $10 billion. The participants of auctions are demand

A. Pepelyshev (✉)
Cardiff University, Sengennydd Road, Cardiff, UK
e-mail: pepelyshevan@cardiff.ac.uk

Y. Staroselskiy
Crimtan, 1 Castle Lane, London, SW1E 6DR, UK
e-mail: yuri@crimtan.com

A. Zhigljavsky
Cardiff University, Sengennydd Road, Cardiff, UK
Lobachevskii State University of Nizhnii Novgorod, Gagarin av. 23, 603950, Nizhny Novgorod, Russia
e-mail: ZhigljavskyAA@cardiff.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_23

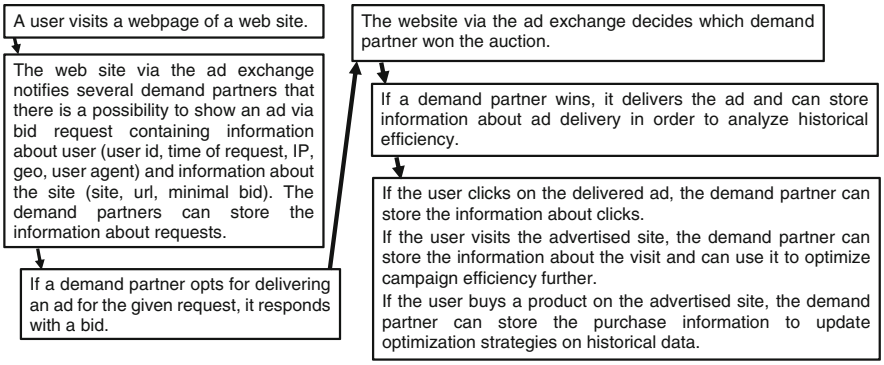

Fig. 1 Process of RTB and actions of a demand partner for delivering an ad

partners, which are essentially advertising platforms whose core business is the design of bidding strategies for ad requests for delivery advertisements. Specifically, each time when an ad exchange sends information about a user visiting a webpage, the demand partner can identify the prospectiveness of the request depending on the parameters (e.g., user id, webpage visited, IP address, user browser agent) and behaviour data (e.g., track record of the user over the latest few months) and propose a bid to compete in the auction with other demand partners. Thus, RTB enables a demand side to find a favourable ad campaign and submit a bid for a request depending on parameters of the request and behaviour data. Supposedly, online advertising brings customers at lower cost which is achieved by targeting narrow groups of users.

The process of showing online advertisements through the RTB systems occurs billions of times every day and consists of the steps displayed in Fig. 1.

The demand partner has to solve the problem of maximizing either the click through rate (CTR, i.e. the proportion of the number of clicks to the number of impressions) or the conversion rate (i.e. the proportion of the number of purchases to the number of impressions) by bidding on a set of requests under several constraints:

C1: Budget (total amount of money available for advertising);
C2: Number of impressions N_{total} (the total amount of ad exposures);
C3: Time (ad campaign is restricted to a certain time period).

In practice, the demand partner designs a strategy which cleverly chooses 5–500 million requests out of 50 billion available ones. To construct a good strategy, the demand partner has to use all log records.

General principles of adaptive designing are considered in [2, 4–6]. The design problem for optimizing the CTR has the following specifics compared to assumptions of the standard response surface methodology.

A1: The demand partner cannot choose requests with desired conditions but can leave an auction or suggest a bid for a user currently visiting a webpage.

A2: The design space is very complicated compared to typical $[0,1]^d$ and $\{0,1\}^d$ cases. Usually, the demand partner considers about 20 categorical factors; some factors (e.g. website, city, behaviour category) have hundreds of levels as well as other factors typically have about 10 levels.

A3: We observe the binary outcome but we have to consider the CTR as a function of the request.

The problem of adaptive targeting for ad campaigns was discussed in dozens of papers, see e.g. [10, 12, 16]. Some papers, for example [1, 15], use the look-alike idea implying that a new request will lead to the click/conversion if the new request is similar to (looks like one of) the previous successful requests. In 2014 two contests were organized at the Kaggle platform (www.kaggle.com), see [8] and [9], on algorithms for predicting the CTR using a dataset with subsampled non-click records so that the CTR for the dataset is about 20 % while for a typical advertising campaign the CTR is about 0.4 % or less. The algorithms, which were proposed by many teams are publicly available and give approximately the same performance with respect to the logarithmic loss criterion

$$\log(loss) = -\sum_{i=1}^{N} \Big\{ y_i \log(p_i) + (1-y_i)\log(1-p_i) \Big\} / N,$$

where N is the size of the data set, p_i is the predicted probability of a click for the i-th request, and $y_i = 1$ if the i-th request leads to a click and $y_i = 0$ otherwise. This criterion, however, does not look very sensible when the probabilities p_i are very small as it pays equal weights to type I and type II error probabilities (as noted above, typical values of p_i's are in the vicinity of 0.004 or even less).

In this paper, we provide a unified approach which comprises the popular methodologies, give a short review of these methodologies and make a comparison of several methods on real data.

2 Formal Statement of the Problem

Suppose that the advertisement we want to show is given and first assume that the price for showing a given ad is fixed; we shall also ignore the time constraint C3. Then the problem can be thought of as an optimization problem for a single optimality criterion which is the CTR. We discuss a generic adaptive targeting strategy which should yield the decision whether or not to show the ad to a request from a webpage visited by a user. If the strategy decides to show the ad, it then has to propose a bid.

An adaptive decision should depend on the current dataset of impressions and clicks which include all the users to whom we have shown the ad before and those who have clicked on the ad. Note that the dataset size N grows with time. We can increase the size of the dataset by including all our previous impressions of the

same or similar advertisements (perhaps applying some calibration to decrease the influence of past ad campaigns), so that N could be very large.

Denote the i-th request by $X_i = (x_{i,1}, \dots, x_{i,m})$, $i = 1, \dots, N$, where m is the number of features (factors); these features include the behavioural characteristics of the user, characteristics of the website, time of exposure, the device used (e.g. mobile telephone, tablet, PC , etc.), see the assumption A2. Let K be the number of the requests leading to a click on the ad, say, $X_{j_1}, \dots, X_{j_K}$, where $1 \le j_1 < \dots < j_K \le N$, among N requests of the current dataset of impressions. Note that K depends on N. The running performance criterion of the advertising campaign is the CTR defined by $P_N = K/N$. It is clear that the CTR changes as N grows.

We impose the following assumption of independence: if we choose a request $X = (x_1, \dots, x_m)$ then the probability of a click is $p(X)$; different events ('click' or 'no click') are independent. The assumption of independence obviously fails on the set of users that have already made a click on the ad at an earlier time (these users comprise the set $L(0)$ defined in Sect. 4) but it seems a reasonable assumption for the general set of users.

We also assume that all possible vectors $X = (x_1, \dots, x_m)$ belong to some set $\mathbb{X}$, which is either partly or fully discrete (see the assumption A2) and whose structure is difficult for determining a distance between different elements of $\mathbb{X}$. We also assume that for any two points X and $X' \in \mathbb{X}$, we can define a similarity measure $d(X, X')$ which does not have to satisfy mathematical axioms of the distance function.

If $\mathbb{X}$ is a discrete set with all possible requests $X = (x_1, \dots, x_m) \in \mathbb{X}$ given on the nominal scale then we can use the Hamming distance

$$d(X, X') = \sum_{j=1}^{m} \delta(x_j, x'_j), \quad \delta(x_j, x'_j) = \begin{cases} 1 & x_j = x'_j, \\ 0 & x_j \neq x'_j, \end{cases}$$

or the weighted Hamming distance $d(X, X') = \sum_{j=1}^{m} w_j \delta(x_j, x'_j)$, where the coefficients w_j are positive and proportional to the importance of the j-th feature (factor), $j = 1, \dots, m$. These weight coefficients can be chosen on the basis of the analysis of previous data of similar advertising campaigns, see Table 1 below.

Alternative ways of defining the similarity measure $d(X, X')$ are a logistic model for p_X (as is done in the so-called 'field-aware factorization machines' (FFM), see [14]) or to use sequential splitting of the set $\mathbb{X}$ based on the values of the most important factors of X ('gradient boosting machines' (GBM), see [7]). For FFM, the distance is defined on the space of parameters of the logistic model but in GBM $d(X, X')$ is small if $d(X, X')$ belongs to the same subset of $\mathbb{X}$ and it is large if the subsets which X and X' belong to have been split at early stages of the sequential splitting algorithm (that is, the values of the most influential features are very different).

Table 1 The CTR multiplied by 10^4 for several sets $L(r)$ with $T = 2015\text{-}02\text{-}08$ and different choices of factors. Abbreviation of factors are Be:behaviour category, We:website, Ex:ad exchange, Ci:city, Po:postcode, De:device type, Ag:user agent

Set of used factors, S	CTR$[L(0)\|S]$	CTR$[L(1)\|S]$	CTR$[L(2)\|S]$	f_i	I_{f_i}
Be,We,Ex,Ci,Po,De,Ag	15.3	5.01	2.36		
We,Ex,Ci,Po,De,Ag	5.13	2.43	2.35	Be	0.71
Be, Ex,Ci,Po,De,Ag	11.69	2.81	2.35	We	0.25
Be,We, Ci,Po,De,Ag	12.29	3.89	2.31	Ex	0.09
Be,We,Ex, Po,De,Ag	7.62	2.46	2.32	Ci	0.51
Be,We,Ex,Ci, De,Ag	14.96	2.45	2.32	Po	0.26
Be,We,Ex,Ci,Po, Ag	15.27	5.09	2.38	De	0.0003
Be,We,Ex,Ci,Po,De	4.87	3.37	2.20	Ag	0.58

2.1 Field-Aware Factorization Machines (FFM)

FFM describes the probability p_X by some sigmoidal parametric function, for example, the logistic function

$$p_X = \frac{1}{1 + e^{-m(X,\theta)}},$$

where θ is a vector of parameters and $m(X, \theta)$ is linear in the parameters. For example, the second-order function $m(X, \theta)$ is given by

$$m(X,\theta) = \theta_0 + \sum_{i=1}^{m}\sum_{k=1}^{n_i}\theta_{i,k}\delta(x_i, l_{i,k}) + \sum_{i=1}^{m-1}\sum_{k=1}^{n_i}\sum_{j=i+1}^{m}\sum_{s=1}^{n_j}\beta_{i,k;j,s}\delta(x_i, l_{i,k})\delta(x_j, l_{j,s}),$$

where $\beta_{i,k;j,s} = \sum_{z=1}^{q}\theta_{i,k,z}\theta_{j,s,z}$ describes a factorization procedure, $l_{i,k}$ are all possible levels of the i^{th} factor, $i = 1, \ldots, m$, $k = 1, \ldots, n_i$, $\delta(x_i, l_{i,k})$ equals 1 if $x_i = l_{i,k}$ and 0 otherwise. The vector of parameters θ consists of $\theta_0, \theta_{i,k}, \theta_{i,k,z}$ and is estimated by an iterative use of the gradient descent method for the logarithmic loss criterion, see [14].

A similar approach is the follow-the-regularized-leader (FTRL) methodology, where the function $m(X, \theta)$ has a simpler expression, see [11].

2.2 Gradient Boosting Machines (GBM)

GBM is a method of iterative approximation of the desired function p_X by a function of the form

$$p_X^{(k)} = \sum_{j=1}^{k}\alpha_j T(X, \theta_k),$$

where the vector θ_k is estimated at the k-th iteration, through minimizing the loss criterion [7]. Tree-based GBM considers the function $T(X, \theta)$ as the indicator function of the form

$$T(X, \theta) = \begin{cases} \theta_{in}, & \theta_{i,low} \le x_i \le \theta_{i,up},\ i = 1, \dots, m, \\ \theta_{out}, & \text{otherwise}, \end{cases}$$

where $\theta = (\theta_{in}, \theta_{out}, \theta_{1,low}, \theta_{1,up}, \dots, \theta_{m,low}, \theta_{m,up})$. Note that levels of categorical variables are encoded by integer numbers.

3 Generic Adaptive Strategy for Maximizing the CTR of an Advertising Campaign

The purpose of the strategy for maximizing the CTR is to employ the training set of past records for the new requests we will be showing the ad, to increase P_N as N increases.

We can always assume that N_{total} defined in the assumption C2 is very large. Mathematically, we can then assume that $N \to \infty$. If we assume that the bid price is the same (that is, we ignore C1) and there is no time constraint (here we ignore C3) then formally our aim becomes devising a strategy such that $\lim_{N\to\infty} P_N$ is maximum. This is simply an optimization problem of $p_X,\ x \in \mathbb{X}$. The algorithms solving this problem do this either in the parameter space (for the factorization machines) or in the original space $\mathbb{X}$ (for GBM and the look-alike strategies).

The main problems for applying these algorithms are as follows:

- Factorization machines: the number of parameters is of the order of billions. In practice, this number is reduced by confounding parameters.
- Gradient boosting: the number of observations with certain ranges of levels for several factors is small.
- Computational time grows very fast for all approaches as the size of training data increases. Consequently, in practice training data are often subsampled.
- All approaches have several tuning parameters which should be carefully chosen.

By the nature of the methods, the look-alike approach is applicable in practice if the number of observed clicks K is at least a few dozens, the GBM approach is applicable if K is at least several hundreds and the FFM approach is applicable if K is at least several thousands.

A generic adaptive strategy is an evolutionary one which chooses new requests in the vicinity of the requests that were successful previously; in marketing these kinds of methods are called look-alike methods. To define the preference criterion, for all N we need an estimator $\hat{p}_N(X)$ of the function $p(X)$, which is defined for all $X \in \mathbb{X}$. We do not need to construct the function $\hat{p}_N(X)$ explicitly; we just need to compute values of $\hat{p}_N(X)$ for a given X, where X is a request which is currently on

offer for a demand partner. We hence suggest the following estimator $\hat{p}_N(X)$:

$$\hat{p}_N(X) = \frac{\sum_{k=1}^{K} \omega_{j_k} \exp\{-\lambda_N d(X, X_{j_k})\}}{\sum_{i=1}^{N} \omega_i \exp\{-\lambda_N d(X, X_i)\}} + \varepsilon_N \,, \tag{1}$$

where λ_N and ε_N are some positive constants (possibly depending on N) and ω_i is the weight of the i-th observation made after a calibration of the data is made (the possibility of making such calibration has been mentioned above). The sum in the numerator in (1) is taken over all users who have clicked on the ad. If all these (good) requests are far away from X then the value $\hat{p}_N(X)$ will be very close to zero. The constant ε_N is a regularization constant. As $\varepsilon_N > 0$ there is always a small probability assigned to each X, even if in the past there were no successful requests that were similar to X. Theoretically, as $N \to \infty$, we may assume that $\varepsilon_N \to 0$.

Note that an estimator $\hat{p}_N(X)$ for $p(X)$ is implicitly constructed in the factorization machines and in gradient boosting machines too. Using an estimator $\hat{p}_N(X)$, we can suggest how much the demand partner can offer for the request X in the bidding procedure (that is, we stop optimizing $p(X)$ and take into account the constraint C2). For example, the demand partner can offer larger bids if $\hat{p}_N(X) \geq p_*$, where p_* is the desired probability we want to reach. Another possible use of an estimator $\hat{p}_N(X)$ can be based on the following idea: the amount of money the advertising platform offers for X is proportional to the difference $\hat{p}_N(X) - K/N$, if this difference is positive, and a very small bid, if the difference is negative. For these strategies we do not obtain $\lim_N \hat{p}_N(X) = \max_X p(X)$ but we construct effective strategies which take into account not only the constraint C2 but also C1 and C3. Note in this respect that it is always a good idea to offer very small bids to the users with small values of $\hat{p}_N(X)$ for the following reasons: (a) learning about $p(X)$ in the subregions of X where we perhaps do not have much data, (b) the difference (ratio) between large values of probabilities $p(X)$ for 'good' X and 'bad' X can be smaller than the difference of the option prices for these 'good' and 'bad' X's, (c) the constraint C3 is easier to satisfy, and (d) by saving some funds on cheap X's we can afford higher prices on X's with large values of $\hat{p}_N(X)$.

4 Analysis of Real Data

In the present section we analyze an ad campaign which was executed by Crimtan from 2015-02-01 to 2015-02-17, the number of impressions is slightly above 3 millions and the number of clicks is slightly above 700, so that the CTR $\hat{p} \cong 2.4 \cdot 10^{-4}$, thus the FFM approach is not applicable.

To investigate the performance of the strategies for the database of requests for the ad campaign, we split the database of impressions into 2 sets: the training set $\mathbb{X}_p(T)$ of past records with dates until a certain time T (where T is interpreted as the present time) and the test set $\mathbb{X}_f(T)$ of future records with dates from the time T.

