Continuous
Univariate
Distributions

Volume 2

Continuous Univariate Distributions

Volume 2

Second Edition

NORMAN L. JOHNSON
University of North Carolina
Chapel Hill, North Carolina

SAMUEL KOTZ
University of Maryland
College Park, Maryland

N. BALAKRISHNAN
McMaster University
Hamilton, Ontario, Canada

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore

This text is printed on acid-free paper.

Library of Congress Cataloging in Publication Data:

Johnson, Norman Lloyd.
Continuous univariate distributions.
(Wiley series in probability and mathematical statistics. Applied probability and statistics)
"A Wiley-Interscience publication."
Includes bibliographical references and indexes.
1. Distribution (Probability theory) **I. Kotz,** Samuel. **II.** Balakrishnan, N., 1956– . **III.** Title. **IV.** Series.
QA273.6J6 1994 519.2'4 93-45436
ISBN 0-471-58494-0 (**v.** 2: acid-free)

Printed in the United States of America

10 9 8

Contents

Preface

The remarks in the Preface to the new edition of ***Continuous Uniuariate Distributions-1*** apply to the present volume also. This second edition differs from the first in the following respects:

- The chapter on **Extreme Value Distributions,** which was the final chapter of the original ***Continuous Uniuariate Distributions-1*** now appears as the first chapter of the present volume.
- The chapter on **Quadratic Forms** has been postponed to a projected volume dealing with ***Continuous Multivariate Distributions.***
- The final chapter on **Miscellaneous Distributions** has been drastically revamped and restricted, and some topics have been given more, and some markedly less emphasis.
- The length of each chapter has been substantially increased (about doubled, on average) and the number of references increased almost threefold.
- In order to mirror recent developments, the authors have, somewhat reluctantly, included descriptions of numerous results relating to approximations. Although these are often computationally ingenious, their practical relevance in an age of high speed computers has been substantially diminished.
- On the other hand, we were happy to include many examples of applications of distributions (such as logistic, Laplace, beta, F, t and noncentral chi-square, F and t) in various new fields of science, business and technology. We welcome this trend towards penetration of more sophisticated models into wider areas of human endeavour.

Since the publication of the new edition of ***Continuous Uniuariate Distributions-I,*** the sixth edition of ***Kendall's Advanced Theory of Statistics, Volume I-Distribution Theory*** by A. Stuart and J. K. Ord has come out, providing a lot of details on univariate as well as multivariate distribution theory. Though it was late for Volume 1, we have tried to coordinate in this Volume (at some

places) with results presented in Stuart and Ord. Our sincere thanks go to Professor Keith Ord for providing us with a copy of page proofs in order to achieve this goal. We record with gratitude a large number of comments received from our colleagues in statistical and engineering communities concerning misprints in and omissions from the first edition of this volume. These were very valuable to us in preparation of this new edition.

We acknowledge with thanks the invaluable assistance of Mrs. Lisa Brooks (University of North Carolina) and Mrs. Debbie Iscoe (Hamilton, Canada) in their skillful typing of the manuscript. We also thank the Librarians of the University of North Carolina, University of Maryland, and McMaster University for their help in library research.

Special thanks are also due to Mrs. Kate Roach and Mrs. Shirley Thomas at John Wiley & Sons in New York for their sincere efforts in the fine production of this volume. We also thank Ms. Dana Andrus for all the efforts in copy-editing the long manuscript.

Thanks are offered to the Institute of Mathematical Statistics, the American Statistical Association, the Biometrika Trustees, the Institute of Electrical and Electronics Engineering, Marcel Dekker, Inc., the Royal Statistical Society, the Australian Statistical Society, the Statistical Society of Canada, the Biometric Society, North Holland, Gordon and Breach Science Publishers, and the editors of ***Naval Research Logistics Quarterly, Water Resources Research, Soochow Journal of Mathematics, Journal of the Operational Research Society, Sankhyā, Decision Sciences, Mathematical and Computer Modelling, International Statistical Review,*** and ***Oxford Bulletin of Economics and Statistics,*** for their kind permission to reproduce previously published tables and figures.

NORMAN L. JOHNSON
SAMUEL KOTZ
N. BALAKRISHNAN

Chapel Hill, North Carolina
College Park, Maryland
Hamilton, Ontario, Canada
February, 1995

List of Tables

CHAPTER 22

Extreme Value Distributions

1 GENESIS

The development of extreme value distributions proceeded to some extent outside the mainstream of statistical distribution theory, with its early stage dominated by work on curve fitting and the later stage by problems encountered in statistical inference. The extreme value theory is a blend of an enormous variety of applications involving natural phenomena such as rainfall, floods, wind gusts, air pollution, and corrosion, and delicate mathematical results on point processes and regularly varying functions. This area of research thus attracted initially the interests of theoretical probabilists as well as engineers and hydrologists, and only recently of the mainstream statisticians. Historically work on extreme value problems may be dated back to as early as **1709** when Nicolas Bernoulli discussed the mean largest distance from the origin when n points lie at random on a straight line of length t [see Gumbel **(1958)]**.

Extreme value theory seems to have originated mainly from the needs of astronomers in utilizing or rejecting outlying observations. The early papers by Fuller **(1914)** and Griffith **(1920)** on the subject were highly specialized both in fields of applications and in methods of mathematical analysis. A systematic development of the general theory may, however, be regarded as having started with a paper by Bortkiewicz **(1922)** that dealt with the distribution of range in random samples from a normal distribution. This has already been pointed out in Chapter **13** and as can be seen in that chapter, subsequent development of this particular topic was quite rapid. From our present point of view, it suffices to say that the importance of the paper by Bortkiewicz **(1922)** resides in the fact that the concept of ***distribution of largest value*** was clearly introduced in it. In the very next year von Mises **(1923)** evaluated the expected value of this distribution, and Dodd **(1923)** calculated its median and also discussed some nonnormal parent distributions. Of more direct relevance to this chapter is a paper by **Fréchet (1927)** in which asymptotic distributions of largest values are considered. In the following year Fisher and Tippett **(1928)** published results of an independent inquiry into the same problem. While Fréchet (1927) had identified one possible limit distribution for the largest order statistic, Fisher and Tippett (1928) showed that extreme limit distributions can only be one of three types. Tippett (1925) had earlier studied the exact cumulative distribution function and moments of the largest order statistic and of the sample range arising from samples from a normal population. von Mises (1936) presented some simple and useful sufficient conditions for the weak convergence of the largest order statistic to each of the three types of limit distributions given earlier by Fisher and Tippett (1928). Seven years later, it was Gnedenko (1943) who presented a rigorous foundation for the extreme value theory and provided necessary and sufficient conditions for the weak convergence of the extreme order statistics. de Haan (1970) refined the work of Gnedenko. Gnedenko's (1943) classical paper has been reproduced in the first volume of ***Breakthroughs in Statistics*** and supplemented by a perceptive introduction written by R. L. Smith in which the influence of the paper and subsequent developments in the extreme value theory have been analyzed.