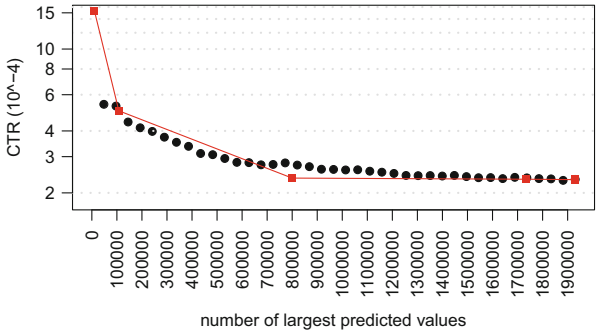

Fig. 2 The CTR for favorable samples of requests of certain sizes for the look-alike approach (*squares*) and the GBM approach (*dots*), $T = 2015\text{-}02\text{-}08$, 7 factors are used

We now compare GBM and the look-alike approach by comparing the CTR for the samples of most favorable requests with the highest chances to click in Fig. 2.

To form the sample of most favourable requests for the look-alike approach, we define the set

$$L(r) = \{X_j \text{ from } \mathbb{X}_f(T)\colon \min_{clicked\ \tilde{X}_i \in \mathbb{X}_p(T)} d(X_j, \tilde{X}_i) \le r\};$$

that is, $L(r)$ is a set of requests where we have shown the ad and the minimal distance to the set of clicked requests from the set of past records is not greater than r. In other words, the set $L(r)$ is an intersection of the set of our requests with the union of balls of radius r centered around the clicked past requests. We consider X_j with 7 factors: website, ad exchange, city, postcode, device type, user agent, user behaviour category. In Fig. 2 the points corresponding to the look-alike approach are given by $\big(\text{size of } L(r), \text{CTR for } L(r)\big)$, $r = 0, \dots, 4$.

To form the sample of most favourable requests for the GBM approach, we construct the GBM model using the training set $\mathbb{X}_p(T)$ and then apply this model to predict the probability to click for each request from the test set $\mathbb{X}_f(T)$. Now we can sort the predicted probabilities and create samples of requests with the highest predicted probabilities to click.

In Fig. 2 we can see that the look-alike approach and the GBM approach have similar possibilities to increase the CTR for the considered ad campaign.

Let us perform a sensitivity analysis of the CTR for the sets $L(r)$. In Table 1 we show the CTR for several sets $L(r)$ with $T = 2015\text{-}02\text{-}08$ and different choices of factors, and the index of the influence of the ith factor

$$I_{f_i} = \sum_{r=0}^{2} \left(1 - \frac{\mathrm{CTR}[L(r)|f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_m]}{\mathrm{CTR}[L(r)|f_1, \dots, f_m]}\right)^2$$

where $\mathrm{CTR}[L(r)|f_1,\ldots,f_m]$ is the CTR for the set $L(r)$ with requests containing only factors $f_1,\ldots,f_m$. We can observe that $I_{\mathrm{De}} = 0.0003$ and $I_{\mathrm{Ex}} = 0.09$; that is, the device type has no influence and the ad exchange has a small influence on the CTR; consequently such factors can be removed from the model (and computations). The postcode has no influence on the CTR for the set $L(0)$ but has some influence on the CTR for the set $L(1)$.

In contrast, the user agent, the user behaviour category, and the city are very influential factors. It is rather surprising that the postcode has no influence but the city has a big influence on the CTR for the set $L(0)$.

Acknowledgements The paper is a result of collaboration of Crimtan, a provider of proprietary ad technology platforms and the University of Cardiff. Research of the third author was supported by the Russian Science Foundation, project No. 15-11-30022 "Global optimization, supercomputing computations, and application".

References

1. Aly, M., Hatch, A., Josifovski, V., Narayanan, V.K.: Web-scale user modeling for targeting. In: Proceedings of the 21st International Conference Companion on World Wide Web, pp. 3–12. ACM, New York (2012)
2. Box, G.E.P., Wilson, K.B.: On the experimental attainment of optimum conditions (with discussion). J. R. Stat. Soc. Ser. B **13**, 1–45 (1951)
3. eMarketer: US programmatic ad spend tops $10 Billion this year, to double by 2016. http://www.emarketer.com/Article/US-Programmatic-Ad-Spend-Tops-10-Billion-This-Year-Double-by-2016/1011312 (2014)
4. Ermakov, S.M., Zhigljavsky, A.A.: Mathematical Theory of Optimal Design. Nauka, Moscow (1987)
5. Fedorov, V.V., Hackl, P.: Model-Oriented Design of Experiments. Springer, New York (2012)
6. Fedorov, V.V., Leonov, S.L.: Optimal Design for Nonlinear Response Models. CRC Press, Boca Raton (2013)
7. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**, 1189–1232 (2001)
8. https://www.kaggle.com/c/avazu-ctr-prediction. Accessed 12 Sept 2015
9. https://www.kaggle.com/c/criteo-display-ad-challenge. Accessed 12 Sept 2015
10. Jansen, B.J., Mullen, T.: Sponsored search: an overview of the concept, history, and technology. Int. J. Electron. Bus. **6**, 114–131 (2008)
11. McMahan, H.B.: Follow-the-regularized-leader and mirror descent: equivalence theorems and l1 regularization. In: International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, pp. 525–533 (2011)
12. McMahan, H.B., Holt, G., Sculley, D., et al.: Ad click prediction: a view from the trenches. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, pp. 1222–1230. ACM (2013)
13. Nicholls, S., Malins, A., Horner, M.: Real-time bidding in online advertising. http://www.gpbullhound.com/wp-content/uploads/2014/09/Real-Time-Bidding-in-Online-Advertising.pdf (2014)
14. Rendle, S.: Factorization machines. In: IEEE 10th International Conference on Data Mining (ICDM 2010), Sydney, pp. 995–1000 (2010)

15. Tu, S., Lu, C.: Topic-based user segmentation for online advertising with latent Dirichlet allocation. In: Advanced Data Mining and Applications, pp. 259–269. Springer, Berlin/Heidelberg (2010)
16. Yang, S., Ghose, A.: Analyzing the relationship between organic and sponsored search advertising: positive, negative or zero interdependence? Mark. Sci. **29**, 602–623 (2010)

Interpolation and Extrapolation in Random Coefficient Regression Models: Optimal Design for Prediction

Maryna Prus and Rainer Schwabe

Abstract The problem of optimal design for the prediction of individual parameters in random coefficient regression in the particular case of a given population mean was considered by Gladitz and Pilz (Statistics 13:371–385, 1982). In the more general situation, where the population parameter is unknown, *D*- and *L*-optimal designs were discussed in Prus and Schwabe (J R Stat Soc Ser B, 78:175–191). Here we present analytical results for designs which are optimal for prediction in the case of interpolation as well as extrapolation of the individual response.

1 Introduction

Hierarchical random coefficient regression models are very popular in many fields of statistical applications. The subject of this paper is the problem of finding optimal designs for interpolation and extrapolation in such models. Optimal designs for interpolation and extrapolation in fixed effects models were considered in detail in Kiefer and Wolfowitz [7] and [8]. Some theory for determining Bayesian optimal designs, which are also optimal for the prediction of individual deviations from the population mean in random coefficients regression models (see Prus [9], ch. 5), was developed by Chaloner [3] (see also Chaloner [2]). The problem of optimal designs for the prediction of individual parameters in random coefficient regression models was considered in Prus and Schwabe [11] for the linear and the generalized *D*-criteria. Here we deal with the *c*-criterion for the prediction and formulate the optimality condition for approximate designs. We consider the problem of optimal designs for interpolation and extrapolation as a particular case of the *c*-criterion for the straight line regression model and illustrate the results by a numerical example.

The paper has the following structure. In the second part the hierarchical random coefficient regression model is specified and the best linear unbiased prediction of the individual parameters is introduced. The third part provides analytical results for *c*-optimal designs. In the fourth part optimal designs for interpolation and

M. Prus (✉) • R. Schwabe
Institute for Mathematical Stochastics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany
e-mail: maryna.prus@ovgu.de; rainer.schwabe@ovgu.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_24

extrapolation are considered. The last section presents some discussion and an outlook.

2 Model Specification and Prediction

In the hierarchical random coefficient regression model the j-th observation of individual i is given by

$$Y_{ij} = \mu_i(x_{ij}) + \varepsilon_{ij}, \quad x_{ij} \in \mathscr{X}, \quad j = 1, .., m_i, \quad i = 1, .., n, \tag{1}$$

at the individual level, where n denotes the number of individuals, m_i is the number of observations on individual i, $\mu_i(x)$ are the response functions of the form $\mu_i(x) = \mathbf{f}(x)^\top \boldsymbol{\beta}_i$, $\mathbf{f} = (f_1, .., f_p)^\top$ is the vector of known regression functions, and $\boldsymbol{\beta}_i = (\beta_{i1}, .., \beta_{ip})^\top$ is the individual parameter vector specifying the individual response. The experimental settings x_{ij} may be chosen from a given experimental region $\mathscr{X}$. Within an individual the observations are assumed to be uncorrelated given the individual parameters. The observational errors ε_{ij} have zero mean $\mathrm{E}(\varepsilon_{ij}) = 0$ and are homoscedastic with common variance $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$.

The individual random parameters $\boldsymbol{\beta}_i$ are assumed to have an unknown population mean $\mathrm{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta}$ and a given covariance matrix $\mathrm{Cov}(\boldsymbol{\beta}_i) = \sigma^2 \mathbf{D}$, where the dispersion matrix $\mathbf{D}$ is assumed to be regular. All individual parameters and all observational errors are assumed to be uncorrelated.

We consider the particular case of the model (1) where the number of observations as well as the experimental settings are the same for all individuals ($m_i = m$ and $x_{ij} = x_j$).

We investigate the predictor of the individual parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$. This predictor is also sometimes called the estimator of the random parameters and can be viewed as an empirical Bayes estimator.

As exhibited in Prus and Schwabe [10] the best linear unbiased predictor $\hat{\boldsymbol{\beta}}_i$ of the individual parameter $\boldsymbol{\beta}_i$ is a weighted average of the individualized estimate $\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{Y}_i$, based on the observations on individual i, and the estimator of the population mean $\hat{\boldsymbol{\beta}} = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \bar{\mathbf{Y}}$,

$$\hat{\boldsymbol{\beta}}_i = \mathbf{D}\{(\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D}\}^{-1} \hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} + (\mathbf{F}^\top \mathbf{F})^{-1} \{(\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D}\}^{-1} \hat{\boldsymbol{\beta}}. \tag{2}$$

Here $\mathbf{F} = (\mathbf{f}(x_1), \ldots, \mathbf{f}(x_m))^\top$ denotes the individual design matrix, which is equal for all individuals, $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{im})^\top$ is the observation vector for individual i, and $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i$ is the average response across all individuals.

The performance of the prediction (2) may be measured in terms of the mean squared error matrix of $(\hat{\boldsymbol{\beta}}_1^\top, \ldots, \hat{\boldsymbol{\beta}}_n^\top)^\top$. The latter matrix has the form

$$\text{MSE} = \sigma^2\{(\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{F}^\top\mathbf{F})^{-1}\}, \tag{3}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{1}_n$ is a n-dimensional vector of ones and "$\otimes$" denotes the Kronecker product of matrices. The mean squared error matrix (3) is a weighted average of the corresponding covariance matrix in the fixed effects model and the Bayesian one.

Note that the response functions $\mu_i(x_0) = \mathbf{f}(x_0)^\top\boldsymbol{\beta}_i$ may be predictable for x_0 from the experimental region $\mathscr{X}$ even if the design matrix $\mathbf{F}$ is not of full rank and consequently the matrix $\mathbf{F}^\top\mathbf{F}$ is singular, as long as $\mathbf{f}(x_0)$ belongs to the column space of $\mathbf{F}^\top$. However, the individual parameters $\boldsymbol{\beta}_i$ themselves are not predictable if the design matrix is not full rank (see Prus [9], ch. 5).

3 *c*-Optimal Design

The mean squared error matrix of a prediction depends crucially on the choice of the observational settings $x_1, \ldots, x_m$, which can be chosen by the experimenter to minimize the mean squared error matrix and which constitute an exact design. Typically the optimal settings will not necessarily all be distinct. Then a design

$$\xi = \begin{pmatrix} x_1 & \ldots, & x_k \\ w_1 & \ldots, & w_k \end{pmatrix} \tag{4}$$

can be specified by its distinct settings $x_1, \ldots, x_k$, $k \leq m$, say, and the corresponding numbers of replications $m_1, \ldots, m_k$ or the corresponding proportions $w_j = m_j/m$.

For analytical purposes, we make use of approximate designs in the sense of Kiefer (see e.g. Kiefer [6]) for which the integer condition on mw_j is dropped and the weights $w_j \geq 0$ may be any real numbers satisfying $\sum_{j=1}^k w_j = 1$ or equivalently $\sum_{j=1}^k m_j = m$. For these approximate designs the standardized information matrix for the model without random effects ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$, i.e. $\mathbf{D} = \mathbf{0}$) is defined as

$$\mathbf{M}(\xi) = \textstyle\sum_{j=1}^k w_j\mathbf{f}(x_j)\mathbf{f}(x_j)^\top = \frac{1}{m}\mathbf{F}^\top\mathbf{F}. \tag{5}$$

Further we introduce the standardized dispersion matrix of the random effects $\boldsymbol{\Delta} = m\mathbf{D}$ for notational ease. With this notation, we define the standardized mean squared error matrix for the prediction of individual parameters as

$$\text{MSE}\,(\xi) = (\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes \{\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1}\}^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes \mathbf{M}(\xi)^{-1}. \tag{6}$$

For any exact design ξ all mw_j are integer. Then the matrix MSE (ξ) coincides with the mean squared error matrices (3) up to a multiplicative factor σ^2/m.

In this paper, we focus on the extended c-criterion for prediction (see Prus [9], ch. 5 or Prus and Schwabe [11]), which is defined as the sum of the variances of $\mathbf{c}^\top\hat{\boldsymbol{\beta}}_i - \mathbf{c}^\top\boldsymbol{\beta}_i$ across all individuals, where $\mathbf{c}$ is a specified vector of dimension p:

$$\mathrm{c}_\beta(\xi) = \sum_{i=1}^{n} \mathrm{var}\,(\mathbf{c}^\top\hat{\boldsymbol{\beta}}_i - \mathbf{c}^\top\boldsymbol{\beta}_i). \tag{7}$$

Using (6), the standardized c-criterion $\Phi_\beta = \frac{m}{\sigma^2}\mathrm{c}_\beta$ can be represented as

$$\Phi_\beta(\xi) = \mathbf{c}^\top\mathbf{M}(\xi)^{-1}\mathbf{c} + (n-1)\,\mathbf{c}^\top\{\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1}\}^{-1}\mathbf{c}, \tag{8}$$

which is a weighted sum of the c-criterion in the fixed effects model and the Bayesian c-criterion.

With the general equivalence theorem (see e.g. Silvey [13], ch. 3), we obtain the following characterization of an optimal design.

Theorem 1 *The approximate design ξ^* with non-singular information matrix $\mathbf{M}(\xi^*)$ is c-optimal for the prediction of individual parameters if and only if*

$$\begin{aligned}&\{\mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathbf{c}\}^2 + (n-1)\,[\mathbf{f}(x)^\top\{\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\}^{-1}\mathbf{c}]^2\\ &\leq\ \mathbf{c}^\top\mathbf{M}(\xi^*)^{-1}\,\mathbf{c} + (n-1)\,\mathbf{c}^\top\{\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\}^{-1}\mathbf{M}(\xi^*)\{\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1}\}^{-1}\mathbf{c}\end{aligned} \tag{9}$$

for all $x \in \mathscr{X}$.

For any experimental setting x_j of ξ^* with $w_j > 0$ equality holds in (9).

Note that optimal designs, which result in singular information matrices, may also exist.

4 Optimal Designs for Interpolation and Extrapolation

For models without random effects the interpolation and extrapolation problem was considered in detail by Kiefer and Wolfowitz [7] and [8]. The Bayesian optimal designs were discussed in Chaloner [3] (see also Chaloner [2]).

In this section we examine the straight line regression

$$Y_{ij} = \beta_{i1} + \beta_{i2}x_j + \varepsilon_{ij} \tag{10}$$

on the experimental region $\mathscr{X} = [0, 1]$. The settings x_j can be interpreted as time or dosage and $x = 0$ means a measurement at baseline. We assume uncorrelated components such that the dispersion matrix $\mathbf{D} = \mathrm{diag}\,(d_1, d_2)$ of the random effects

is diagonal with entries d_1 and d_2 for the variance of the intercept and slope, respectively. The variance of the intercept is assumed to be small, $d_1 < 1/m$.

The problem of optimal designs for interpolation and extrapolation of the response functions $\mu_i(x_0) = \mathbf{f}(x_0)^\top \boldsymbol{\beta}_i$ at some given point $x_0 \in \mathscr{X} = [0, 1]$ and $x_0 \notin \mathscr{X} = [0, 1]$, respectively, may be recognized as a special case of the c-criterion (8) for prediction with $c = \mathbf{f}(x_0) = (1, x_0)^\top$.

It follows from Theorem 1 that for $c = (1, x_0)^\top$ the c-optimal designs (with non-singular information matrices) only take observations at the endpoints $x = 0$ and $x = 1$ of the design region, since the sensitivity function, given by the left hand side of inequality (9), is then a polynomial in x of degree 2 with positive leading term. Hence, the optimal design ξ^* is of the form

$$\xi_w = \begin{pmatrix} 0 & 1 \\ 1-w & w \end{pmatrix}, \tag{11}$$

and only the optimal weight w^* has to be determined. For designs ξ_w the criterion function (8) is calculated with $\gamma_k = 1/(m\, d_k)$ for $k = 1, 2$ as

$$\Phi_\beta(\xi_w) = \frac{x_0^2 - 2wx_0 + w}{w(1-w)} + (n-1)\frac{x_0^2(1+\gamma_1) - 2wx_0 + w + \gamma_2}{(1+\gamma_1)(w+\gamma_2) - w^2}. \tag{12}$$

To obtain numerical results, the number of individuals and the number of observations on each individual are fixed to $n = 100$ and $m = 10$. For the variance d_1 of the intercept, we use the value 0.001. Figure 1 illustrates the dependence of the optimal weight w^* on the rescaled variance parameter $\rho = d_2/(1 + d_2)$, which in a way mimics the intraclass correlation and has the advantage of being bounded, so that the whole range of slope variances d_2 can be shown. We use the values 0.9, 0.5 and 0.3 for the interpolation point and the values 1.2, 2 and 100 for the extrapolation point x_0. Note that, for $\rho = 0$, the optimal weights w^* for both interpolation and

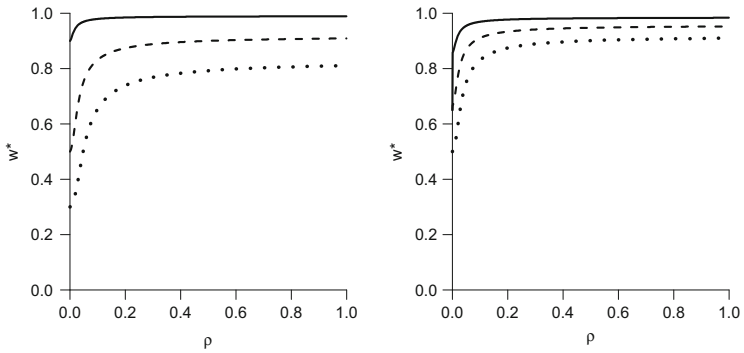

Fig. 1 Optimal weights w^* for interpolation (*left panel*) at $x_0 = 0.9$ (*solid line*), $x_0 = 0.5$ (*dashed line*) and $x_0 = 0.3$ (*dotted line*) and extrapolation (*right panel*) at $x_0 = 1.2$ (*solid line*), $x_0 = 2$ (*dashed line*) and $x_0 = 100$ (*dotted line*) as functions of $\rho = d_2/(1+d_2)$

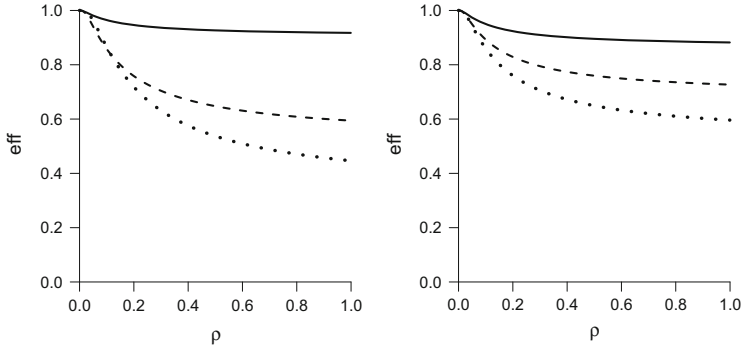

Fig. 2 Efficiency of the optimal designs in fixed effects model for interpolation (*left panel*) at $x_0 = 0.9$ (*solid line*), $x_0 = 0.5$ (*dashed line*) and $x_0 = 0.3$ (*dotted line*) and extrapolation (*right panel*) at $x_0 = 1.2$ (*solid line*), $x_0 = 2$ (*dashed line*) and $x_0 = 100$ (*dotted line*) as functions of $\rho = d_2/(1+d_2)$

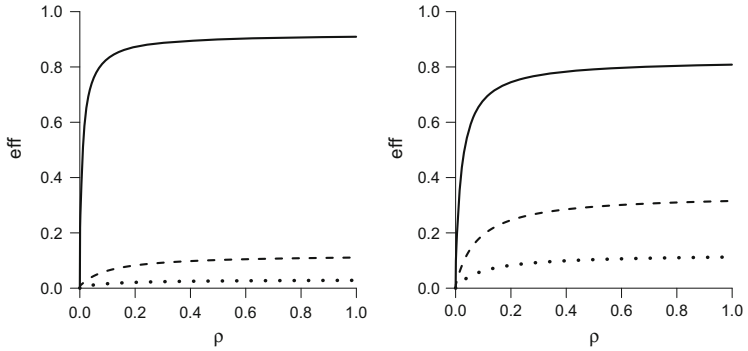

Fig. 3 Efficiency of the Bayesian optimal designs for interpolation (*left panel*) at $x_0 = 0.9$ (*solid line*), $x_0 = 0.5$ (*dashed line*) and $x_0 = 0.3$ (*dotted line*) and extrapolation (*right panel*) at $x_0 = 1.2$ (*solid line*), $x_0 = 2$ (*dashed line*) and $x_0 = 100$ (*dotted line*) as functions of $\rho = d_2/(1+d_2)$

extrapolation are the same as those in the model without random effects, namely: $w^* = x_0$ and $w^* = x_0/(2x_0 - 1)$, respectively (see e.g. Schwabe [12], ch. 2).

In Fig. 2 the efficiency $\mathrm{eff}(\xi) = \Phi_\beta(\xi_{w^*})/\Phi_\beta(\xi_w)$ is plotted for the optimal designs ξ_w with $w = x_0$ and $w = x_0/(2x_0 - 1)$ in the fixed effects model for interpolation and extrapolation, respectively.

Figure 3 represents the efficiency of the Bayesian optimal designs. The efficiency has been computed as $\mathrm{eff}(\xi) = \Phi_\beta(\xi_{w^*})/\Phi_\beta(\xi_{w_B})$, where ξ_{w_B} denotes the Bayesian optimal designs. Note that the latter designs are optimal for the prediction of individual deviations $\mathbf{b}_i = \boldsymbol{\beta}_i - \boldsymbol{\beta}$ from the population mean in the random coefficient regression model (1) (see Prus [9], ch. 5).

5 Discussion and Outlook

We have proposed an analytical method for determining optimal interpolation and extrapolation designs for the prediction of individual parameters in hierarchical random coefficient regression. This problem was considered as a particular case of the c-criterion for prediction. The criterion function of the c-criterion is a weighted sum of the c-criterion in the fixed effects models and the Bayesian c-criterion and can be recognized as a special case of a compound criterion (see e.g. Cook and Wong [4] or Atkinson et al. [1], ch. 21).

It was established that the optimal two-point designs in the fixed effects models are in general not optimal for prediction. Note that the one-point designs, which take all observations at point x_0 and are also optimal for estimation in models without random effects, result in even larger values of the criterion function than the two-points designs mentioned above and perform hence worse. In the numerical example for the straight line regression with a diagonal dispersion matrix the Bayesian optimal designs show a similar behavior as the optimal designs for prediction. They lead, however, for $x_0 \to 1$ to a singular information matrix and cannot be used for prediction.

The optimality condition for the c-criterion proposed here was formulated for approximate designs, which are not directly applicable and the optimal weights have to be appropriately rounded. The analytical results presented in this paper are based on the assumption that all individuals (observational units) get the same treatment. In one of the next steps of the research, the design optimality problem will be considered for a more general case of the hierarchical random coefficient regression, where different designs are allowed for different individuals.

References

1. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)
2. Chaloner, K.: Optimal Bayesian experimental design for linear models. Department of Statistics Technical report 238, Carnegie-Mellon University, Pittsburgh (1982)
3. Chaloner, K.: Optimal Bayesian experimental design for linear models. Ann. Stat. **12**, 283–300 (1984)
4. Cook, R.D., Wong, W.K.: On the equivalence of constrained and compound optimal designs. J. Am. Stat. Assoc. **89**, 687–692 (1994)
5. Gladitz, J., Pilz, J.: Construction of optimal designs in random coefficient regression models. Statistics **13**, 371–385 (1982)
6. Kiefer, J.: General equivalence theory for optimum designs (approximate theory). Ann. Stat. **2**, 849–879 (1974)
7. Kiefer, J., Wolfowitz, J.: Optimum extrapolation and interpolation designs I. Ann. Inst. Stat. Math. **16**, 79–108 (1964)
8. Kiefer, J., Wolfowitz, J.: Optimum extrapolation and interpolation designs II. Ann. Inst. Stat. Math. **16**, 295–303 (1964)

9. Prus, M.: Optimal designs for the prediction in hierarchical random coefficient regression models. Ph.D. thesis, Otto-von-Guericke University Magdeburg (2015)
10. Prus, M., Schwabe, R.: Optimal designs for individual prediction in random coefficient regression models. In: Melas, V., Nachtmann, G., Rasch, D. (eds.) Optimal Design of Experiments – Theory and Application: Proceedings of the International Conference in Honor of the late Jagdish Srivastava, pp. 122–129. Center of Experimental Design, University of Natural Resources and Life Sciences, Vienna (2011)
11. Prus, M., Schwabe, R.: Optimal designs for the prediction of individual parameters in hierarchical models. J. R. Stat. Soc.: Ser. B. **78**, 175–191 (2016)
12. Schwabe, R.: Optimum Designs for Multi-Factor Models. Springer, New York (1996)
13. Silvey, S.D.: Optimal Design. Chapman & Hall, London (1980)

Invariance and Equivariance in Experimental Design for Nonlinear Models

Martin Radloff and Rainer Schwabe

Abstract In this note we exhibit the usefulness of invariance considerations in experimental design in the context of nonlinear models. Therefor we examine the equivariance of locally optimal designs and their criteria functions and establish the optimality of invariant designs with respect to robust criteria like weighted or maximin optimality, which avoid parameter dependence.

1 Introduction

Invariance and equivariance are powerful tools in the construction of optimal experimental designs. In this context equivariance means that simultaneous transformations on the design region and on the design considered leave the performance of the design unchanged. In particular, this implies that an optimal design on the original design region will be transformed to an optimal design on the transformed design region. If a design problem is additionally invariant with respect to a whole group of transformations (*symmetries*), then the search for an optimal design can be restricted to the invariant (*symmetrized*) ones.

For standard linear models these properties are well-known and widely investigated (see e.g. Pukelsheim [7, Chapter 13], Gaffke and Heiligers [4] or Schwabe [10, Chapter 3]). In nonlinear setups, however, additional features occur which need some clarification. In particular, because of the parameter dependence of the information matrix, the introduction of an additional transformation on the parameter space is required. Starting with this parameter transformation, Ford et al. [3] introduced in their seminal paper equivariance under the name *canonical transformation* in the context of generalized linear models. Using this concept Ford et al. [3] could characterize locally optimal designs for arbitrary pre-specified parameter values on the basis of standardized models. This approach can readily be extended to other nonlinear situations as long as the intensity function of the information depends only on the value of the linear predictor.

M. Radloff (✉) • R. Schwabe
Institute for Mathematical Stochastics, Otto-von-Guericke-University, PF 4120, 39016 Magdeburg, Germany
e-mail: martin.radloff@ovgu.de; rainer.schwabe@ovgu.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_25

Here we put their approach bottom-up and start with transformation of the design region which makes the arguments more straightforward. First we introduce the model specifications and exhibit the local optimality of transformed designs in Sect. 2. In Sect. 4 we establish the optimality of invariant designs with respect to robust criteria like weighted or maximin optimality, which avoid parameter dependence, under the natural additional assumption that also the weight function or the parameter region of interest, respectively, are invariant.

2 Model, Transformation and Local Optimality

In the following sections it is required that the one-support-point information matrix $M(x, \beta)$ can be written in the form

$$M(x, \beta) = \lambda\{f(x)^\top \beta\} f(x) f(x)^\top,$$

with an intensity function λ which only depends on the value of the linear predictor. The design point x is in the design region $\mathscr{X} \subseteq \mathbb{R}^k$ and the parameter β is in $\mathscr{B} \subseteq \mathbb{R}^p$. $f : \mathscr{X} \to \mathbb{R}^p$ is the regression function.

Hence the information matrix of the (generalized) design ξ with independent observations is

$$M(\xi, \beta) = \int_{\mathscr{X}} M(x, \beta)\xi(\mathrm{d}x) = \int_{\mathscr{X}} \lambda\{f(x)^\top \beta\} f(x) f(x)^\top \xi(\mathrm{d}x).$$

In generalized linear models (see McCulloch and Searle [6]) or for example in censored data models (see Schmidt and Schwabe [9]) this prerequisite is fulfilled.

Let g be a one-to-one-mapping of the design region $\mathscr{X}$. If there is a $p \times p$ matrix Q_g with $f\{g(x)\} = Q_g f(x)$ for all $x \in \mathscr{X}$, g induces a linear transformation of the regression function f and f is linearly equivariant.

These one-to-one-mappings of the design region $\mathscr{X}$ provide a group G, if every $g \in G$ induces a linear transformation of the regression function f. One example is the orthogonal group $O(k)$ or subgroups of it for a k-dimensional circle.

Here we only want to focus on finite groups of transformation, like the group generated by a 90° rotation in a 2-dimensional space.

The linear predictor $f(x)^\top \beta$ should be equivariant with respect to the transformation g of the design space. So it is necessary to transform the parameter space as well and to find an analogous parameter transformation $\tilde{g}$ with

$$f\{g(x)\}^\top \tilde{g}(\beta) = f(x)^\top \beta;$$

$\tilde{g}(\beta) = (Q_g^\top)^{-1}\beta$ is such a transformation for regular Q_g.

If not only the parameter but also the design points are transformed in this way, the information matrix has the form

$$\begin{aligned}
M\{\xi^g, \tilde{g}(\beta)\} &= \int_{\mathscr{X}} M\{x, \tilde{g}(\beta)\}\xi^g(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\{f(x)^\top \tilde{g}(\beta)\} f(x) f(x)^\top \xi^g(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\big[f\{g(x)\}^\top \tilde{g}(\beta)\big] f\{g(x)\} f\{g(x)\}^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\big[\{Q_g f(x)\}^\top (Q_g^\top)^{-1}\beta\big]\, Q_g f(x)\{Q_g f(x)\}^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\{f(x)^\top Q_g^\top (Q_g^\top)^{-1}\beta\}\, Q_g f(x) f(x)^\top Q_g^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\{f(x)^\top \beta\}\, Q_g f(x) f(x)^\top Q_g^\top \xi(\mathrm{d}x) \\
&= Q_g \int_{\mathscr{X}} \lambda\{f(x)^\top \beta\} f(x) f(x)^\top \xi(\mathrm{d}x)\; Q_g^\top \\
&= Q_g M(\xi, \beta) Q_g^\top \quad .
\end{aligned}$$

In linear models it is commonly known, that optimality criteria Φ are invariant,

$$\Phi[M\{\xi^g, \tilde{g}(\beta)\}] = \Phi\{Q_g M(\xi, \beta) Q_g^\top\} = \Phi\{M(\xi, \beta)\},$$

if the information matrix is transformed by the matrices Q_g or the group G satisfying special properties. So, for example, Φ_q-optimality, $0 \le q \le \infty$, is invariant if the transformation matrix Q_g is orthonormal (see Schwabe [10, Chapter 3]). In the special case of D-optimality it is enough to have unimodal matrices Q_g, i.e. $|\det(Q_g)| = 1$.

Local optimality is not invariant, in general, because the information matrix depends on the parameter β.

There are two consequences. If ξ is a locally optimal design for β with respect to Φ then ξ^g is locally optimal for $\tilde{g}(\beta)$. Otherwise, if there is a better design ξ^* for $\tilde{g}(\beta)$, $(\xi^*)^{g^{-1}}$ will be better for β—contradiction.