The theoretical developments of the 1920s and mid 1930s were followed in the late 1930s and 1940s by a number of papers dealing with practical applications of extreme value statistics in distributions of human lifetimes, radioactive emissions [Gumbel (1937a, b)], strength of materials [Weibull (1939)], flood analysis [Gumbel (1941, 1944, 1945, 1949a), Rantz and Riggs (1949)], seismic analysis [Nordquist (1945)], and rainfall analysis [Potter (1949)]. From the application point of view, Gumbel made several significant contributions to the extreme value analysis; most of them are detailed in his book-length account of statistics of extremes [Gumbel (1958)] which is an extension of his earlier brochure [Gumbel (1954)]. Many more applications are listed in Section 14.

The bibliography at the end of this chapter contains about 350 references. This impressive number is, however, only a small part of publications pertaining to this subject. The bibliography in Gumbel's (1958) book, not including the developments during the last 35 years, is even more extensive. While this extensive literature serves as a testimony to the vitality and applicability of the extreme value distributions and processes, it also reflects to some extent on the lack of coordination between researches and the inevitable duplication (or even triplication) of results appearing in a wide range of publications.

2 INTRODUCTION

Extreme value distributions are generally considered to comprise the three following families:

Type 1:

$$\Pr[X \le x] = \exp\{-e^{-(x-\xi)/\theta}\}. \tag{22.1}$$

Type 2:

$$\Pr[X \le x] = \begin{cases} 0, & x < \xi, \\ \exp\left\{-\left(\dfrac{x-\xi}{\theta}\right)^{-k}\right\}, & x \ge \xi. \end{cases} \tag{22.2}$$

Type 3:

$$\Pr[X \le x] = \begin{cases} \exp\left\{-\left(\dfrac{\xi-x}{\theta}\right)^{k}\right\}, & x \le \xi, \\ 1, & x > \xi, \end{cases} \tag{22.3}$$

where ξ, $\theta(>0)$, and $k(>0)$ are parameters. The corresponding distributions of $(-X)$ are also called ***extreme value distributions.***

Of these families of distributions type 1 is by far the one most commonly referred to in discussions of extreme value distributions. Indeed some authors call (22.1) ***the*** extreme value distribution. In view of this, and the fact that distributions (22.2) and (22.3) can be transformed to type 1 distributions by the simple transformations

$$Z = \log(X - \xi), \quad Z = -\log(\xi - X),$$

respectively, we will, for the greater part of this chapter, confine ourselves to discussion of type 1 distributions. We may also note that the type 3 distribution of $(-X)$ is a ***Weibull*** distribution. These distributions have been discussed in Chapter 21, and so there is no need to discuss them in detail here.

Of course, types 1 and 2 are also closely related to the Weibull distribution, by the simple formulas relating Z and X, just quoted. Type 1 is sometimes called the ***log-Weibull*** distribution [see White (1964, 1969)].

Type 1 distributions are also sometimes called ***doubly exponential*** distributions, on account of the functional form of (22.1). We do not use this term to avoid confusion with Laplace distributions (Chapter 24), which are also called double exponential.

The term "extreme value" is attached to such distributions because they can be obtained as limiting distributions (as $n \to \infty$) of the greatest value among n independent random variables each having the same continuous distribution (see Section 3). By replacing X by $-$X, limiting distributions of ***least*** values are obtained. As already mentioned, these are also extreme value distributions, so they do not need separate treatment.

Although the distributions are known as extreme value, it is to be borne in mind (1) that they do not represent distributions of ***all*** kinds of extreme values (e.g., in samples of finite size), and (2) they can be used empirically (i.e., without an extreme value model) in the same way as other distributions.

In this last connection we note that the type 1 distribution may be regarded as an approximation to a Weibull distribution with large value of c. Also if X has a type 1 distribution, $Z = \exp[-(X - \xi)/\theta]$ has an exponential distribution with probability density function:

$$p_Z(z) = e^{-z}, \qquad 0 \le z.$$

The type 2 distribution is also referred to as the ***Fréchet-type distribution;*** the type 3 distribution is called as the ***Weibull-type distribution;*** and the type 1 distribution as the ***Gumbel-type distribution.*** We may note that the Fréchet and Weibull distributions are also related by a simple change of sign. The ***Gompertz distribution*** of lifetimes introduced in 1825, and already in use for about a century before Fisher and Tippett's results appeared is a type 1 distribution even though it is not generally regarded to be of this group (see Section 8 for details).

Although the three types of the distributions in (22.1)–(22.3) appear to be unrelated, they may all be represented as members of a single family of generalized distributions with cumulative distribution function

$$\Pr[X \le x] = e^{-\{1+[(x-\xi)/\theta]/\alpha\}^{-\alpha}}, \qquad 1 + \frac{1}{\alpha}\left(\frac{x-\xi}{\theta}\right) > 0,$$

$$-\infty < \alpha < \infty, \qquad \theta > 0. \tag{22.4}$$

For $\alpha > 0$ the distribution (22.4) is of the same form as (22.2). For $\alpha < 0$ the distribution (22.4) becomes of the same form as (22.3). Finally, when $\alpha \to \infty$ or $-\infty$, the distribution (22.4) becomes the same form as the type 1 extreme value distribution in (22.1). For this reason the distribution function in (22.4) is known as the ***generalized extreme value distribution*** and is also sometimes referred to as the ***von Mises type extreme value distribution*** or the ***von Mises–Jenkinson-type distribution.*** More details on this distribution will be presented in Section 15.

Mann and Singpurwalla (1982) have provided a brief review of the extreme value distributions. A similar review of the Gumbel distribution has been made by Tiago de Oliveira (1983).

3 LIMITING DISTRIBUTIONS OF EXTREMES

Extreme value distributions were obtained as limiting distributions of greatest (or least) values in random samples of increasing size. To obtain a nondegenerate limiting distribution, it is necessary to "reduce" the actual greatest value by applying a linear transformation with coefficients which depend on the sample size. This process is analogous to standardization (as in central limit theorems; see Chapter 13, Section 2) though not restricted to this particular sequence of linear transformations.

If X,, $X_2, \cdots, X_n$ are independent random variables with common probability density function

$$p_{X_j}(x) = f(x), \qquad j = 1, 2, \ldots, n,$$

then the cumulative distribution function of $X_n' = \max(X_1, X_2, \cdots, X_n)$ is

$$F_{X_n'}(x) = [F(x)]^n, \tag{22.5}$$

where

$$F(x) = \int_{-\infty}^{x} f(x)\, dt.$$

As n tends to infinity, it is clear that for any fixed value of x

$$\lim_{n\to\infty} F_{X_n'}(x) = \begin{cases} 1 & \text{if } F(x) = 1, \\ 0 & \text{if } F(x) < 1. \end{cases}$$

Even if it is proper, this limiting distribution would be "trivial" and of no special interest. If there is a limiting distribution of interest, we must find it as the limiting distribution of some sequence of transformed "reduced" values, such as (a, $X_n' + b_n$), where a,, b, may depend on n but not on x.