And, in situations, where $\tilde{g}(\beta) = \beta$ for all $\tilde{g}$, the local criterion Φ is invariant at β:

$$\Phi\{M(\xi^g, \beta)\} = \Phi[M\{\xi^g, \tilde{g}(\beta)\}] = \Phi\{M(\xi, \beta)\} \quad \text{for all } g \in G.$$

By defining the symmetrized design of ξ with respect to the finite group G

$$\overline{\xi} := \frac{1}{\#G} \sum_{g \in G} \xi^g \,,$$

which is an invariant design (with respect to G, $\overline{\xi}^g = \overline{\xi}$ for all $g \in G$), it can be shown, as in the theory of linear models (see Schwabe [10, Chapter 3]), that for a concave and (with respect to G) invariant criterion function Φ, $\Phi(\overline{\xi}) \geq \Phi(\xi)$ for every possible design ξ. Hence locally optimal designs can be found in the class of invariant designs. This is in particular the case when $\beta = 0$, which is not surprising, because then the local criterion coincides with that in the corresponding linear model without intensity function.

The first consequence is illustrated in a small example.

3 Example

In Russell et al. [8] a theorem to find locally D-optimal designs for Poisson regression on a cuboid is given. We only want to consider the 2-dimensional case with $f(x_i)^\top \beta = (1, x_{1i}, x_{2i})(\beta_0, \beta_1, \beta_2)^\top = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ on the square design region $[-1, 1]^2$. With an initial guess $(\tilde{\beta}_0, 1, 2)$, $\tilde{\beta}_0$ arbitrary, for the parameter vector β a (locally) D-optimal design is given by the 3 equally weighted support points $(1, 1)$, $(-1, 1)$ and $(1, 0)$.

A 90° rotation g of the design space induces a transformation of f with $f\{g(x)\} = Q_g f(x)$ and

$$Q_g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Because $(Q_g^\top)^{-1} = Q_g$, $\tilde{g}(\beta) = Q_g \beta = (\beta_0, -\beta_2, \beta_1)^\top$ and the transformed initial guess $\tilde{g}\{(\tilde{\beta}_0, 1, 2)^\top\} = (\beta_0, -2, 1)^\top$. The transformed design contains the 3 equally weighted support points $(-1, 1)$, $(-1, -1)$ and $(0, 1)$, which is identical to the (locally) D-optimal design given by Russell et al. [8].

4 Weighted Optimality and Maximin Designs

Locally optimal designs depend on the initial values of the parameter β. A wrong choice, like an initial guess which is too far from the real parameter, may show a bad performance. So we regard two concepts to avoid this problem: the weighted

optimality or (pseudo)-Bayesian designs and the maximin approach. (see Atkinson et al. [1])

If a prior (probability) distribution π on the parameter space $\mathscr{B}$ with σ-algebra $\sigma_{\mathscr{B}}$ is available, a Bayesian optimal design can be found by maximizing

$$\Psi_\pi(\xi) = \int_{\mathscr{B}} \Phi\{M(\xi,\beta)\}\pi(\mathrm{d}\beta)$$

with a criterion function Φ on the space of all information matrices.

If moreover Φ is invariant for all $g \in G$, used in $(*)$, the Bayesian design criterion is equivariant, too. In this case it means, if ξ is optimal with respect to the prior distribution π, the transformed design ξ^g will be optimal for $\pi^{\tilde{g}}$.

$$\begin{aligned}
\Psi_{\pi^{\tilde{g}}}(\xi^g) &= \int_{\mathscr{B}} \Phi\{M(\xi^g,\beta)\}\pi^{\tilde{g}}(\mathrm{d}\beta) \\
&= \int_{\mathscr{B}} \Phi(M[\xi^g, \tilde{g}\{\tilde{g}^{-1}(\beta)\}])\pi^{\tilde{g}}(\mathrm{d}\beta) \\
&\overset{(*)}{=} \int_{\mathscr{B}} \Phi[M\{\xi,\tilde{g}^{-1}(\beta)\}]\pi^{\tilde{g}}(\mathrm{d}\beta) \\
&= \int_{\tilde{g}^{-1}(\mathscr{B})} \Phi\{M(\xi,\beta)\}(\pi^{\tilde{g}})^{\tilde{g}^{-1}}(\mathrm{d}\beta) \\
&\overset{(**)}{=} \int_{\mathscr{B}} \Phi\{M(\xi,\beta)\}\pi(\mathrm{d}\beta) \\
&= \Psi_\pi(\xi).
\end{aligned}$$

One assumption used in $(**)$ is that the parameter space $\mathscr{B}$ is invariant under $\tilde{g}$ or $\tilde{g}^{-1}$, that is $\tilde{g}^{-1}(\mathscr{B}) = \mathscr{B}$.

The prior distribution π is invariant under g, respectively $\tilde{g}$, if $\pi^{\tilde{g}} = \pi$, or to be precise, $\pi\{\tilde{g}^{-1}(B)\} = \pi(B)$ for all $B \in \sigma_{\mathscr{B}}$. As a consequence the Bayesian design is invariant under transformations of G, i.e. $\Psi_\pi(\xi^g) = \Psi_\pi(\xi)$.

In the case of the Bayesian D-criterion

$$\Psi_D(\xi) = \int_{\mathscr{B}} \log[\det\{M(\xi,\beta)\}]\pi(\mathrm{d}\beta),$$

which is concave according to Firth and Hinde [2], the prerequisites in Sect. 2 for the symmetrization are fulfilled. Hence the optimal designs can be found in the class of invariant designs.

The second possibility to avoid the dependence of designs on the unknown value of the parameter β is the maximin design which corresponds to the invariant criterion Φ and maximizes

$$\Psi(\xi) := \inf_{\beta\in\mathscr{B}} \Phi\{M(\xi,\beta)\}.$$

If $\mathscr{B}$ is also invariant under $\tilde{g}$, or equivalently under $\tilde{g}^{-1}$, then

$$\begin{aligned}\Psi(\xi^g) &= \inf_{\beta\in\mathscr{B}} \Phi\{M(\xi^g,\beta)\} = \inf_{\beta\in\mathscr{B}} \Phi(M[\xi^g,\tilde{g}\{\tilde{g}^{-1}(\beta)\}])\\ &= \inf_{\beta\in\mathscr{B}} \Phi[M\{\xi,\tilde{g}^{-1}(\beta)\}] = \inf_{\beta\in\tilde{g}^{-1}(\mathscr{B})} \Phi\{M(\xi,\beta)\} = \inf_{\beta\in\mathscr{B}} \Phi\{M(\xi,\beta)\}\\ &= \Psi(\xi).\end{aligned}$$

Graßhoff and Schwabe [5] used this fact, not in the direct sense of optimality, but of efficiency of maximin designs for the Bradley-Terry paired comparison model. The efficiency with respect to β corresponding to the criterion Φ has the form

$$\mathrm{eff}_\Phi(\xi,\beta) = \frac{h[\Phi\{M(\xi,\beta)\}]}{h[\Phi\{M(\xi^*_\beta,\beta)\}]},$$

with a special concave function h and the locally (with respect to β) Φ-optimal design ξ^*_β. The function h is chosen so that $h\circ\Phi$ is homogeneous, i.e. $h\{\Phi(aM)\} = ah\{\Phi(M)\}$ for all information matrices M and $a\in\mathbb{R}$. For example in the D-optimal case $h(x) = x^{1/p}$ and so $\mathrm{eff}_D(\xi,\beta) = \left[\frac{\det\{M(\xi,\beta)\}}{\det\{M(\xi^*_\beta,\beta)\}}\right]^{1/p}$.

If the criterion function Φ is concave, then the corresponding efficiency $\mathrm{eff}_\Phi(\xi,\beta)$ is obviously also concave in ξ for every β. As pointed out in Graßhoff and Schwabe [5], then the criterion of maximin efficiency shares the property of concavity. This together with the invariance of the maximin efficiency criterion

$$\Psi_{\mathrm{eff}_\Phi}(\xi) = \inf_{\beta\in\mathscr{B}} \frac{h[\Phi\{M(\xi,\beta)\}]}{h[\Phi\{M(\xi^*_\beta,\beta)\}]}$$

establishes that each design ξ is dominated by its symmetrization $\overline{\xi}$. Hence, globally maximin efficient designs can be found in the class of invariant designs if the parameter region $\mathscr{B}$ is invariant.

The invariance of the maximin efficiency criterion can easily be seen by

$$\begin{aligned}\Psi_{\mathrm{eff}_\Phi}(\xi^g) &= \inf_{\beta\in\mathscr{B}} \frac{h[\Phi\{M(\xi^g,\beta)\}]}{h[\Phi\{M(\xi^*_\beta,\beta)\}]} = \inf_{\beta\in\mathscr{B}} \frac{h(\Phi[M\{\xi^g,\tilde{g}(\beta)\}])}{h(\Phi[M\{\xi^*_{\tilde{g}(\beta)},\tilde{g}(\beta)\}])}\\ &= \inf_{\beta\in\mathscr{B}} \frac{h(\Phi[M\{\xi^g,\tilde{g}(\beta)\}])}{h(\Phi[M\{\xi^{*g}_\beta,\tilde{g}(\beta)\}])} = \inf_{\beta\in\mathscr{B}} \frac{h[\Phi\{M(\xi,\beta)\}]}{h[\Phi\{M(\xi^*_\beta,\beta)\}]} = \Psi_{\mathrm{eff}_\Phi}(\xi).\end{aligned}$$

5 Example (Continued)

We want to revisit the example in Sect. 3—the 2-dimensional Poisson regression without interactions. But the focus is to find a Bayesian D-optimal design containing exactly 3 equally weighted support points on the design region $[-1, 1]^2$. However, we do not want to find an optimal design over all possible designs. The prior distributions for β_1 and β_2 should be a uniform distribution being symmetrical around 0, that is $\beta_1, \beta_2 \sim \mathrm{U}(-a, a)$, i.i.d., with $a \in (0, \infty)$. The prior distribution for β_0 can be any distribution. Then it can be shown that these 3 points have the form $\{(-1, 1), (-1, -1), (1, d)\}$, $\{(-1, -1), (1, -1), (d, 1)\}$, $\{(1, -1), (1, 1), (-1, d)\}$ or $\{(1, 1), (-1, 1), (d, -1)\}$ with $d \in [-1, 1]$.

Consider the group G generated by the 90° rotation $g_{90} : \mathbb{R}^2 \to \mathbb{R}^2$, $g_{90}(x_1, x_2) = (-x_2, x_1)$, and the reflection $g_{\text{ref}} : \mathbb{R}^2 \to \mathbb{R}^2$, $g_{\text{ref}}(x_1, x_2) = (-x_1, x_2)$. Then there are 8 transformations in G. The design space is invariant under g and the prior distribution of the parameter space is invariant under $\tilde{g}$ for all $g \in G$. So the designs $\{(-1, 1), (-1, -1), (1, d)\}$, $\{(-1, 1), (-1, -1), (1, -d)\}$, $\{(-1, -1), (1, -1), (d, 1)\}$, $\{(-1, -1), (1, -1), (-d, 1)\}$, $\{(1, -1), (1, 1), (-1, d)\}$, $\{(1, -1), (1, 1), (-1, -d)\}$, $\{(1, 1), (-1, 1), (d, -1)\}$ and $\{(1, 1), (-1, 1), (-d, -1)\}$ with fixed $d \in [-1, 1]$ have the same value of the Bayesian criterion function. And if one of these designs is optimal (in the set of all 3-point designs), then all of these designs are optimal (in the set of all 3-point designs).

Changing only the design region to $\{-1, 1\}^2$, the group G of transformations is only generated by g_{90} and the Bayesian D-optimal 3-point designs are $\{(-1, 1), (-1, -1), (1, -1)\}$, $\{(-1, -1), (1, -1), (1, 1)\}$, $\{(1, -1), (1, 1), (-1, 1)\}$ and $\{(1, 1), (-1, 1), (-1, -1)\}$. The symmetrized design consists of the 4 equally weighted support points $(1, 1)$, $(-1, 1)$,$(-1, -1)$ and $(1, -1)$. This is the only invariant design with respect to G, so it is optimal over all possible designs.

6 Discussion

Although locally optimal designs are not invariant, in general, symmetry considerations can be useful for weighted optimality or maximin criteria, when the weight function or the parameter region is invariant. This may substantially simplify the optimization problem since attention may be restricted to the class of invariant (*symmetric*) designs and helps to find analytical and numerical/algorithmic solutions of the design problem.

The above findings may be extended to infinite transformation groups like rotations on a circle or on a k-dimensional sphere. Then the averaging (*symmetrization*) has to be performed by the corresponding Haar (Lebesgue) measure which is uniform on the orbits. The resulting invariant designs are continuous and have to be discretized. For rotations this can typically be achieved without loss of information.

References

1. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)
2. Firth, D., Hinde, J.: On Bayesian D-optimum design criteria and the equivalence theorem in non-linear Models. J. R. Stat. Soc. Ser. B Methodol. **59**, 793–797 (1997)
3. Ford, I., Torsney, B., Wu, C.F.J.: The use of a canonical form in the construction of locally optimal designs for non-linear problems. J. R. Stat. Soc. Ser. B Methodol. **54**, 569–583 (1992)
4. Gaffke, N., Heiligers, B.: Approximate designs for polynomial regression: invariance, admissibility, and optimality. In: Ghosh, S., Rao, C.R. (eds.) Handbook of Statistics, vol. 13, pp. 1149–1199. Elsevier, Amsterdam (1996)
5. Graßhoff, U., Schwabe, R.: Optimal designs for the Bradley-Terry paired comparison model. Stat. Methods Appl. **17**, 275–289 (2008)
6. McCulloch, C.E., Searle, S.R.: Generalized, Linear, and Mixed Models. Wiley, New York (2001)
7. Pukelsheim, F.: Optimal Design of Experiments. Wiley-Interscience, New York (1993)
8. Russell, K.G., Woods, D.C., Lewis, S.M., Eccleston, J.A.: D-optimal designs for Poisson regression models. Stat. Sin. **19**, 721–730 (2009)
9. Schmidt, D., Schwabe, R.: On optimal designs for censored data. Metrika **78**, 237–257 (2015)
10. Schwabe, R.: Optimum Designs for Multi-Factor Models. Springer, New York (1996)

Properties of the Random Block Design for Clinical Trials

Hui Shao and William F. Rosenberger

Abstract To avoid deterministic treatment allocations in the permuted block design (PBD), many clinical trialists prefer randomizing the block sizes as well. While such a procedure is rarely formalized, it is generally assumed that the design will be less predictable. In this paper, we formalize the random block design by assuming a discrete uniform distribution for block size. The aim of this study is to provide a statistical understanding of the RBD, by investigating its distributional properties, including the degree of predictability and variability of treatment imbalance.

1 Introduction

The permuted block design (PBD) is the most popular randomization procedure. In a PBD procedure, first a number of blocks with equal even block size is established, then the treatment assignments are randomized within each block. The main advantage of using the PBD is that it assures balance or approximately balance for each treatment group throughout the course of the trial, especially when the sample size is small. A drawback of this design is that one or more treatment assignments in each block are deterministic and predictable when the trial is not blinded. Therefore, the random block design (RBD), in which the block size is randomly selected from a sequence of even integers, is proposed to reduce the predictability of future assignments and achieve balanced treatment allocation. However, this procedure is rarely carefully defined when it is employed, [5, 7] and there is no theoretical support that the RBD is less predictable than the PBD. We formalize the procedure by selecting block sizes according to a discrete uniform distribution on the even integers, and quantify the predictability of the RBD in two-arm clinical trials.

H. Shao (✉) • W.F. Rosenberger
Department of Statistics, George Mason University, 4400 University Drive, MS 4A7, Fairfax, VA 22030, USA
e-mail: hshao2@gmu.edu; wrosenbe@gmu.edu

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_26

2 Random Block Design

We formalize the random block design as follows. Blocks sizes are randomly selected from even integers $2, 4, \ldots, 2B_{max}$ with *an equal probability 1/*B_{max}, where B_{max} is predefined by investigators, and then patients are randomized within each block. Each block can be filled by either the random allocation rule (RAR) or truncated binomial design (TBD). A random allocation rule can be considered as drawing balls without replacement from an urn containing equal numbers of two types of ball. A truncated binomial rule first uses a probability of $1/2$ until half the patients have been assigned to one of the two treatments, then assigns the remaining patients to the opposite treatment. When referencing the RBD, it is necessary to indicate the range of block size, in other words, the value of B_{max}. It is also necessary to indicate by which method each block is filled. We refer to each design as RBD(B_{max}; R) for those filled using the RAR and RBD(B_{max}; T) for those filled using the TBD. In this paper, we only focus on the RBD(B_{max}; R); reference [8] gives results for the RBD($B_{\max}$; T). The following is the notation used in this paper:

- There are two treatments, E ("experimental") and C ("control").
- n is the total number of patients. We do not assume all the blocks are filled.
- T_j is treatment assignment for the jth patient. T_j=1 if the jth patient is assigned to treatment E and 0 if the patient is assigned to treatment C.
- $2B_j$ is the block size of the jth subject, $2B_j = 2, \ldots, 2B_{max}$.
- R_j is the position number within the block, and takes the values $1, \ldots, 2B_j$. For example, if the block sizes are 4, 2, 4, $R_8 = 2$. If $R_n = 2B_n$, then every block is filled; otherwise, the last block is unfilled.
- $N_E(j)$ and $N_C(j)$ are the total number of patients assigned to treatment E and C after assigning the jth patient, respectively.
- $D_j = N_E(j) - N_C(j)$ is the imbalance in the number of patients after assigning the jth patient.
- ϕ_j is the probability assigning the jth patient to treatment E.
- $\mathbb{I}$ is the indicator function.
- $\lfloor a \rfloor$ denotes the integer part of a.

3 Exact Distribution of D_j

Throughout all the mathematical development, we will treat summations as 0 if the upper limit of the summation is smaller than the lower limit. All the proofs are by induction and can be found in [8].

Lemma 1 *Let b be an integer from 1 to* B_{max}. *Let r be an integer from 1 to* 2*b. Let r and j have the same parity. For* $j \leq 2B_{max} + 2$, *the joint distribution of* B_j *and* R_j *for*

the RBD is given by

$$P(R_j = r, B_j = b) = \begin{cases} \dfrac{1}{B_{max}}, & 1 \le r = j \le 2b, \\ \dfrac{1}{B_{max}(B_{max}+1)} \left(1 + \dfrac{1}{B_{max}}\right)^{\frac{j-r}{2}}, & 1 \le r \le \min(j-2, 2b), \\ 0, & r > \min(j, 2b). \end{cases}$$

The limiting joint distribution of B_j and R_j is given by

$$\lim_{j\to\infty} P(R_j = r, B_j = b) = \frac{2}{B_{max}(B_{max}+1)}.$$

Theorem 1 *Let d be a nonnegative integer from 0 up to B_{max}. Let d and j have the same parity. For $j \le 2B_{max} + 2$, the distribution of D_j for RBD(B_{max}; R) is given by*

$$P(D_j = \pm d) = \sum_{b=1}^{B_{max}} \sum_{\substack{r=1 \\ r,j \text{ have the} \\ \text{same parity}}}^{\min(2b,j)} \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{1}{B_{max}} \times \left\{ \frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}}\right)^{\frac{j-r}{2}} \right\}^{\mathbb{I}_{(r<j)}} \mathbb{I}_{[d \le \min(r, 2b-r)]}.$$

For $j \ge 2B_{max} + 3$, the distribution of D_j is given by

$$P(D_j = \pm d) \approx \sum_{b=1}^{B_{max}} \sum_{\substack{r=1 \\ r,j \text{ have the} \\ \text{same parity}}}^{2b} \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{2}{B_{max}(B_{max}+1)} \mathbb{I}_{[d \le \min(r, 2b-r)]}.$$

4 Selection Bias

Blackwell and Hodges developed a simple model to measure the potential for selection bias [2]. They proposed a convergence strategy for guessing the upcoming assignment, which was to guess the treatment that has fewer prior allocations, or to guess one of the treatments if both treatments have an equal number of prior allocations. Their model calculates the expected selection bias factor, $E(F)$, which is the expected excess number of correct guesses of treatment assignments beyond that expected by chance when the investigator uses the convergence strategy. It is

also equivalent to half of the difference between the expected number of correct and incorrect guesses among all the guesses made when the two treatment groups have different prior assignments.

Based on the fact that a larger degree of randomness corresponds to less predictability, we propose a metric for measuring the degree of randomness of a restricted randomization procedure, which is the degree of predictability

$$\rho_{PRED} = \sum_{j=1}^{n} E\left|\phi_j - \frac{1}{2}\right|.$$

This is similar to the metric proposed by Chen [3] and Berger [1]. This degree of predictability describes the expected deviation of a randomization procedure from complete randomization. A large value of this indicates a low degree of randomness, hence a high chance of correct prediction.

It turns out that the degree of predictability is mathematically equivalent to Blackwell-Hodges expected selection bias factor. Now call a correct guess a hit and an incorrect guess a miss when the two treatment arms have different prior allocations. Letting H and M denote the number of hits and misses, respectively, then the expected number of hits and misses, given $\phi_1, \ldots, \phi_n$ is calculated as

$$E(H|\phi_1, \ldots, \phi_n) = \sum_{j=1}^{n} \left\{ \frac{1}{2}\mathbb{I}\left(\phi_j = \frac{1}{2}\right) + \phi_j \mathbb{I}\left(\phi_j > \frac{1}{2}\right) + (1-\phi_j)\mathbb{I}\left(\phi_j < \frac{1}{2}\right) \right\},$$

and

$$E(M|\phi_1, \ldots, \phi_n) = \sum_{j=1}^{n} \left\{ \frac{1}{2}\mathbb{I}\left(\phi_j = \frac{1}{2}\right) + (1-\phi_j)\mathbb{I}\left(\phi_j > \frac{1}{2}\right) + \phi_j \mathbb{I}\left(\phi_j < \frac{1}{2}\right) \right\}.$$

The expected selection bias factor is then given by

$$\begin{aligned} E(F) =& EE(F|\phi_1, \ldots, \phi_n) = E\left(\frac{E(H|\phi_1, \ldots, \phi_n) - E(M|\phi_1, \ldots, \phi_n)}{2}\right) \\ =& E\sum_{j=1}^{n} \left\{ \left(\phi_j - \frac{1}{2}\right)\mathbb{I}\left(\phi_j > \frac{1}{2}\right) + \left(\frac{1}{2} - \phi_j\right)\mathbb{I}\left(\phi_j < \frac{1}{2}\right) \right\} \\ =& \sum_{j=1}^{n} E\left|\phi_j - \frac{1}{2}\right|. \end{aligned}$$

Theorem 2 *The ρ_{PRED} in n, $n \geq 1$, trials for the RBD(B_{max}; R) is given by*

$$\rho_{PRED} \approx \sum_{j=1}^{\min(n,2B_{max}+2)} \sum_{b=1}^{B_{max}} \left\{ \sum_{\substack{r=1 \\ r,j \text{ have the} \\ \text{same parity}}}^{\min(b,j)} \sum_{t=0}^{r-1} \left| \frac{b-t}{2b-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}} \right.$$

$$\left. + \sum_{\substack{r=b+1 \\ r,j \text{ have the} \\ \text{same parity}}}^{\min(2b,j)} \sum_{t=r-b-1}^{b} \left| \frac{b-t}{2b-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}} \right\}$$

$$\times \frac{1}{B_{max}} \left(\frac{1}{B_{max}+1} \left(1 + \frac{1}{B_{max}} \right)^{\frac{j-r}{2}} \right)^{\mathbb{I}(r<j)}$$

$$+ \sum_{j=2B_{max}+3}^{n} \sum_{b=1}^{B_{max}} \left\{ \sum_{\substack{r=1 \\ r,j \text{ have the} \\ \text{same parity}}}^{b} \sum_{t=0}^{r-1} \left| \frac{b-t}{2b-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}} \right.$$

$$\left. + \sum_{\substack{r=b+1 \\ r,j \text{ have the} \\ \text{same parity}}}^{2b} \sum_{t=r-b-1}^{b} \left| \frac{b-t}{2b-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2b-r+1}{b-t}}{\binom{2b}{b}} \right\} \frac{2}{B_{max}(B_{max}+1)}.$$

We refer to the PBD with block size $2B$ using the RAR to fill each block as PBD(B; R). Theorem 2 can be simplified for the degree of predictability of the PBD(B; R).

Theorem 3 *The ρ_{PRED} in n, $n \geq 1$, trials for the PBD(B; R) is given by*

$$\rho_{PRED} = \left\lfloor \frac{n}{2B} \right\rfloor \left\{ \frac{2^{2B-1}}{\binom{2B}{B}} - \frac{1}{2} \right\}$$

Table 1 Degree of predictability of the RBD and PBD for different values of n, B_{max} and B

	RBD(B_{max}; R)						PBD(B; R)					
n	2	3	4	5	6	7	2	3	4	5	6	7
5	0.87	0.76	0.66	0.58	0.52	0.47	0.83	0.60	0.40	0.30	0.24	0.20
10	2.14	1.90	1.71	1.60	1.45	1.32	1.83	1.50	1.40	1.53	0.99	0.77
15	3.08	2.78	2.55	2.36	2.21	2.09	2.83	2.40	2.16	1.83	1.81	1.89
20	4.36	3.92	3.59	3.33	3.11	2.93	4.17	3.40	2.93	3.06	2.33	2.18
25	5.31	4.81	4.43	4.13	3.88	3.66	5.00	4.40	3.99	3.37	3.43	2.81
50	11.03	10.01	9.24	8.62	8.12	7.69	10.17	8.90	8.04	7.66	6.91	6.16
75	16.42	14.95	13.84	12.95	12.22	11.60	15.33	13.40	12.10	11.02	10.39	9.64
100	22.14	20.15	18.64	17.45	16.46	15.64	20.83	18.00	16.21	15.32	13.90	13.25

$$+ \sum_{\substack{r=1 \\ r,j \text{ have the} \\ \text{same parity}}}^{m} \left\{ \mathbb{I}_{(r \le B)} \sum_{t=0}^{r-1} \left| \frac{B-t}{2B-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2B-r+1}{B-t}}{\binom{2B}{B}} \right.$$

$$\left. + \mathbb{I}_{(r \ge B+1)} \sum_{t=r-B-1}^{B} \left| \frac{B-t}{2B-r+1} - \frac{1}{2} \right| \frac{\binom{r-1}{t}\binom{2B-r+1}{B-t}}{\binom{2B}{B}} \right\},$$

where $m = n - 2B \left\lfloor \frac{n}{2B} \right\rfloor$.

The degree of predictability of the RBD and PBD for different values of n, B_{max} and B is provided in Table 1. As expected, the degree of predictability decreases with the block size. It also shows that when $B_{max} = B$, the degree of predictability of the RBD(B_{max}; R) is higher than the PBD(B; R). This agrees with the simulation results of Zhao et al. [9]. In order to generate an allocation sequence with degree of predictability comparable to the PBD(B; R), a larger value of B_{max} has to be selected for the RBD(B_{max}; R).

5 Balancing Properties

We investigate the balancing properties of the RBD by calculating the variance of the terminal treatment imbalance, $Var(D_n)$. For a RBD with block sizes ranging from 2 to $2B_{max}$, the final imbalance D_n varies from $-B_{max}$ to B_{max}. A small variance of D_n indicates that repeating the RBD produces similar values of imbalance.

Theorem 4 *Let $n \geq 1$. The variance of the terminal imbalance of the RBD(B_{max}; R) is given by the following.*
If $n \leq 2B_{max} + 2$,

$$Var(D_n) = 2\sum_{b=1}^{B_{max}} \sum_{\substack{r=1 \\ r,n \text{ have the} \\ \text{same parity}}}^{\min(2b,n)} \sum_{\substack{d=0 \\ d,n \text{ have the} \\ \text{same parity}}}^{\min(r,2b-r)} \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \times \frac{d^2}{B_{max}} \left\{ \frac{1}{B_{max}+1}\left(1+\frac{1}{B_{max}}\right)^{\frac{n-r}{2}} \right\}^{\mathbb{I}_{(r<n)}}.$$

If $n \geq 2B_{max} + 3$,

$$Var(D_n) \approx 2\sum_{b=1}^{B_{max}} \sum_{\substack{r=1 \\ r,n \text{ have the} \\ \text{same parity}}}^{2b} \sum_{\substack{d=0 \\ d,n \text{ have the} \\ \text{same parity}}}^{\min(r,2b-r)} d^2 \frac{\binom{r}{\frac{d+r}{2}}\binom{2b-r}{b-\frac{d+r}{2}}}{\binom{2b}{b}} \frac{2}{B_{max}(B_{max}+1)}.$$

Theorem 5 *Let $n \geq 1$. The variance of terminal imbalance of the PBD (B_{max}; R) is given by the following.*

$$Var(D_n) = 2 \sum_{\substack{d=0 \\ d,n \text{ have the} \\ \text{same parity}}}^{\min(m,2B-m)} d^2 \frac{\binom{m}{\frac{m+d}{2}}\binom{2B-m}{B-\frac{m+d}{2}}}{\binom{2B}{B}},$$

where $m = n - 2B\left\lfloor \frac{n}{2B} \right\rfloor$.

The variance of the imbalance of the RBD and PBD for different values of n, B_{max}, and B is provided in Table 2. For the RBD, the variance of the imbalance is an increasing function of B_{max} for each n. The variance of the imbalance for the PBD is 0 when all blocks are filled. When $B_{max} = B$, the overall variance of the imbalance of the RBD is smaller than that of the PBD.

We plot the variance of the terminal imbalance of the RBD(5; R) and PBD(3; R) for different values of n in Fig. 1, since these two procedures generate treatment allocations with similar degrees of predictability. We see the variance of the terminal imbalance of the RBD is more stable than that of the PBD. The PBD(3; R) has on average smaller variance of imbalance than the RBD(5; R).

Table 2 Variance of the imbalance of the RBD and PBD for different values of n, B_{max} and B

	RBD(B_{max}; R)						PBD(B; R)					
n	2	3	4	5	6	7	2	3	4	5	6	7
5	1.00	1.09	1.41	1.72	1.98	2.21	1.00	1.00	2.14	2.78	3.18	3.46
10	0.44	0.76	1.08	1.10	1.40	1.78	1.33	1.60	1.71	0	1.82	3.08
15	1.00	1.13	1.31	1.50	1.70	1.84	1.00	1.80	1.00	2.78	2.45	1.00
20	0.44	0.76	1.02	1.28	1.52	1.75	0	1.60	2.29	0	2.91	3.69
25	1.00	1.13	1.31	1.50	1.70	1.91	1.00	1.00	1.00	2.78	1.00	2.54
50	0.44	0.76	1.02	1.28	1.52	1.75	1.33	1.60	1.71	0	1.82	3.69
75	1.00	1.13	1.31	1.50	1.70	1.91	1.00	1.80	2.14	2.78	2.45	3.46
100	0.44	0.76	1.02	1.28	1.52	1.75	0	1.60	2.29	0	2.91	1.85

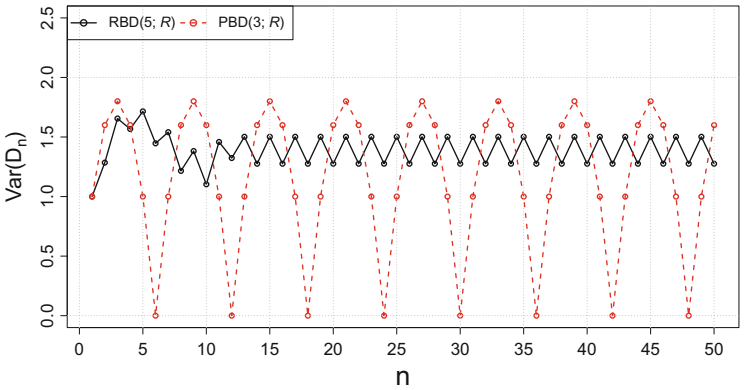

Fig. 1 Variance of the imbalance as a function of n

6 Discussion

In this study, we have formalized the RBD and quantified its degree of predictability and balancing properties in closed-form formulas. This provides a statistical understanding of the RBD that can be used in comparisons with other restricted randomization procedures.

Compared with the PBD, the RBD does not reduce predictability as one might expect. One has to choose a relatively larger value of B_{max} for the RBD(B_{max}; R) to produce a comparable or less predictable allocation sequence than that under the PBD(B; R). In this case, the PBD has an overall relative smaller variance of terminal imbalance than the RBD. If the total number of patients is known, the clinical trialist can choose a PBD design with 0 variance for terminal imbalance. Alternatively, if one can establish a fixed sample size that is divisible by all possible realizations of the block sizes that result in all blocks being filled. from the RBD, one can eliminate the possibility of imbalance [4].

Matts and Lachin concluded that the selection bias factor of a design employing multiple block sizes is approximately the same as that of a design with a block size equal to the average of all the block sizes of the former design [6]. For example, the selection bias factor of the RBD(3; R) is approximately the same as that of the PBD(2; R), and the selection bias factor of the RBD(5; R) is approximately the same as that of the PBD (3; R). Our results confirm their conclusion and show that the degree of predictability of RBD(3; R) is slightly smaller than that of the PBD (2; R).