To distinguish the limiting cumulative distribution of the "reduced" greatest value from $F(x)$, we will denote it by $G(x)$. Then since the greatest of Nn values $X_1, X_2, \ldots, X_{Nn}$ is also the greatest of the N values

$$\max(X_{(j-1)n+1}, X_{(j-1)n+2}, \cdots, X_{jn}), \qquad j = 1, 2, \ldots, N,$$

it follows that $G(x)$ must satisfy the equation

$$[G(x)]^N = G(a_N x + b_N). \tag{22.6}$$

This equation was obtained by Fréchet (1927) and also by Fisher and Tippett (1928). It is sometimes called the stability postulate.

Type 1 distributions are obtained by taking a, = 1; types 2 and 3 by taking a, # 1. In this latter case

$$x = a_N x + b_N \qquad \text{if } x = b_N(1 - a_N)^{-1},$$

and from (22.6) it follows that $G(b_N(1 - a_N)^{-1})$ must equal 1 or 0. Type 2 corresponds to 1, and type 3 to 0.

We now consider the case a, = 1 (type 1) in some detail. Equation (22.6) is now

$$[G(x)]^N = G(x + b_N). \tag{22.7}$$

Since $G(x + b_N)$ must also satisfy (22.6),

$$[G(x)]^{NM} = [G(x + b_N)]^M = G(x + b_N + b_M). \tag{22.8}$$

But, also from (22.6),

$$[G(x)]^{NM} = G(x + b_{NM}) \tag{22.9}$$

and from (22.8) and (22.9) we have

$$b_N + b_M = b_{NM},$$

whence

$$b_N = 6 \log N, \qquad \theta \text{ a constant.} \tag{22.10}$$

Taking logarithms of (22.7) twice and inserting the value of b_N from (22.10), we have (noting that $G \leq 1$)

$$\log N + \log\{-\log G(x)\} = \log\{-\log G(x + \theta \log N)\}. \tag{22.11}$$

In other words, when the argument of

$$h(x) = \log\{-\log G(x)\}$$

increases by 6 log N, $h(x)$ decreases by log N. Hence

$$h(x) = h(0) - \frac{x}{\theta}. \tag{22.12}$$

Since $h(x)$ decreases as x increases, $6 > 0$. From (22.12),

$$\begin{aligned} -\log G(x) &= \exp\left[-\frac{x - \theta h(0)}{\theta}\right] \\ &= \exp\left(-\frac{x - \xi}{\theta}\right) \end{aligned}$$

where $\xi = \theta \log(-\log G(0))$. Hence

$$G(x) = \exp[-e^{-(x-\xi)/\theta}],$$

which is in agreement with (22.1). We will not enter into details of derivation for types 2 and 3, and interested readers may refer to Galambos (1978,1987).

Gnedenko (1943) established certain correspondences between the parent distribution [$F(x)$ in the above analysis] and the type to which the limiting distribution belongs. It should be noted that the conditions relate essentially to the behavior of $F(x)$ for high (low) values of x if the limiting distribution of greatest (least) values is to be considered. It is quite possible for greatest and least values, corresponding to the same parent distribution, to have different limiting distributions.

We now summarize Gnedenko's results:

For type 1 distribution: Defining X_α by the equation

$$F(X_\alpha) = \alpha,$$

the condition is

$$\lim_{n \to \infty} n\left[\left(1 - F\left(X_{1-n^{-1}} + y\left(X_{1-(ne)^{-1}} - X_{1-n^{-1}}\right)\right)\right] = e^{-y}. \quad (22.13)$$

For type 2 distribution:

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(cx)} = c^k, \qquad c > 0, k > 0. \qquad (22.14)$$

For type 3 distribution:

$$\lim_{x \to 0^-} \frac{1 - F(cx + w)}{1 - F(x + \omega)} = c^k, \qquad c > 0, k > 0, \qquad (22.15)$$

where $F(\omega) = 1, F(x) < 1$ for $x < \omega$.

Gnedenko also showed that these conditions are necessary, as well as sufficient, and that there are ***no other*** distributions satisfying the stability postulate. An alternative interpretation of these conditions was given by Clough and Kotz *(1965)* who also illustrated a special queueing model application for the extreme value distributions. Among distributions satisfying the type 1 condition *(22.13)* are normal, exponential, and logistic; the type 2 condition *(22.14)* is satisfied by Cauchy; the type 3 condition is satisfied by nondegenerate distributions with range of variation bounded above.

Gnedenko's (1943) results have been generalized by several authors. Results for order statistics of fixed and increasing rank were obtained by

Smirnov (1952) who completely characterized the limiting types and their domains of attraction. Generalizations for the maximum term have been made by Juncosa (1949) who dropped the assumption of a common distribution, Watson (1954) who proved that under mild restrictions the limiting distribution of the maximum term of a stationary sequence of m-dependent random variables is the same as in the independent case, Berman (1962) who studied exchangeable random variables and samples of random size, and Harris (1970) who extended the classical theory by introducing a model from reliability theory (essentially a series system with replaceable components). Weinstein (1973) generalized the basic result of Gnedenko dealing with the asymptotic distribution of the exponential case with the initial distribution $V(x) = 1 - e^{-x^{\upsilon}}$ $(x \geq 0)$. He showed that

$$\lim_{n\to\infty} V^n\left\{\left(x_n^{\upsilon} + \frac{u}{d_n}\right)^{1/\upsilon}\right\} = e^{-e^{-u}}, \qquad \upsilon > 0,$$

if and only if

$$\lim_{n\to\infty} n\left[1 - V\left\{\left(x_n^{\upsilon} + \frac{u}{d_n}\right)^{1/\upsilon}\right\}\right] = e^{-u},$$

where

$$V(x_n) = 1 - \frac{1}{n},$$

$$V\left(x_n + \frac{1}{c_n}\right) = 1 - \frac{1}{ne},$$

$$d_n = \frac{c_n}{\upsilon |x_n|^{\upsilon - 1}},$$

$$x^{\upsilon} \equiv |x|^{\upsilon}\,\mathrm{sgn}(x).$$

[Gnedenko's (1943) result is for $\upsilon = 1$.] See also Jeruchim (1976) who has warned that the additional parameter o must be treated cautiously in applications.

The necessary and sufficient conditions in (22.13)–(22.15) are often difficult to verify. In such instances the following sufficient conditions established by von Mises (1936) may be useful (though they are applicable only for absolutely continuous parent distributions):

For type 1 distribution. For $r(x) = f(x)/[1 - F(x)]$ nonzero and differentiable for x close to $F^{-1}(1)$ [or for large x if $F^{-1}(1) = \infty$], the condition is

$$\lim_{x \to F^{-1}(1)^-} \frac{d}{dx}\left\{\frac{1}{r(x)}\right\} = 0. \qquad (22.16)$$

For type 2 distribution. For $r(x) > 0$ for large x and for some a > 0, the condition is

$$\lim_{x \to \infty} xr(x) = a. \qquad (22.17)$$

For type 3 distribution. For $F^{-1}(1) < \infty$ and for some a > 0, the condition is

$$\lim_{x \to F^{-1}(1)^-} \{F^{-1}(1) - x\}r(x) = a. \qquad (22.18)$$

de Haan (1976) has provided a simple proof of this result. The function $r(x) = f(x)/\{1 - F(x)\}$ appearing in conditions (22.16)–(22.18) is the failure rate or the hazard function (see Chapter 1, Section B2).

The choice of the normalizing constants a_N and $b_N > 0$ (which are not unique) depends on the type of the three limiting distributions. In general, convenient choices for a_N and b_N are as follows:

For type 1 distribution.

$$a_N = F^{-1}\left(1 - \frac{1}{N}\right),$$

$$b_N = F^{-1}\left(1 - \frac{1}{Ne}\right) - F^{-1}\left(1 - \frac{1}{N}\right). \qquad (22.19)$$

For type 2 distribution.

$$a_N = 0,$$

$$b_N = F^{-1}\left(1 - \frac{1}{N}\right). \qquad (22.20)$$

For type 3 distribution.

$$a_N = F^{-1}(1),$$

$$b_N = F^{-1}(1) - F^{-1}\left(1 - \frac{1}{N}\right). \qquad (22.21)$$

Analogous results for the limiting distributions of the sample minimum can be stated in a straightforward manner. There are several excellent books that deal with the asymptotic theory of extremes and statistical applications of extremes. David (1981) and Arnold, Balakrishnan, and Nagaraja (1992) provide a compact account of the asymptotic theory of extremes, and Galambos (1978, 1987), Resnick (1987), and Leadbetter, Lindgren, and Rootzén (1983) present elaborate treatments of this topic. Reiss (1989) discusses various convergence concepts and rates of convergence associated with the extremes (and also with the order statistics). Castillo (1988) has updated Gumbel (1958) and presented many statistical applications of the extreme value theory. Harter (1978) prepared an authoritative bibliography of the extreme value theory.

With $F_X(x;\xi,\theta)$ denoting the extreme value distribution for the sample minimum with cdf given by

$$F_X(x;\xi,\theta) = 1 - e^{-e^{(x-\xi)/\theta}}, \qquad \theta > 0, \xi \in \mathbb{R},$$

and $G_X(x;a,b,c)$ denoting the three-parameter Weibull distribution with cdf (see Chapter 21)

$$G_X(x;a,b,c) = \begin{cases} 0 & \text{if } x \le c, \\ 1 - e^{-[(x-c)/b]^a} & \text{if } x \ge c, \end{cases}$$

for $a, b > 0$ and $c \in \mathbb{R}$, Davidovich (1992) established some bounds for the difference between the two cdfs. Specifically he showed that

$$F_X\left(x; b + c, \frac{b}{a}\right) - G_X(x;a,b,c) < \begin{cases} e^{-a} & \text{if } x \le c, \\ \dfrac{2e^{-2}}{a-2} & \text{if } c \le x \le c + 2b, \\ e^{a-2^a} & \text{if } x \ge c + 2b. \end{cases}$$

Thus, if $a \to \infty$, $b \to \infty$, and $c \to -\infty$ such that $b + c \to d$ ($|d| < \infty$) and $b/a \to f$ ($0 < f < \infty$), then the Weibull distribution above uniformly approaches the extreme value distribution for the minimum with $\xi = d$ and $\theta = f$.

It is easily proved that if $Y_1, Y_2, \cdots$, are independent variables, each having the exponential distribution (see Chapter 19, Section 1),

$$\Pr[Y \le y] = 1 - e^{-y}, \qquad y > 0, \tag{22.22}$$

and if L is the zero-truncated Poisson variable (see Chapter 4, Section 10)

$$\Pr[L = l] = \frac{(e^{\lambda} - 1)^{-1}\lambda^{l}}{l!}, \qquad l = 1, 2, \dots, \tag{22.23}$$

the random variable defined by

$$X = \max(Y_1, \cdots, Y_L)$$

has the extreme value distribution with cdf

$$\Pr(X \le x) = (e^{\lambda} - 1)^{-1}[\exp\{\lambda(1 - e^{-x})\}] = c\exp[-\lambda e^{-x}]. \tag{22.24}$$

In a similar manner the **Fréchet** distribution can be generated from the Pareto distribution (see Chapter 20) and the Weibull from the power function distribution (see Chapter 20). In fact Sibuya (1967) has suggested a method of generating pseudorandom numbers from the extreme value distribution by using the method described above, based on the exponential distribution.

4 DISTRIBUTION FUNCTION AND MOMENTS

In this section we will consider type 1 distributions (22.1) exclusively. Corresponding to (22.1) is the probability density function

$$p_X(x) = \theta^{-1} e^{-(x-\xi)/\theta} \exp[-e^{-(x-\xi)/\theta}]. \tag{22.25}$$

If $\xi = 0$ and $\theta = 1$, or equivalently, the distribution of $Y = (X - \xi)/\theta$, we have the ***standard form***

$$p_Y(y) = \exp(-y - e^{-y}). \tag{22.26}$$

Since, as we pointed out in Section 1, the variable $Z = \exp[-(X - \xi)/\theta] = e^{-Y}$ has the exponential distribution

$$p_Z(z) = e^{-z}, \qquad z \ge 0,$$

it follows that

$$E[e^{t(X-\xi)/\theta}] = E[Z^{-t}] = \Gamma(1-t)$$

for $t < 1$. Replacing t by θt, we see that the moment generating function of X is

$$E[e^{tX}] = e^{t\xi}\Gamma(1-\theta t), \qquad \theta|t| < 1, \tag{22.27}$$

and the cumulant generating function is

$$\Psi(t) = \xi t + \log\Gamma(1-\theta t). \tag{22.28}$$

The cumulants of X are

$$\kappa_1(X) = E[X] = \xi - \theta\psi(1) = \xi + \gamma\theta \simeq \xi + 0.57722\theta, \tag{22.29}$$

where γ is Euler's constant, and

$$\kappa_r(X) = (-\theta)^r\psi^{(r-1)}(1), \qquad r \geq 2. \tag{22.30}$$

In particular

$$\operatorname{var}(X) = \tfrac{1}{6}\pi^2\theta^2 \simeq 1.64493\theta^2, \tag{22.31}$$

$$\text{Std. dev.}(X) \simeq 1.282559, \tag{22.31$'$}$$

and the moment ratios are

$$\alpha_3^2(X) = \beta_1(X) \simeq 1.29857, \quad \alpha_4(X) = \beta_2(X) = 5.4. \tag{22.32}$$

Note that ξ and *9* are purely location and scale parameters, respectively. All distributions (22.25) have the same shape.

The distribution is unimodal. Its mode is at $X = \xi$, and there are points of inflection at

$$X = \xi \pm \theta\log\left[\tfrac{1}{2}(3+\sqrt{5})\right] \simeq \xi \pm 0.96242\theta. \tag{22.33}$$

For $0 < p < 1$ the pth quantile defined by $F(X_p) = p$ readily becomes from (22.1)

$$X_p = \xi - \theta\log(-\log p). \tag{22.34}$$

From (22.34) we immediately obtain the lower quartile, median, and upper quartile to be

$$X_{0.25} = \xi - \theta \log(\log 4) \simeq \xi - 0.32663\theta, \tag{22.35}$$

$$X_{0.50} = \xi - \theta \log(\log 2) \simeq \xi + 0.36611\theta, \tag{22.36}$$

$$X_{0.75} = \xi - \theta \log(-\log 0.75) \simeq \xi + 1.24590\theta, \tag{22.37}$$

respectively.