Acknowledgements Professor Rosenberger was supported by a scholarship from the German-American Fulbright Kommission, and also by the Department of Medical Statistics, RWTH Aachen University.

References

1. Berger, V.W.: Selection Bias and Covariate Imbalance in Randomized Clinical Trials. Wiley, Chichester (2005)
2. Blackwell, D., Hodges, J.L.: Design for the control of selection bias. Ann. Math. Stat. **28**, 449–460 (1957)
3. Chen, Y.P.: Which design is better? Ehrenfest urn versus biased coin. Adv. Appl. Probab. **32**, 738–749 (2000)
4. Heussen, N.: Der Einfluss der Randomisierung in Blöcken zufälliger Länge auf die Auswertung klinischer Studien mittels Randomisationstest. RWTH Aachen University, Aachen (2004)
5. Macian, N., Pereira, B., Shinjo, C., Dubray, C., Pickering, G.: Fibromyalgia, milnacipran and experimental pain modulation: study protocol for a double blind randomized controlled trial. Trials **16**, 134 (2015)
6. Matts, J.P., Lachin, J.M.: Properties of permuted-block randomization in clinical trials. Control. Clin. Trials **9**, 327–344 (1988)
7. Shah, B., Berger, J.S., Amoroso, N.S., Mai, X., Lorin, J.D., Danoff, A., Schwartzbard, A.Z., Lobach, I., Guo, Y., Feit, F., Slater, J., Attubato, M.J., Sedlis, S.P.: Periprocedural glycemic control in patients with diabetes mellitus undergoing coronary angiography with possible percutaneous coronary intervention. Am. J. Cardiol. **113**, 1474–1480 (2014)
8. Shao, H.: Exact properties of restricted randomization procedures. George Mason University, Fairfax (2015)
9. Zhao, W., Weng, Y., Wu, Q., Palesch, Y.: Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. Pharm. Stat. **11**, 39–48 (2012)

Functional Data Analysis in Designed Experiments

Bairu Zhang and Heiko Großmann

Abstract F-type tests for functional ANOVA models implicitly assume that the response curves are generated by a completely randomized design. By using the split-plot design as an example it is illustrated how these tests can be extended to more complex ANOVA models. In order to derive the test statistics and their approximate null distributions, Hasse diagrams for representing the structure of the experiment are combined with a stochastic process perspective. The application of the more general F-type tests is illustrated for simulated data.

1 Introduction

Functional data arise in designed experiments when the response for an experimental unit is a curve that is densely sampled over a continuum. Despite this more complicated response the basic structure of the experiment in terms of treatment allocation is often relatively simple and essentially that of a regression or analysis of variance (ANOVA) model. In these models, the parameters are also functions and traditional F tests are no longer appropriate.

In [4, 11] and [12], functional F tests or F-type tests have been developed for testing fixed effects in functional linear models. This work adopts the so-called stochastic process perspective [6] in which the response curves are regarded as sample paths of a stochastic process and the null distribution of the test statistic is derived by using the Karhunen-Loève expansion [7] and Satterthwaite's approximation [10]. However, these results essentially assume that the response curves are obtained from a completely randomized experiment.

In this paper, we illustrate how, by applying well-known experimental design principles, functional data from more complex experiments can be analyzed under the stochastic process perspective. In particular, we are interested in ANOVA models

B. Zhang (✉)
Queen Mary University of London, London, UK
e-mail: bairu.zhang@qmul.ac.uk

H. Großmann
Otto-von-Guericke-University Magdeburg, Magdeburg, Germany
e-mail: heiko.grossmann@ovgu.de

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_27

for functional data which take the structure of the observational units into account. Although the method in this paper is applicable to more complex situations, we focus on split-plot designs with two treatment factors in different strata as an example.

The basic idea is to adapt Bailey's ANOVA approach [2] to the functional data setting and to derive F-type tests for testing fixed effects in different strata. More precisely, we use Hasse diagrams to develop a mixed-effects model for functional data and to derive the analysis of variance table. We then apply the approach in [12] in order to derive the null distributions of the F-type tests in the different strata. Mixed-effects functional models have been considered before [1, 3, 5, 8, 9], but none of these papers has emphasized the link between the design and the analysis.

The remainder of the paper is organized as follows. Section 2 briefly recaps the structure of the split-plot design and presents the corresponding mixed-effects functional model and the ANOVA table. F-type tests of the treatment effects are derived in Sect. 3 and applied to simulated functional data in Sect. 4. The final Sect. 5 offers some conclusions.

2 Model and Analysis for Split-Plot Design

A classic split-plot design (for example, see [2]) has b large blocks, each of which consists of s small blocks (whole-plots), each of which contains m observational units (sub-plots). Thus the total number of observational units is $N = bsm$. There are two treatment factors H with s levels and A with m levels. Within each block, each level of H is applied to one entire whole-plot and, within each whole-plot, each level of A is applied to one sub-plot. The Hasse diagrams [2] representing the structure of the experiment are shown in Fig. 1.

In the functional data setting, the N responses are curves $Y_{ijk}(t)$ with arguments t in a real interval $\mathcal{T}$. If we assume that all blocks and whole-plots have random

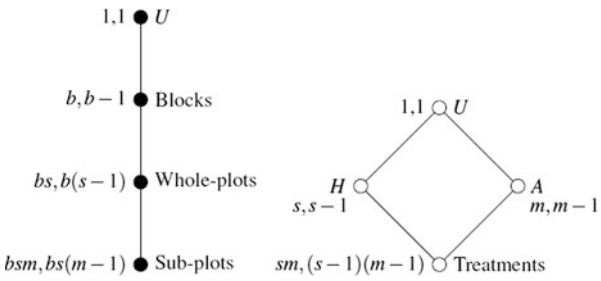

Fig. 1 Hasse diagrams for plot structure (*left*) and treatment structure (*right*) of split-plot design

effects, then the mixed-effects functional model can be expressed as

$$Y_{ijk}(t) = \mu(t) + \tau_j(t) + \omega_k(t) + \phi_{jk}(t) + \beta_i(t) + \alpha_{j(i)}(t) + \epsilon_{k(ij)}(t), \tag{1}$$

where $i = 1, \ldots, b, j = 1, \ldots, s$ and $k = 1, \ldots, m$. The terms $\mu(t)$, $\tau_j(t)$, $\omega_k(t)$ and $\phi_{jk}(t)$ respectively indicate an overall mean function and functions representing the fixed effect of the jth level of factor H, the fixed effect of the kth level of factor A and the interaction of these two levels. Moreover, $\beta_i(t)$, $\alpha_{j(i)}(t)$ and $\epsilon_{k(ij)}(t)$ represent functions for the random effect of the ith block, the random effect of the jth whole-plot, nested in the ith block, and the error term, which is equal to the random effect of the kth sub-plot of whole-plot j within block i. We assume that $\beta_i(t)$, $\alpha_{j(i)}(t)$ and $\epsilon_{k(ij)}(t)$ are independent Gaussian processes with zero mean and smooth covariance functions $D(s,t) = cov[\beta_i(s), \beta_i(t)]$, $P(s,t) = cov[\alpha_{j(i)}(s), \alpha_{j(i)}(t)]$ and $\gamma(s,t) = cov[\epsilon_{k(ij)}(s), \epsilon_{k(ij)}(t)]$. These can be summarized as $\beta_i(t) \sim GP(0, D)$, $\alpha_{j(i)}(t) \sim GP(0, P)$ and $\epsilon_{k(ij)} \sim GP(0, \gamma)$, where GP stands for 'Gaussian process'.

At each $t \in \mathcal{T}$, the ANOVA table for the model represented by the Hasse diagrams in Fig. 1 and Equation (1) can be calculated as described in [2, p. 151] and is shown in Table 1. In the table, $\bar{y}_{\cdots}(t) = \frac{1}{bsm} \sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} y_{ijk}(t)$ and averages $\bar{y}_{i\cdot\cdot}(t), \bar{y}_{\cdot j\cdot}(t), \bar{y}_{\cdot\cdot k}(t), \bar{y}_{ij\cdot}(t), \bar{y}_{\cdot jk}(t)$ are calculated in a similar way. Moreover, $SS(t)$ and $EMS(t)$ denote functional sums of squares and functional expected mean squares,

Table 1 Analysis of variance table for split-plot design with functional response

Stratum	Source	df	$SS(t)$	$EMS(t)$
Mean	Mean	1	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \bar{y}_{\cdots}(t)^2$	–
Blocks	Blocks	$b-1$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{\bar{y}_{i\cdot\cdot}(t) - \bar{y}_{\cdots}(t)\}^2$	–
Whole-plots	H	$s-1$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{\bar{y}_{\cdot j\cdot}(t) - \bar{y}_{\cdots}(t)\}^2$	$\frac{bm}{s-1} \sum_{j=1}^{s} \tau_j(t)^2 +$ $\{mP(t,t) + \gamma(t,t)\}$
	Residual	$(b-1)(s-1)$	By subtraction	$mP(t,t) + \gamma(t,t)$
	Total	$b(s-1)$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{\bar{y}_{ij\cdot}(t) - \bar{y}_{i\cdot\cdot}(t)\}^2$	–
Sub-plots	A	$m-1$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{\bar{y}_{\cdot\cdot k}(t) - \bar{y}_{\cdots}(t)\}^2$	$\frac{bs}{m-1} \sum_{k=1}^{m} \omega_k(t)^2 + \gamma(t,t)$
	$A \wedge H$	$(s-1)(m-1)$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{\bar{y}_{\cdot jk}(t) -$ $\bar{y}_{\cdots}(t)\}^2 - SS_A(t) - SS_H(t)$	$\frac{b}{(m-1)(s-1)} \sum_{j=1}^{s} \sum_{k=1}^{m} \phi_{jk}(t)^2 +$ $\gamma(t,t)$
	Residual	$(b-1)s(m-1)$	By subtraction	$\gamma(t,t)$
	Total	$bs(m-1)$	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} \{y_{ijk}(t) - \bar{y}_{ij\cdot}(t)\}^2$	–
Total		bsm	$\sum_{i=1}^{b} \sum_{j=1}^{s} \sum_{k=1}^{m} y_{ijk}(t)^2$	

which are calculated as point-wise sums of squares and mean squares for every $t \in \mathscr{T}$. Similarly, the degrees of freedom in the third column of Table 1 are for a point-wise analysis with fixed $t \in \mathscr{T}$. As will be seen in the next section, these degrees of freedom are however different from those for the functional model (1) and the corresponding hypothesis tests.

3 Hypothesis Tests

F-type tests of fixed effects have been derived in [12] for functional ANOVA models under the implicit assumption that the response curves are obtained from a completely randomized design. Here we present corresponding results for model (1). The main difference in this setting is that treatment effects are not tested in a single but in two strata. More precisely, as for classic split-plot designs, the effect of treatment factor H is tested in the whole-plots stratum, whereas the effect of factor A and the interaction $H \wedge A$ of the two factors H and A are tested in the sub-plot stratum.

In all three cases the test statistics are based on integrated sums of squares from Table 1. Following the approach in [11, 12] the null distribution is derived by using the Karhunen-Loève expansion. Here we only present the results.

3.1 F-Type Test for Factor H

In order to test the effect of treatment factor H we test the null hypothesis that the s functions corresponding to the effects of the s levels of H are equal, that is we test $H_0 : \tau_1(t) = \cdots = \tau_s(t)$ for all $t \in \mathscr{T}$ against $H_1 : \neg H_0$. It can be shown that under the null hypothesis

$$\frac{\int SS_H(t)dt/(s-1)}{\int SS_{\textit{whole-plot residual}}(t)dt/(b-1)(s-1)} \overset{\textit{approx.}}{\sim} F(df_H, df_{res}), \tag{2}$$

where $SS_H(t)$ and $SS_{\textit{whole-plot residual}}(t)$ denote the functional sums of squares for H and the residual within the whole-plot stratum in Table 1.

The degrees of freedom of the F distribution in (2) are equal to

$$df_H = \frac{(\sum_{r=1}^{\infty} \lambda_r)^2}{\sum_{r=1}^{\infty} \lambda_r^2}(s-1) \quad \text{and} \quad df_{res} = \frac{(\sum_{r=1}^{\infty} \lambda_r)^2}{\sum_{r=1}^{\infty} \lambda_r^2}(b-1)(s-1),$$

where $\lambda_1, \ldots, \lambda_\infty$ are the eigenvalues of the covariance function $mP(s,t) + \gamma(s,t)$ with $s, t \in \mathscr{T}$. Following the approach of [12], $mP(s,t)+\gamma(s,t)$ can be estimated by

$$\frac{\sum_{i=1}^{b}\sum_{j=1}^{s}\sum_{k=1}^{m}\{\bar{y}_{ij.}(s) - \bar{y}_{i..}(s) - \bar{y}_{.j.}(s) + \bar{y}_{...}(s)\}\{\bar{y}_{ij.}(t) - \bar{y}_{i..}(t) - \bar{y}_{.j.}(t) + \bar{y}_{...}(t)\}}{(b-1)(s-1)}.$$

We now show how the approximation in (2) is derived. To this end, we use the following auxiliary result adapted from [12]: If $x_1(t), \ldots, x_n(t), t \in \mathscr{T}$, are independent Gaussian processes with zero mean function and common covariance function $\delta(s,t)$, then

$$\int [\sum_{i=1}^{n}\{x_i(t) - \bar{x}(t)\}^2]dt \sim \sum_{r=1}^{\infty}\theta_r \chi^2_{n-1},$$

where $\theta_1, \ldots, \theta_\infty$ are the eigenvalues of $\delta(s,t)$ and $\bar{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$.

Under the null hypothesis that $H_0 : \tau_1(t) = \cdots = \tau_s(t)$ for all $t \in \mathscr{T}$, $SS_H(t)$ in (2) can be calculated as

$$\begin{aligned} SS_H(t) =& \sum_{i=1}^{b}\sum_{j=1}^{s}\sum_{k=1}^{m}\{\bar{y}_{.j.}(t) - \bar{y}_{...}(t)\}^2 \\ =& bm\sum_{j=1}^{s}[\{\bar{\alpha}_{j(.)}(t) + \bar{\epsilon}_{.(.j)}(t)\} - \{\bar{\alpha}_{.(.)}(t) + \bar{\epsilon}_{.(..)}(t)\}]^2, \end{aligned}$$

where $\bar{\alpha}_{j(.)}(t)$, $\bar{\epsilon}_{.(.j)}(t)$, $\bar{\alpha}_{.(.)}(t)$ and $\bar{\epsilon}_{.(..)}(t)$ are obtained by taking averages over the subscripts represented by the dots. The assumptions in Sect. 2 imply that $\bar{\alpha}_{j(.)}(t) + \bar{\epsilon}_{.(.j)}(t), j = 1, \ldots, s$, are independent Gaussian processes with zero mean function and common covariance function $\frac{1}{bm}(mP + \gamma)$. By applying the auxiliary result, it follows that

$$\int SS_H(t)dt \sim \sum_{r=1}^{\infty}\lambda_r \chi^2_{s-1},$$

where $\lambda_1, \ldots, \lambda_\infty$ are the eigenvalues of the covariance function $mP(s,t) + \gamma(s,t)$. Similarly, we can prove that $\int SS_{\textit{small blocks residual}}(t)dt \sim \sum_{r=1}^{\infty}\lambda_r \chi^2_{(b-1)(s-1)}$, where $\lambda_1, \ldots, \lambda_\infty$ are the same eigenvalues as before.

As in [11], the ratio $\frac{\sum_{r=1}^{\infty}\lambda_r\chi^2_{s-1}/(s-1)}{\sum_{r=1}^{\infty}\lambda_r\chi^2_{(b-1)(s-1)}/(b-1)(s-1)}$ can be approximated by $F(df_1, df_2)$, where $df_1 = \frac{(\sum_{r=1}^{\infty}\lambda_r)^2}{\sum_{r=1}^{\infty}\lambda_r^2}(s-1)$ and $df_2 = \frac{(\sum_{r=1}^{\infty}\lambda_r)^2}{\sum_{r=1}^{\infty}\lambda_r^2}(b-1)(s-1)$, which gives the result in (2).