Quantiles of the distribution are easy to compute from (22.34), even with a pocket calculator. Most of the standard distribution (22.26) is contained in the interval $(-2, 7)$. As a matter of fact, for the distribution function in (22.1) we find the probability between ξ – 28 and ξ + 78 to be 0.998. That is, 99.8% of the distribution lies between Mean – 2.0094 · (Standard deviation) and Mean + 5.0078 · (Standard deviation). More details on properties of the distribution are presented by Lehman (1963).

The standard probability density function (22.26) is shown in Figure 22.1. Its shape is very closely mimicked by a lognormal distribution with $e^{\sigma^2} =$ 1.1325 (in the notation of Chapter 14). (The β_1, β_2 values of this lognormal distribution are 1.300, 5.398, respectively; (22.32) *cf.*) In Table 22.1 the standard cumulative distribution functions are compared.

Table 22.2 gives standardized percentile points (i.e., for a type 1 extreme value distribution with expected value zero and standard deviation 1, corresponding to $\theta = \sqrt{6}/\pi = 0.77970$; $\xi = -\gamma\theta = -0.45006$). The positive skewness of the distribution is clearly indicated by these values. The usefulness of this distribution to model time-to-failure data in reliability studies has been discussed by Canfield (1975) and Canfield and Borgman (1975).

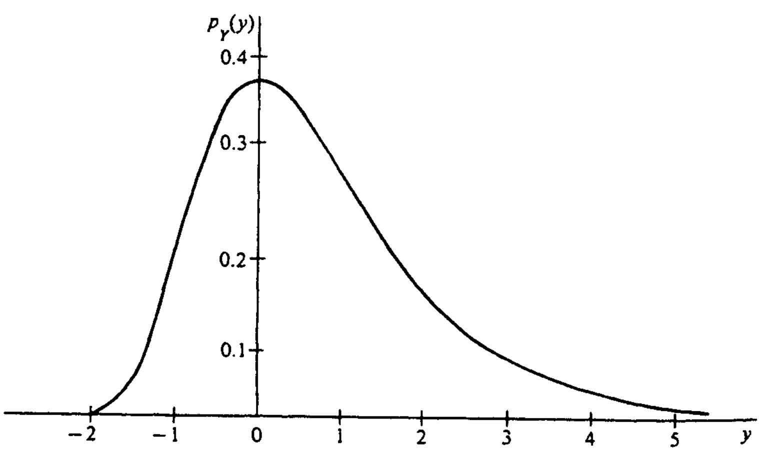


Figure 22.1 Standard type 1 probability density function, $p_Y(y) = e^{-y}\exp(-e^{-y})$

Table 22.1 Standard cumulative distribution functions

x	$F(x)$	
	Type 1 Extreme Value Distribution[a]	Lognormal Distribution[b]
−2.0	0.00068	0.00022
−1.5	0.02140	0.01959
−1.0	0.1321	0.1342
−0.5	0.3443	0.3471
0.0	0.5704	0.5700
0.5	0.7440	0.7423
1.0	0.8558	0.8546
1.5	0.92237	0.92096
2.0	0.95774	0.95792
2.5	0.97752	0.97730
3.0	0.98810	0.98837
3.5	0.99371	0.99389
4.0	0.99668	0.99677

[a]Where $F(x) = \exp[-\exp\{-1.28254x - 0.57722\}]$;
[b]Where $F(x) = (\sqrt{2\pi})^{-1}\int_{-\infty}^{u(x)} \exp(-u^2/2)\,du$ with $u(x) = 6.52771 \log_{10}(x + 2.74721) - 2.68853$.

Table 22.2 Standardized percentiles for Type 1 extreme value distribution

a	Percentile
0.0005	−2.0325
0.0001	−1.9569
0.0025	−1.8460
0.005	−1.7501
0.01	−1.6408
0.025	−1.4678
0.05	−1.3055
0.1	−1.1004
0.25	−0.7047
0.5	−0.1643
0.75	0.5214
0.9	1.3046
0.95	1.8658
0.975	2.4163
0.99	3.1367
0.995	3.6791
0.9975	4.2205
0.999	4.9355
0.9995	5.4761

5 ORDER STATISTICS

If $Y_1' \le Y_2' \le \dots \le Y_n'$ are the order statistics corresponding to n independent random variables each having the standard type 1 extreme value distribution (22.26), then the probability density function of Y_r' $(1 \le r \le n)$ is

$$p_{Y_r'}(y) = \frac{n!}{(r-1)!(n-r)!}\left(e^{-e^{-y}}\right)^{r-1}\left(1 - e^{-e^{-y}}\right)^{n-r} e^{-y} e^{-e^{-y}}$$

$$= \frac{n!}{(r-1)!(n-r)!}\sum_{j=0}^{n-r}(-1)^j \binom{n-r}{j} e^{-y-(j+r)e^{-y}},$$

$$-\infty < y < \infty. \qquad (22.38)$$

From (22.38), the kth moment of Y_r' can be written as [Lieblein (1953)]

$$E\left[Y_r'^k\right] = \frac{n!}{(r-1)!(n-r)!}\sum_{j=0}^{n-r}(-1 \qquad r) g_k(r+j), \qquad (22.39)$$

where

$$g_k(c) = \int_{-\infty}^{\infty} y^k e^{-y-ce^{-y}}\, dy$$

$$= (-1)^k \int_{-\infty}^{\infty} (\log u)^k e^{-cu}\, du \qquad (\text{with } u = e^{-y}).$$

For nonnegative integers k, $g_k(c)$ may be written

$$g_k(c) = (-1)^k \frac{d^k}{dt^k}\int_0^{\infty} u^{t-1} e^{-cu}\, du\Big|_{t=1}$$

$$= (-1)^k \frac{d^k}{dt^k}\{\Gamma(t)c^{-t}\}\Big|_{t=1}. \qquad (22.40)$$

The functions $g_1(\cdot)$ and $g_2(\cdot)$ needed for the expressions of the first two moments of order statistics are obtained from *(22.40)* to be

$$g_1(c) = -\frac{\Gamma'(1)}{c} + \frac{\Gamma(1)}{c}\log c = \frac{1}{c}(\gamma + \log c) \qquad (22.41)$$

and

$$g_2(c) = \frac{1}{c}\left\{\frac{\pi^2}{6} + (\gamma + \log c)^2\right\}; \qquad (22.42)$$

here γ is Euler's constant.

Proceeding similarly, the product moment of Y'_r and Y'_s $(1 \le r < s \le n)$ can be shown to be

$$E[Y'_r Y'_s] = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \times \sum_{i=0}^{s-r-1} \sum_{j=0}^{n-s} (-1)^{i+j} \binom{s-r-1}{i} \binom{n-s}{j} \times \phi(r+i, s-r-i+j), \tag{22.43}$$

where the function ϕ is the double integral

$$\phi(t,u) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} xy\, e^{x-te^x} e^{y-ue^y}\, dx\, dy, \qquad t,u > 0. \tag{22.44}$$

Lieblein (1953) derived an explicit expression for the ϕ function in (22.44) in terms of Spence's function which has been tabulated quite extensively by Newman (1892) and Abramowitz and Stegun (1965).

Means and variances of order statistics for sample sizes up to 20 have been provided by White (1969); see also Lieblein and Salzer (1957) and McCord (1964). Covariances of order statistics for sample sizes up to 6 have been tabulated by Lieblein (1953,1962) and Lieblein and Zelen (1956). It is of interest to mention that the variance of the largest order statistic is $\pi^2/6$, irrespective of the value of n. Kimball (1946a, 1949) presented an alternative expression for the expected value of Y'_{n-r+1} as

$$E[Y'_{n-r+1}] = \gamma + \sum_{j=1}^{r-1} (-1)^j \binom{n}{j} \Delta^j \log n, \tag{22.45}$$

where A' represents forward difference of the ith order (see Chapter 1, Section A3). Balakrishnan and Chan (1992a) presented tables of means, variances, and covariances of all order statistics for sample sizes n = 1(1)15(5)30. [Although their tables are for order statistics of the distribution of $-Y$, the tables for order statistics of Y are easily obtained from their tables, since $E[Y'_i] = -E[(-Y)'_{n-i+1}]$ and $\operatorname{cov}(Y'_i, Y'_j) = \operatorname{cov}((-Y)'_{n-j+1}, (-Y)'_{n-i+1})$.] Complete tables for all sample sizes up to 30 have also been prepared by Balakrishnan and Chan (1992c). Mahmoud and Ragab (1975) and Provasi (1987) have provided further discussions on order statistics from the extreme value distribution and their moments. The last author has also discussed some approximations to the means, variances, and covariances of order statistics.

In Table 22.3 the means and variances of order statistics are presented for sample sizes up to 10. The covariances of order statistics for sample sizes up to 10 are presented in Table 22.4.

Table 22.3 Means and variances of order statistics from extreme value distribution

n	r	Mean	Variance	n	r	Mean	Variance
1	1	0.57722	1.64493	7	7	2.52313	1.64493
2	1	−0.11593	0.68403	8	1	−0.90212	0.19956
2	2	1.27036	1.64493	8	2	−0.45279	0.18355
3	1	−0.40361	0.44850	8	3	−0.10288	0.19837
3	2	0.45943	0.65852	8	4	0.23121	0.23166
3	3	1.67583	1.64493	8	5	0.58818	0.29005
4	1	−0.57351	0.34402	8	6	1.01107	0.39840
4	2	0.10608	0.41553	8	7	1.58841	0.64642
4	3	0.81278	0.65180	8	8	2.65666	1.64493
4	4	1.96351	1.64493	9	1	−0.94934	0.18395
5	1	−0.69017	0.28486	9	2	−0.52438	0.16390
5	2	−0.10689	0.30850	9	3	−0.20220	0.17158
5	3	0.42555	0.40598	9	4	0.09575	0.19275
5	4	1.07094	0.64907	9	5	0.40053	0.22869
5	5	2.18665	1.64493	9	6	0.73829	0.28844
6	1	−0.77729	0.24658	9	7	1.14745	0.39758
6	2	−0.25453	0.24855	9	8	1.71439	0.64609
6	3	0.18839	0.29762	9	9	2.77444	1.64493
6	4	0.66272	0.40186	10	1	−0.98987	0.17143
6	5	1.27505	0.64770	10	2	−0.58456	0.14879
6	6	2.36898	1.64493	10	3	−0.28369	0.15192
7	1	−0.84596	0.21964	10	4	−0.01204	0.16581
7	2	−0.36531	0.21021	10	5	0.25745	0.18958
7	3	0.02240	0.23701	10	6	0.54361	0.22686
7	4	0.40969	0.29271	10	7	0.86808	0.28739
7	5	0.85248	0.39969	10	8	1.26718	0.39702
7	6	1.44407	0.64691	10	9	1.82620	0.64586
				10	10	2.87980	1.64493

6 RECORD VALUES

Suppose that $Y_1, Y_2, \ldots$ is a sequence of i.i.d. standard type 1 extreme value random variables with density *(22.26)* and that $Y_{L(1)} \equiv Y_1, Y_{L(2)}, \ldots$ are the corresponding lower record values. That is, with $L(1) \equiv 1$ and $L(n) = \min\{j : j > L(n-1), Y_j < Y_{L(n-1)}\}$ for $n = 2, 3, \ldots, \{Y_{L(n)}\}_{n=1}^{\infty}$ forms the lower record value sequence. Then the density function of $Y_{L(n)}, n \geq 1,$ is given by

$$p_{Y_{L(n)}}(y) = \frac{1}{(n-1)!}\{-\log F_Y(y)\}^{n-1} p_Y(y)$$

$$= \frac{1}{(n-1)!} e^{-ny} e^{-e^{-y}}, \qquad -\infty < y < \infty. \qquad (22.46)$$

Table 22.4 Covariances of order statistics from extreme value distribution

n	r	s	Covariance	n	r	s	Covariance	n	r	s	Covariance
2	1	2	0.48045	6	2	6	0.13619	8	1	2	0.12233
3	1	2	0.30137	6	3	4	0.25617	8	1	3	0.09447
3	1	3	0.24376	6	3	5	0.22888	8	1	4	0.07953
3	2	3	0.54629	6	3	6	0.20925	8	1	5	0.07001
4	1	2	0.22455	6	4	5	0.36146	8	1	6	0.06332
4	1	3	0.17903	6	4	6	0.33205	8	1	7	0.05832
4	1	4	0.15389	6	5	6	0.59986	8	1	8	0.05440
4	2	3	0.33721	7	1	2	0.13618	8	2	3	0.14306
4	2	4	0.29271	7	1	3	0.10578	8	2	4	0.12103
4	3	4	0.57432	7	1	4	0.08941	8	2	5	0.10686
5	1	2	0.18203	7	1	5	0.07893	8	2	6	0.09685
5	1	3	0.14359	7	1	6	0.07155	8	2	7	0.08931
5	1	4	0.12258	7	1	7	0.06601	8	2	8	0.08340
5	1	5	0.10901	7	2	3	0.16497	8	3	4	0.16868
5	2	3	0.24677	7	2	4	0.14020	8	3	5	0.14941
5	2	4	0.21227	7	2	5	0.12419	8	3	6	0.13570
5	2	5	0.18967	7	2	6	0.11283	8	3	7	0.12534
5	3	4	0.35267	7	2	7	0.10427	8	3	8	0.11719
5	3	5	0.31716	7	3	4	0.20262	8	4	5	0.20599
5	4	5	0.58992	7	3	5	0.18017	8	4	6	0.18759
6	1	2	0.15497	7	3	6	0.16412	8	4	7	0.17362
6	1	3	0.12122	7	3	7	0.15195	8	4	8	0.16256
6	1	4	0.10292	7	4	5	0.26155	8	5	6	0.26509
6	1	5	0.09116	7	4	6	0.23906	8	5	7	0.24600
6	1	6	0.08285	7	4	7	0.22190	8	5	8	0.23081
6	2	3	0.19671	7	5	6	0.36717	8	6	7	0.37119
6	2	4	0.16806	7	5	7	0.34211	8	6	8	0.34937
6	2	5	0.14945	7	6	7	0.60675	8	7	8	0.61182