3.2 F-Type Tests for Factor A and for Interaction $H \wedge A$

Using a similar method, we can also derive test statistics and corresponding null distributions for treatment factor A and for the interaction of the two treatment factors. Under the null hypothesis $H_0 : \omega_1(t) = \cdots = \omega_m(t)$ for all $t \in \mathscr{T}$, we have

$$\frac{\int SS_A(t)dt/(m-1)}{\int SS_{\text{sub-plot residual}}(t)dt/(b-1)s(m-1)} \overset{approx.}{\sim} F(df_A, df_{res*}),$$

and under the null hypothesis $H_0 : \phi_{11}(t) = \cdots = \phi_{sm}(t)$ for all $t \in \mathscr{T}$, we have

$$\frac{\int SS_{H\wedge A}(t)dt/(s-1)(m-1)}{\int SS_{\text{sub-plot residual}}(t)dt/(b-1)s(m-1)} \overset{approx.}{\sim} F(df_{H\wedge A}, df_{res*}),$$

where $SS_A(t)$, $SS_{H\wedge A}(t)$ and $SS_{\text{sub-plot residual}}(t)$ denote functional sums of squares for treatment factor A, the interaction of H and A, and the residual within the sub-plot stratum. The degrees of freedom of the F distributions in the above equations are respectively equal to $df_A = \frac{(\sum_{r=1}^{\infty}\lambda_r^*)^2}{\sum_{r=1}^{\infty}\lambda_r^{*2}}(s-1)$, $df_{H\wedge A} = \frac{(\sum_{r=1}^{\infty}\lambda_r^*)^2}{\sum_{r=1}^{\infty}\lambda_r^{*2}}(s-1)(m-1)$ and $df_{res*} = \frac{(\sum_{r=1}^{\infty}\lambda_r^*)^2}{\sum_{r=1}^{\infty}\lambda_r^{*2}}(b-1)s(m-1)$, where $\lambda_1^*, \ldots, \lambda_\infty^*$ are the eigenvalues of the covariance function $\gamma(s,t)$ with $s,t \in \mathscr{T}$. Finally, $\gamma(s,t)$ can be estimated similarly to $mP(s,t) + \gamma(s,t)$.

4 Simulation Study

We illustrate the application of the tests in Sect. 3 with simulated data for a split-plot design with two blocks, three whole-plots per block and four sub-plots within each whole-plot. Values of the response curves are generated for 1000 equally spaced values t in the interval $\mathscr{T} = [0,1]$. The functions for generating the fixed effects are $\mu(t) = 2\sin(2\pi t)$, $\tau_j(t) = je^{(t-0.5)^2}$, $\omega_k(t) = e^{\sin(k\pi t)}$ and $\phi_{jk}(t) = \tau_j(t) \times \omega_k(t)$, where $i = 1,2$, $j = 1,2,3$ and $k = 1,2,3,4$. In the simulations, for simplicity, we use the covariance functions $D = 0.5\Lambda$, $P = 0.2\Lambda$ and $\gamma = 0.3\Lambda$, where $\Lambda(s,t) = e^{-|1000(s-t)|}$. The stochastic processes representing the random effects in Equation (1) are then simulated as $\beta_i(t) \sim GP(0, 0.5\Lambda)$, $\alpha_{j(i)}(t) \sim GP(0, 0.2\Lambda)$ and $\epsilon_{k(ij)}(t) \sim GP(0, 0.3\Lambda)$.

Figure 2 depicts the functional data from the simulation. Results of the functional analysis of variance are given in Table 2. Note that the integrated sums of squares in the fourth column of the table are approximated by summation over the 1000 values of t. The F-type tests show that the treatment factors and the interaction have significant effects on the response curves.

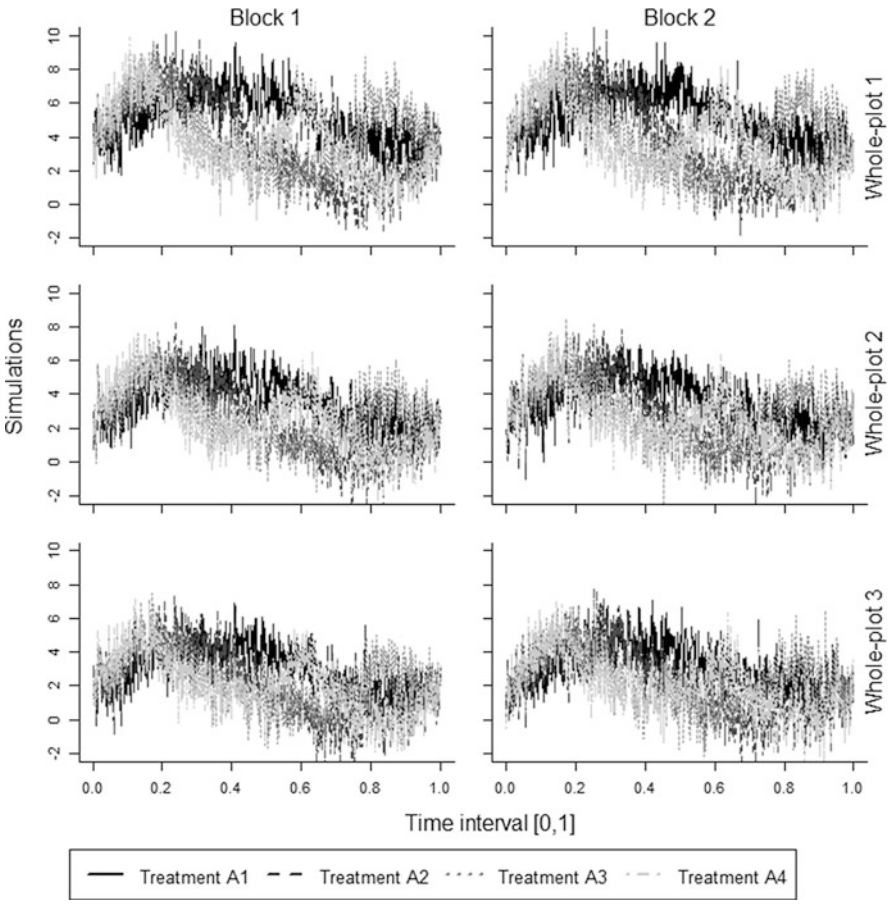

Fig. 2 Functional data simulated by using model (1) for split-plot design

Table 2 Analysis of variance table for simulated functional data

Stratum	Source	df	$\int SS(t)dt$	Hypothesis test
Mean	Mean	1	279,873.3	–
Blocks	Blocks	1	7337.6	–
Whole-plots	H	2	16,727.5	$7.6004 > F_{0.05}(3.98, 3.98)$
	Residual	2	2200.9	–
	Total	4	18,928.4	–
Sub-plots	A	3	36,045.3	$40.9802 > F_{0.05}(26.72, 80.16)$
	$A \wedge H$	6	2871.5	$1.6323 > F_{0.05}(53.44, 80.16)$
	Residual	9	2638.7	–
	Total	18	41,555.5	–
Total		24	347,694.8	

5 Conclusions

F-type tests in [4, 11] and [12] for functional ANOVA models using a completely randomized design can be extended to more complex ANOVA models. Here we have presented results for split-plot designs. By using the approach of [2] a point-wise ANOVA table can be derived from Hasse diagrams representing the structure of the observational units and the treatments.

The sums of squares and degrees of freedom in that table provide the basis for the test statistics of F-type tests which under the null hypothesis of no effects have an approximate F distribution. The degrees of freedom of this distribution are however different from the degrees of freedom in the point-wise ANOVA table. Both the numerator and the denominator of the test statistics are integrated sums of squares and in practice the corresponding integrals have to be approximated by taking sums over the values at which the functional data are observed.

An important difference from the results for completely randomized designs is that, depending on the structure of the experiment, the tests are performed in different strata using different integrated sums of squares in the denominator. Failing to account for the structure of the experiment may result in incorrect conclusions due to false replication. Although in this paper we have only considered the split-plot design, we conjecture that the approach presented here can be generalized to the whole class of orthogonal designs discussed in [2].

References

1. Antoniadis, A., Sapatinas, T.: Estimation and inference in functional mixed-effects models. Comput. Stat. Data Anal. **51**(10), 4793–4813 (2007)
2. Bailey, R.A.: Design of Comparative Experiments. Cambridge University Press, Cambridge (2008)
3. Fan, J.Q., Zhang, J.T.: Two-step estimation of functional linear models with applications to longitudinal data. J. R. Stat. Soc. Ser. B. **62**(2), 303–322 (2000)
4. Faraway, J.J.: Regression analysis for a functional response. Technometrics **39**(3), 254–261 (1997)
5. Guo, W.: Functional mixed effects models. Biometrics **58**(1), 121–128 (2002)
6. Hsing, T., Eubank, R.: Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. John Wiley, Sussex (2015)
7. Loève, M.: Probability Theory II. Springer, New York (1978)
8. Morris, J.S., Carroll, R.J.: Wavelet-based functional mixed models. J. R. Stat. Soc. Ser. B. **68**(2), 179–199 (2006)
9. Qin, L., Guo, W.: Functional mixed-effects model for periodic data. Biostatistics **7**(2), 225–234 (2006)
10. Satterthwaite, F.E.: Synthesis of variance. Psychometrika **6**(5), 309–316 (1941)
11. Shen, Q., Faraway, J.J.: An F test for linear models with functional responses. Stat. Sinica. **14**(4), 1239–1257 (2004)
12. Zhang, J.T.: Analysis of Variance for Functional Data. CRC, New York (2013)

Analysis and Design in the Problem of Vector Deconvolution

Anatoly Zhigljavsky, Nina Golyandina, and Jonathan Gillard

Abstract We formulate the problem of deconvolution of a given vector as an optimal design problem and suggest numerical algorithms for solving this problem. We then discuss an important application of the proposed methods for problems of time series analysis and signal processing and also to the low-rank approximation of structured matrices.

1 Introduction

For two vectors $\alpha = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}} \in \mathbb{R}^n$ and $\beta = (\beta_1, \ldots, \beta_m)^{\mathrm{T}} \in \mathbb{R}^m$, their convolution is the vector $\gamma = \alpha \star \beta = (\gamma_1, \ldots, \gamma_N)^{\mathrm{T}} \in \mathbb{R}^N$ of size $N = m + n - 1$ with elements

$$\gamma_i = \sum_j \alpha_j \beta_{i-j+1}\,; \quad i = 1, \ldots, N\,. \tag{1}$$

The summation range for j in (1) is defined so that the terms $\alpha_j \beta_{i-j+1}$ make sense; that is, $\max\{1, i-m+1\} \le j \le \min\{n, i\}$.

In this paper, we shall discuss the following problems.

A. Zhigljavsky (✉)
Cardiff University, Cardiff, UK

Lobachevskii Nizhnii Novgorod State University, Nizhny Novgorod, Russia
e-mail: ZhigljavskyAA@cardiff.ac.uk

N. Golyandina
St.Petersburg State University, St. Petersburg, Russia
e-mail: n.golyandina@spbu.ru

J. Gillard
Cardiff University, Cardiff, UK
e-mail: GillardJ@cardiff.ac.uk

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_28

Estimation Problem 1. Given $Y \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}^n \setminus \{0\}$, find

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 , \tag{2}$$

where $m = N + 1 - n$ and $\|\cdot\|$ is the usual Euclidean norm.

Estimation Problem 1a. Given $Y \in \mathbb{R}^N$ and $\beta \in \mathbb{R}^m \setminus \{0\}$, find

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^n}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 , \tag{3}$$

where $n = N + 1 - m$.

Estimation Problem 2. Given $Y \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}^n \setminus \{0\}$, find

$$\hat{\beta}_\delta = \underset{\beta \in \mathbb{R}^m_\delta}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 , \tag{4}$$

where $\mathbb{R}^m_\delta = \big\{\beta = (\beta_1, \ldots, \beta_m)^{\mathrm{T}} \in \mathbb{R}^m$ such that $\beta_j \geq \delta$ for all $j = 1, \ldots, m\big\}$. Here $\delta \in [0, 1)$ is some number (usually either 0 or a small positive number).

Estimation Problem 2a. Given $Y \in \mathbb{R}^N$ and $\beta \in \mathbb{R}^m \setminus \{0\}$, find

$$\hat{\alpha}_\delta = \underset{\alpha \in \mathbb{R}^n_\delta}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 , \tag{5}$$

where $\mathbb{R}^n_\delta = \big\{\alpha = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}} \in \mathbb{R}^n$ such that $\alpha_i \geq \delta$ for all $i = 1, \ldots, n\big\}$.

Design Problem 1. Given $Y \in \mathbb{R}^N \setminus \{0\}$ and $n < N$, find

$$(\hat{\alpha}, \hat{\beta}) = \underset{(\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}^m}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 . \tag{6}$$

Design Problem 2. Given $Y \in \mathbb{R}^N \setminus \{0\}$, $\delta \in \mathbb{R}$ and $n < N$, find

$$(\hat{\alpha}, \hat{\beta})_\delta = \underset{(\alpha,\beta) \in \mathbb{R}^n_\delta \times \mathbb{R}^m_\delta}{\operatorname{argmin}} \|Y - \alpha \star \beta\|^2 . \tag{7}$$

We believe that the most practically important, among the problems formulated above, is the problem (7). In Sect. 4, we show that the use of (approximate) solutions to the problem (7) could significantly increase accuracy of a wide range of methods of time series analysis and signal processing.

2 Solving the Estimation Problems

2.1 Matrix Form of the Convolution Operation

Let us express the convolution $\gamma = \alpha \star \beta$ in a matrix form. For given $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^m$, define the matrices $\mathbf{F}_\alpha \in \mathbb{R}^{N\times m}$ and $\mathbf{F}_\beta \in \mathbb{R}^{N\times n}$ as follows: for $i = 1, \ldots, m$, the i-th column of the matrix $\mathbf{F}_\alpha$ is $\begin{pmatrix} 0_{i-1} \\ \alpha \\ 0_{m-i} \end{pmatrix}$, where $0_u \in \mathbb{R}^u$ is a vector of zeros of size u. Similarly, the j-th column of the matrix $\mathbf{F}_\beta$ is $\begin{pmatrix} 0_{j-1} \\ \beta \\ 0_{n-j} \end{pmatrix}$, $j = 1, \ldots, n$. The matrices $\mathbf{F}_\alpha$ and $\mathbf{F}_\beta$ look like this:

$$\mathbf{F}_\alpha = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ \alpha_2 & \alpha_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \alpha_{n-1} & \ddots & \vdots \\ 0 & \alpha_n & \ldots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \alpha_n \end{pmatrix}, \quad \mathbf{F}_\beta = \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ \beta_2 & \beta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_m & \beta_{m-1} & \ddots & \vdots \\ 0 & \beta_m & \ldots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \beta_m \end{pmatrix}. \tag{8}$$

Two (equivalent) matrix forms of (1) are

$$\text{(i)}\ \ \alpha \star \beta = \mathbf{F}_\alpha \beta \quad \text{and} \quad \text{(ii)}\ \ \alpha \star \beta = \mathbf{F}_\beta \alpha\,. \tag{9}$$

Remark 1 If $\alpha \neq 0$ then $\text{rank}(\mathbf{F}_\alpha) = m$, hence in this case the matrix $(\mathbf{F}_\alpha^{\mathrm{T}}\mathbf{F}_\alpha)^{-1}$ exists and $(\mathbf{F}_\alpha^{\mathrm{T}}\mathbf{F}_\alpha)^{-1}\mathbf{F}_\alpha^{\mathrm{T}} Y \neq 0$ for any $Y \in \mathbb{R}^N \setminus \{0\}$. Similarly, if $\beta \neq 0$ then $\text{rank}(\mathbf{F}_\beta) = n$ and $(\mathbf{F}_\beta^{\mathrm{T}}\mathbf{F}_\beta)^{-1}\mathbf{F}_\beta^{\mathrm{T}} Y$ is uniquely defined and is different from 0 for any $Y \in \mathbb{R}^N \setminus \{0\}$.

2.2 Solving the Estimation Problems

Using (i) in (9) we can express Estimation Problem 1 as a problem of parameter estimation in the simple regression model

$$Y = \mathbf{F}_\alpha \beta + \epsilon\,, \tag{10}$$

where ϵ is a vector of errors. Using standard results from linear regression, the least squares estimate (LSE) of β is

$$\hat{\beta} = (\mathbf{F}_\alpha^{\mathrm{T}}\mathbf{F}_\alpha)^{-1}\mathbf{F}_\alpha^{\mathrm{T}}Y\,, \tag{11}$$

which is exactly the required vector (2).

To solve Estimation Problem 1a we write it as a problem of parameter estimation in the regression model $Y = \mathbf{F}_\beta\alpha + \epsilon$, by using (ii) in (9). LSE of α gives the required solution:

$$\hat{\alpha} = (\mathbf{F}_\beta^{\mathrm{T}}\mathbf{F}_\beta)^{-1}\mathbf{F}_\beta^{\mathrm{T}}Y\,. \tag{12}$$

In view of Remark 1 the matrices $(\mathbf{F}_\alpha^{\mathrm{T}}\mathbf{F}_\alpha)^{-1}$ in (11) and $(\mathbf{F}_\beta^{\mathrm{T}}\mathbf{F}_\beta)^{-1}$ in (12) exist.