This is the density function of a log-gamma population when the shape parameter $\kappa = n$ (see Section 16 or Chapter 17, Section 8.7). Thus, for $n = 1, 2, \dots,$

$$E[Y_{L(n)}] = \gamma - \sum_{i=1}^{n-1} \frac{1}{i}, \quad \operatorname{var}(Y_{L(n)}) = \frac{\pi^2}{6} - \sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{22.47}$$

The joint density function of $Y_{L(m)}$ and $Y_{L(n)}$, $1 \le m < n$, is given by

$$\begin{aligned} p_{Y_{L(m)}, Y_{L(n)}}(y_1, y_2) &= \frac{1}{(m-1)!(n-m-1)!} \{-\log F_Y(y_1)\}^{m-1} \frac{p_Y(y_1)}{F_Y(y_1)} \\ &\quad \times \{-\log F_Y(y_2) + \log F_Y(y_1)\}^{n-m-1} p_Y(y_2) \\ &= \frac{1}{(m-1)!(n-m-1)!} e^{-my_1}(e^{-y_2} - e^{-y_1})^{n-m-1} e^{-y_2} e^{-e^{-y_2}}, \\ &\qquad -\infty < y_2 < y_1 < \infty. \end{aligned} \tag{22.48}$$

Table 22.4 *(Continued)*

n	r	s	Covariance	n	r	s	Covariance	n	r	s	Covariance
9	1	2	0.11167	9	5	7	0.19267	10	3	5	0.11282
9	1	3	0.08580	9	5	8	0.18033	10	3	6	0.10200
9	1	4	0.07199	9	5	9	0.17027	10	3	7	0.09387
9	1	5	0.06322	9	6	7	0.26763	10	3	8	0.08749
9	1	6	0.05706	9	6	8	0.25105	10	3	9	0.08232
9	1	7	0.05246	9	6	9	0.23745	10	3	10	0.07803
9	1	8	0.04887	9	7	8	0.37418	10	4	5	0.14641
9	1	9	0.04597	9	7	9	0.35488	10	4	6	0.13262
9	2	3	0.12700	9	8	9	0.61569	10	4	7	0.12221
9	2	4	0.10703	10	1	2	0.10319	10	4	8	0.11403
9	2	5	0.09424	10	1	3	0.07893	10	4	9	0.10738
9	2	6	0.08522	10	1	4	0.06603	10	4	10	0.10185
9	2	7	0.07846	10	1	5	0.05785	10	5	6	0.17211
9	2	8	0.07315	10	1	6	0.05213	10	5	7	0.15888
9	2	9	0.06886	10	1	7	0.04786	10	5	8	0.14842
9	3	4	0.14525	10	1	8	0.04453	10	5	9	0.13991
9	3	5	0.12825	10	1	9	0.04184	10	5	10	0.13282
9	3	6	0.11620	10	1	10	0.03962	10	6	7	0.20986
9	3	7	0.10712	10	2	3	0.11471	10	6	8	0.19637
9	3	8	0.09998	10	2	4	0.09635	10	6	9	0.18536
9	3	9	0.09419	10	2	5	0.08463	10	6	10	0.17615
9	4	5	0.17074	10	2	6	0.07639	10	7	8	0.26954
9	4	6	0.15503	10	2	7	0.07021	10	7	9	0.25489
9	4	7	0.14315	10	2	8	0.06538	10	7	10	0.24260
9	4	8	0.13377	10	2	9	0.06148	10	8	9	0.37650
9	4	9	0.12615	10	2	10	0.05824	10	8	10	0.35919
9	5	6	0.20823	10	3	4	0.12812	10	9	10	0.61876

Upon writing the joint density of $Y_{L(m)}$ and $Y_{L(n)}$, $1 \le m < n$, in (22.48) as

$$p_{Y_{L(m)}, Y_{L(n)}}(y_1, y_2) = \frac{(n-1)!}{(m-1)!(n-m-1)!} e^{-m(y_1-y_2)} \left(1 - e^{-(y_1-y_2)}\right)^{n-m-1} \times \frac{1}{(n-1)!} e^{-ny_2} e^{-e^{-y_2}}, \qquad -\infty < y_2 < y_1 < \infty, \tag{22.49}$$

we readily observe that $Y_{L(m)} - Y_{L(n)}$ and $Y_{L(n)}$ (for $1 \le m < n$) are statistically independent. As a result we immediately get

$$\operatorname{cov}(Y_{L(m)}, Y_{L(n)}) = \operatorname{var}(Y_{L(n)}) = \frac{\pi^2}{6} - \sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{22.50}$$

These properties are similar to those of order statistics arising from standard exponential random variables (Chapter 19, Section 6). It follows from (22.49) that $Y_{L(m)} - Y_{L(n)}$ is distributed as the $(n-m)$th order statistic in a sample of size n – 1 from the standard exponential distribution, say $Z_{n-m:n-1}$. For the special case when m = 1, we then have $Y_{L(1)} - Y_{L(n)} = Y_1 - Y_{L(n)} \overset{d}{=} Z_{n-1:n-1}$ which, when used with the known results that (see Chapter 19, Section 6)

$$E[Z_{n-1:n-1}] = \sum_{i=1}^{n-1} \frac{1}{i}, \quad \mathrm{var}(Z_{n-1:n-1}) = \sum_{i=1}^{n-1} \frac{1}{i^2}, \tag{22.51}$$

gives easily the expressions for the mean and variance of $Y_{L(n)}$ in (22.47).

Ahsanullah (1990,1991) has used these expressions to develop inference procedures for the location and scale parameters, ξ and θ, of the type 1 extreme value distribution (22.25) based on the first n lower record values $X_{L(1)}, X_{L(2)}, \dots, X_{L(n)}$ observed.

For the standard type 1 distribution in (22.26), we may note the relationship

$$p_Y(y) = F_Y(y)\{-\log F_Y(y)\}, \qquad -\infty < y < \infty. \tag{22.52}$$

By making use of this relationship, Balakrishnan, Ahsanullah, and Chan (1992) established several recurrence relations for single as well as product moments of lower record values from this distribution. For example, consider for n ≥ 1 and $r = 0, 1, 2, \dots,$

$$\begin{aligned} E\left[Y_{L(n)}^r\right] &= \frac{1}{(n-1)!}\int_{-\infty}^{\infty} y^r\{-\log F_Y(y)\}^{n-1} p_Y(y)\, dy \\ &= \frac{1}{(n-1)!}\int_{-\infty}^{\infty} y^r\{-\log F_Y(y)\}^{n} F_Y(y)\, dy \end{aligned}$$

upon using (22.52). Integration by parts yields

$$\begin{aligned} E\left[Y_{L(n)}^r\right] &= \frac{1}{(n-1)!(r+1)}\left[n\int_{-\infty}^{\infty} y^{r+1}\{-\log F_Y(y)\}^{n-1} p_Y(y)\, dy \right. \\ &\qquad \left. - \int_{-\infty}^{\infty} y^{r+1}\{-\log F_Y(y)\}^{n} p_Y(y)\, dy\right] \\ &= \frac{n}{r+1}\left\{E\left[Y_{L(n)}^{r+1}\right] - E\left[Y_{L(n+1)}^{r+1}\right]\right\}. \end{aligned}$$