The optimization problem (4) and (5) are quadratic programming problems with linear constraints. There are no explicit formulas for the solutions of (4) and (5) but there are many fast iterative methods for approximating these solutions, see e.g. [1, 3].

3 Solving the Design Problems

First, we explain why have we called the problems (6) and (7) 'design problems'. Consider the problem (6); a similar interpretation holds for the problem (7). In view of (i) in (9), the problem (6) can be written as

$$(\hat{\alpha}, \hat{\beta}) = \operatorname*{argmin}_{(\alpha,\beta)\in\mathbb{R}^n\times\mathbb{R}^m} \|Y - \mathbf{F}_\alpha\beta\|^2\,, \tag{13}$$

where $\mathbf{F}_\alpha$ is a design matrix in the linear regression model (10). This matrix has size $N \times m$ and the special form shown in (8). Therefore, optimization with respect to β is equivalent to the parameter estimation in the model (10) but optimization with respect to α is equivalent to the optimal choice of the design matrix $\mathbf{F}_\alpha$. Formally, the latter looks like a typical problem of optimal design although common asymptotic considerations including the equivalence theorem do not have much sense here as the original problem is essentially discrete so that there is no sense in letting $N \to \infty$. Moreover, as we have established numerically, the problems (6) and (7) are not convex and require the use of techniques of global optimization for their solution.

3.1 Solving Design Problem 1

For fixed $\alpha \neq 0$ (and hence the design matrix $\mathbf{F}_\alpha$), the value of optimal β is given by (11), the solution to Regression Problem 1. Similarly, for fixed β, the value of

optimal α is given by (12), the solution to Regression Problem 2. By alternating these two steps we obtain the following iterative algorithm

$$\begin{cases} \hat{\beta}_s = \left(\mathbf{F}_{\alpha_{s-1}}^{\mathrm{T}} \mathbf{F}_{\alpha_{s-1}}\right)^{-1} \mathbf{F}_{\alpha_{s-1}}^{\mathrm{T}} Y \\ \hat{\alpha}_s = \left(\mathbf{F}_{\hat{\beta}_s}^{\mathrm{T}} \mathbf{F}_{\hat{\beta}_s}\right)^{-1} \mathbf{F}_{\hat{\beta}_s}^{\mathrm{T}} Y \,, \end{cases} \tag{14}$$

where $s = 1, 2, \ldots$ and $\alpha_0 \in \mathbb{R}^n \setminus \{0\}$ is a starting vector. In view of Remark 1, for all s the vectors $\hat{\beta}_s$ and $\hat{\alpha}_s$ in (14) are well-defined and different from zero.

The algorithm (14) is monotonic since at each iteration the objective function in (6), $SS(\alpha, \beta) = \|Y - \alpha \star \beta\|^2$, is decreasing. However, despite $SS(\alpha, \beta)$ being quadratic in both sets of variables (α and β), it is not convex and the algorithm (14) does not necessarily converge to the global minimizer in (6). A repeated application of (14) from different starting points can be a good option for reaching a good or even an optimal solution of (6).

3.2 Solving Design Problem 2

For fixed α (or β), the value of optimal β (respectively, α) can be approximately found as discussed in Sect. 2.2. This leads to an iterative algorithm which extends the algorithm (14):

$$\begin{cases} \hat{\beta}_s = \operatorname{argmin}_{\beta \in \mathbb{R}_\delta^m} SS(\alpha_{s-1}, \beta) \\ \hat{\alpha}_s = \operatorname{argmin}_{\alpha \in \mathbb{R}_\delta^n} SS(\alpha, \beta_s) \,, \end{cases} \tag{15}$$

where $s = 1, 2, \ldots$, $SS(\alpha, \beta) = \|Y - \alpha \star \beta\|^2$ and $\alpha_0 \in \mathbb{R}^n$ is a starting vector. Similarly to the algorithm (14), the algorithm (15) is monotonic but the limiting point is not necessarily the global minimizer in (7). A repeated application of (15) from different starting points is therefore recommended.

We will now briefly discuss the choice of the initial starting vector α_0.

Consider the algorithm (15) which generalizes the algorithm (14) from the case $\delta = -\infty$ to general δ. This algorithm is a local descent algorithm applied to a problem of global optimization. Any general global optimization technique can be used for choosing initial vectors in (15). Stochastic global optimization techniques could be of special interest, see [11, 12] for comprehensive reviews of these techniques. Alternatively, for certain types of vectors Y, theoretical considerations can be used for choosing α_0. As argued in Sect. 4, the case $Y = (1, 1, \ldots, 1)^{\mathrm{T}}$ is of special importance in problems of matrix structured low-rank approximation. There are some particular combinations of N and n where the exact analytic solution providing $SS(\hat{\alpha}, \hat{\beta}) = 0$ is available but for a generic pair (N, n) there is no such solution.

Table 1 Values of $\mathrm{SS}(\hat{\alpha},\hat{\beta})$ and $\mathrm{DEV}(\hat{\gamma})$ computed by running algorithm (15) started at α_0^*

(n,m)	$(4,4)_\star$	$(4,5)_\star$	$(4,6)_\star$	$(4,7)_\star$	$(4,8)_\star$	$(4,9)_\star$	(4,10)	$(4,15)_\star$	$(4,20)_\star$	(4,100)
$\mathrm{SS}(\hat{\alpha},\hat{\beta})$	0.3333	0.6667	0.1295	0.1691	0.1149	0	0.3765	0	0.0542	0.0988
$\mathrm{DEV}(\hat{\gamma})$	0.3333	0.3333	0.1710	0.2184	0.1429	0	0.2758	0	0.0739	0.1479
(n,m)	$(5,5)_\star$	(5,10)	$(5,20)_\star$	(5,50)	(5,100)	$(10,10)_\star$	(10,50)	(10,100)	(20,100)	(50,100)
$\mathrm{SS}(\hat{\alpha},\hat{\beta})$	0.3333	0.1363	0	0.1342	0.1149	0.3333	0.1395	0.3162	0.3743	1.271
$\mathrm{DEV}(\hat{\gamma})$	0.3333	0.1351	0	0.1376	0.1429	0.3333	0.1419	0.2067	0.2154	0.3437

Assume $Y = (1,1,\ldots,1)^{\mathrm{T}} \in \mathbb{R}^N$ and assume $\delta = 0$ (or some small positive number). Assume also $n \le m$ (otherwise we change the notation $\alpha \leftrightarrow \beta$). If for some reasons only one run of the algorithm (15) is planned then we suggest to choose $\alpha_0^* = (1,\kappa,\kappa,\ldots,\kappa,\kappa,1)^{\mathrm{T}}$, where all entries α_i of α_0^* are κ except for $\alpha_1 = \alpha_n = 1$; κ is some number in $(0,1)$. If $m \ge 2n$ then the vector β_0^* associated with α_0^* has values

$$\beta_i = \begin{cases} (1-\kappa)^{i-1}, & i = 1,2,\ldots,n \\ 1/(2+(n-2)\kappa), & i = n+1,\ldots,m-n \\ (1-\kappa)^{m-i}, & i = m-n+1, m-n+2,\ldots,m\,. \end{cases}$$

All the elements γ_i of $\gamma_0^* = \alpha_0^* \star \beta_0^*$ are equal to 1 except for the elements γ_i with $i = n, n+1,\ldots,2n-1$ and $i = N-2n+1,\ldots,N-n$. The largest elements of γ_0^* are $\gamma_n = \gamma_{N-n} = 2-\kappa$. If $m < 2n$ then we set $\beta_i = (1-\kappa)^{i-1}$, $i = 1,2,\ldots,\lceil m/2\rceil$ and $\beta_{m-i+1} = \beta_i$ for $i > \lceil m/2\rceil$.

We chose $\kappa = 0.2$ and $\delta = 0$. Table 1 shows the values of $\mathrm{SS}(\hat{\alpha},\hat{\beta})$, where $(\hat{\alpha},\hat{\beta})$ is the limiting value of the iterative process (15) started at α_0^*. Table 1 also shows the value of the maximum deviation $\mathrm{DEV}(\hat{\gamma}) = \max_{1\le i\le N}|\hat{\gamma}_i - 1|$, where $\hat{\gamma} = \hat{\alpha} \star \hat{\beta}$. Note that for a large proportion of the values of n and m (with $n \le m$ and $N = n+m-1$) presented in Table 1, we have not managed to improve the values $\mathrm{SS}(\hat{\alpha},\hat{\beta})$ by starting at different (random) starting vectors α_0. The pairs (m,n) where we believe the values of SS are globally optimal have an added asterisk. Table 1 also demonstrates that although in many cases we cannot get the exact deconvolution of the vector of ones, there are approximate deconvolutions which are quite accurate.

4 Application: Low-Rank Approximation of Hankel Matrices

Assume we are given a vector $X = (x_1,\ldots,x_N)^{\mathrm{T}} \in \mathbb{R}^N$ which we interpret as a time series. We want to decompose this series as $x_n = s_j + \epsilon_j$, $j = 1,\ldots,N$, where $S = (s_1,\ldots,s_N)^{\mathrm{T}}$ is a signal and $\epsilon_1,\ldots,\epsilon_N$ denote noise. We assume that the signal S satisfies a linear recurrence relation (LRR) of order $\le r$; that is, $s_j = a_1 s_{j-1} + \ldots +$

$a_r s_{j-r}$, for all $j = r+1, r+2, \ldots, N$, where $a_1, \ldots, a_r$ are some coefficients of the linear recurrence. Formally, the problem we want to solve is finding the minimizer

$$S^* = \arg\min_S \rho(S, X)\,, \tag{16}$$

where $\rho(\cdot,\cdot)$ is a distance on $\mathbb{R}^N \times \mathbb{R}^N$ and the minimum in (16) is taken over the set of vectors in $\mathbb{R}^N$ which satisfy an LRR of order $\le r$. In most applications, the most natural choice of the metric ρ in (16) would be Euclidean.

The optimization problem (16) is usually formulated as a matrix optimization problem, where vectors in (16) are represented by $n \times m$ Hankel matrices. With a vector $Z = (z_1, z_2, \ldots, z_N)^{\mathrm{T}}$ of size N and given $n < N$, we associate an $n \times m$ Hankel matrix

$$\mathbf{H}_Z = \begin{pmatrix} z_1 & z_2 & \cdots & z_m \\ z_2 & z_3 & \cdots & z_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ z_n & z_{n+1} & \cdots & z_N \end{pmatrix},$$

where $m = N + 1 - n$. A matrix version of the optimization problem (16) is

$$H^* = \arg\min_{\mathbf{H}} d(\mathbf{H}, \mathbf{H}_X)\,, \tag{17}$$

where $d(\cdot,\cdot)$ is a distance on $\mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m}$ and the minimum in (17) is taken over the set of all $n \times m$ Hankel matrices of rank $\le r$. The optimization problem (17) is the general problem of Hankel structured low-rank approximation (HSLRA), see e.g. [2, 4, 6].

The optimization problems (16) and (17) are equivalent if the distance functions $\rho(\cdot,\cdot)$ in (16) and $d(\cdot,\cdot)$ in (17) are such that $\rho(U, V) = c \cdot d(\mathbf{H}_U, \mathbf{H}_V)$ for all $U, V \in \mathbb{R}^N$, where $c > 0$ is arbitrary and can be assumed to be 1 without loss of generality.

The standard choice of the distance $d(\cdot,\cdot)$ in the HSLRA problem (17) is $d(\mathbf{X}, \mathbf{X}') = \|\mathbf{X} - \mathbf{X}'\|_F$, where $\|\cdot\|_F$ is the Frobenius norm. The primary reason for this choice is the availability of the singular value decomposition (SVD) which forms the essential part of many algorithms attempting to solve the HSLRA problem, see e.g. [4, 5, 10, 13]. For a software which provides an efficient implementation of the standard methods of the singular spectrum analysis, see [8, 10]; for the particular case of HSLRA we refer to [7]. However, if the distance $d(\mathbf{X}, \mathbf{X}')$ in (17) is $d(\mathbf{X}, \mathbf{X}') = \|\mathbf{X} - \mathbf{X}'\|_F$, then ρ in (16) takes a very particular form, which is far from Euclidean, see e.g. [8, 10].

One would prefer to define the distance function $\rho(\cdot,\cdot)$ in (16) and acquire the distance $d(\cdot,\cdot)$ for (17). As mentioned above, the most natural distance $\rho(\cdot,\cdot)$ in (16) is Euclidean and hence we want to be close to it but at the same time we would like to have the SVD operation available for the matrix low-rank approximation

algorithms. We now show that an optimal answer to this question is given by the solution of (7) with $Y = (1, 1, \ldots, 1)^{\mathrm{T}}$.

Define the so-called $(\mathbf{Q}, \mathbf{R})$-norm in $\mathbb{R}^{n\times m}$ by

$$||\mathbf{U}||^2_{\mathbf{Q},\mathbf{R}} = \mathrm{Trace}\,\mathbf{QURU}^{\mathrm{T}}, \tag{18}$$

where $\mathbf{U} \in \mathbb{R}^{n\times m}$, $\mathbf{Q} \in \mathbb{R}^{n\times n}$, $\mathbf{R} \in \mathbb{R}^{m\times m}$, $\mathbf{Q} = \mathrm{diag}(Q)$, $\mathbf{R} = \mathrm{diag}(R)$ with diagonals $Q = (q_1, \ldots, q_n)^{\mathrm{T}}$, and $R = (r_1, \ldots, r_m)^{\mathrm{T}}$ respectively; all components of Q and R are positive. The $(\mathbf{Q}, \mathbf{R})$-norms admit SVD expansions for low-rank matrix approximations, see e.g. [9], so we shall seek for the distance in (17) which is defined through a $(\mathbf{Q}, \mathbf{R})$-norm with some Q and R. The relation between the $(\mathbf{Q}, \mathbf{R})$-norms in (17) and the Euclidean norm in (16) is established by the following lemma, which is not difficult to prove.

Lemma 1 *For any $X \in \mathbb{R}^N$, we have*

$$||\mathbf{H}_X||^2_{\mathbf{Q},\mathbf{R}} = X^{\mathrm{T}}\mathbf{W}X, \tag{19}$$

where $\mathbf{W} = \mathrm{diag}(W)$ and the vector $W = (w_1, \ldots, w_N)^{\mathrm{T}}$ is the convolution of the vectors Q and R: $W = Q \star R$.

Lemma 1 implies that if vectors Q and R are chosen so that $Q \star R \simeq W = (1, 1, \ldots, 1)^{\mathrm{T}}$, then we get the required norm in (17). This is exactly the version of the problem (7) which we have discussed in Sect. 3.2.

Acknowledgements (i) The authors are grateful to all three referees for their interesting and valuable comments. (ii) Research of the first author was supported by the Russian Science Foundation, project No. 15-11-30022 "Global optimization, supercomputing computations, and applications".

References

1. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge/New York (2004)
2. Chu, M.T., Funderlic, R.E., Plemmons, R.J.: Structured low rank approximation. Linear Algebra Appl. **366**, 157–172 (2003)
3. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Nav. Res. Logist. Q. **3**, 95–110 (1956)
4. Gillard, J., Zhigljavsky, A.: Optimization challenges in the structured low rank approximation problem. J. Glob. Optim. **57**, 733–751 (2013)
5. Gillard, J., Zhigljavsky, A.: Application of structured low-rank approximation methods for imputing missing values in time series. Stat. Interface **8**, 321–330 (2015)
6. Gillard, J., Zhigljavsky, A.: Stochastic algorithms for solving structured low-rank matrix approximation problems. Commun. Nonlinear Sci. Numer. Simul. **21**, 70–88 (2015)
7. Golyandina, N., Korobeynikov, A.: Basic singular spectrum analysis and forecasting. R. Comput. Stat. Data Anal. **71**, 934–954 (2014)

8. Golyandina, N., Nekrutkin, V., Zhigljavsky, A.: Analysis of Time Series Structure: SSA and Related Techniques. CRC Press, Boca Raton (2001)
9. Golyandina, N., Shlemov, A.: Variations of singular spectrum analysis for separability improvement: non-orthogonal decompositions of time series. Stat. Interface **8**, 277–294 (2015)
10. Golyandina, N., Zhigljavsky, A.: Singular Spectrum Analysis for Time Series. Springer, New York (2013)
11. Zhigljavsky, A.A.: Theory of Global Random Search. Kluwer Academic Press, Dordrecht (1991)
12. Zhigljavsky, A.A., Zilinskas, A.G.: Stochastic Global Optimization. Springer, New York (2008)
13. Zvonarev, N., Golyandina, N.: Iterative algorithms for weighted and unweighted finite-rank time-series approximations. Stat. Interface **9**(4), (2016, to appear). arxiv http://arxiv.org/abs/1507.02751

Index

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8