The equation above, when simply rewritten, yields the recurrence relation

$$E\left[Y_{L(n+1)}^{r+1}\right] = E\left[Y_{L(n)}^{r+1}\right] - \frac{r+1}{n} E\left[Y_{L(n)}^{r}\right] \qquad \text{for } n \geq 1, r = 0, 1, \dots . \tag{22.53}$$

By repeated application of the recurrence relation in (22.53), Balakrishnan, Ahsanullah, and Chan (1992) established the relation

$$E\left[Y_{L(n+1)}^{r+1}\right] = E\left[Y_{L(1)}^{r+1}\right] - (r+1) \sum_{i=1}^{n} \frac{E\left[Y_{L(i)}^{r}\right]}{i}$$

$$\text{for } n = 1, 2, \dots, \ r = 0, 1, 2, \dots . \tag{22.54}$$

From (22.54) one may also easily derive the expressions for the mean and variance of $Y_{L(n)}$ in (22.47).

Proceeding similarly, Balakrishnan, Ahsanullah, and Chan (1992) also established the following recurrence relations for the product moments:

$$E\left[Y_{L(m)}^{r+1} Y_{L(m+1)}^{s}\right] = E\left[Y_{L(m+1)}^{r+s+1}\right] + \frac{r+1}{m} E\left[Y_{L(m)}^{r} Y_{L(m+1)}^{s}\right],$$

$$m \geq 1; r, s = 0, 1, 2, \dots . \tag{22.55}$$

$$E\left[Y_{L(m)}^{r+1} Y_{L(n)}^{s}\right] = E\left[Y_{L(m+1)}^{r+1} Y_{L(n)}^{s}\right] + \frac{r+1}{m} E\left[Y_{L(m)}^{r} Y_{L(n)}^{s}\right],$$

$$1 \leq m \leq n-2; r, s = 0, 1, 2, \dots . \tag{22.56}$$

$$E\left[Y_{L(m)}^{r+1} Y_{L(n)}^{s}\right] = E\left[Y_{L(n)}^{r+s+1}\right] + (r+1) \sum_{i=m}^{n-1} \frac{E\left[Y_{L(i)}^{r} Y_{L(n)}^{s}\right]}{i},$$

$$1 \leq m \leq n-1; r, s = 0, 1, 2, \dots . \tag{22.57}$$

$$E\left[Y_{L(m)}^{r+1} Y_{L(m+1)}^{s}\right] = \sum_{i=0}^{r+1} (r+1)^{(i)} \frac{E\left[Y_{L(m+1)}^{r+s+1-i}\right]}{m^{i}},$$

$$m \geq 1; r, s = 0, 1, 2, \dots . \tag{22.58}$$

and

$$E\left[Y_{L(m)}^{r+1} Y_{L(n)}^{s}\right] = \sum_{i=0}^{r+1} (r+1)^{(i)} \frac{E\left[Y_{L(m+1)}^{r+1-i} Y_{L(n)}^{s}\right]}{m^{i}},$$

$$1 \leq m \leq n-2; r, s = 0, 1, 2, \dots . \tag{22.59}$$

In the equations above

$$k^{(i)} = \begin{cases} 1 & \text{if } i = 0, \\ k(k-1)\cdots(k-i+1) & \text{if } i \geq 1. \end{cases}$$

Suppose that $X'_{i:j}$ is the ith order statistic in a random sample of size j from a distribution $F(\cdot)$. If the distribution function of $(X'_{j:j} - a_j)/b_j$ converges weakly to a nondegenerate distribution function $G(\cdot)$ as $j \to \infty$ for sequences of constants a_j and positive b_j, then Nagaraja (1982) showed that the joint distribution function of $(X'_{j-i+1:j} - a_j)/b_j$, $1 \leq i \leq n$, converges to that of $X_{L(i)}$, $1 \leq i \leq n$. As we have already seen in Section 3, $G(\cdot)$ must be one of the three types of extreme value distributions. Hence, as pointed out by Nagaraja (1988), some inference procedures based on asymptotic theory of extreme order statistics are equivalent to those based on record values from the extreme value distributions. Consequently the asymptotic linear prediction of extreme order statistics discussed by Nagaraja (1984) is the same as predicting a future record value from the distribution $F(\cdot)$. It is also apparent from this discussion that the estimation of parameters of $F(\cdot)$ based on k largest observations discussed by Weissman (1978) is effectively the same as the estimation of parameters based on record values from one of the three extreme value distributions $G(\cdot)$. Smith (1988) has provided a detailed discussion on forecasting records by the maximum likelihood method.

Ballerini and Resnick (1985, 1987a) have discussed upper records arising from the simple linear regression model

$$Z_n = X_n + cn, \qquad n = 1, 2, \ldots, c > 0,$$

where $\{X_n\}$ is i.i.d. type 1 extreme value random variables with density (22.25). They referred to this model as the linear-drift Gumbel record model. Then, for this model, Ballerini and Resnick (1987b) established that the random variables

$$M_n + \max\{Z_1, \cdots, Z_n\} \quad \text{and}$$

$$I_n = \text{Indicator whether record occurs at time n} = I_{[Z_n > M_{n-1}]} \qquad \textbf{(22.60)}$$

are statistically independent for each n (see Section 8 for some additional comments).

Balakrishnan, Balasubramanian, and Panchapakesan (1995) have discussed properties of 6-exceedance records arising from the type 1 extreme value distribution. In this model a new variable will be declared a record only if it is smaller than the previous lower record by at least 6.

7 GENERATION, TABLES, AND PROBABILITY PAPER

The following tables are included in Gumbel (1953):

1. Values of the standard cumulative distribution function, $\exp(-e^{-y})$, and probability density function, $\exp(-y - e^{-y})$, to seven decimal places for $y = -3(0.1) - 2.4(0.05)0.00(0.1)4.0(0.2)8.0(0.5)17.0$
2. The inverse of the cumulative distribution function (i.e., percentiles), $y = -\log(-\log F)$ to five decimal places for

$$F = 0.0001(0.0001)0.0050(0.001)0.988(0.0001)0.9994(0.00001)0.99999.$$

In Owen's tables (1962) there is a similar table, to four decimal places for

$$F = 0.0001(0.0001)0.0010(0.0010)0.0100(0.005)0.100(0.010)0.90(0.005)$$
$$0.990(0.001)0.999(0.0001)1 - 10^{-4(1)7}, 1 - \tfrac{1}{2} \cdot 10^{-4(1)7}.$$

[The special interest in very high values of F, by both Gumbel (1953) and Owen (1962), may be associated with the genesis of the distribution, though it seems rather risky to rely on practical applicability so far out in the tails of a distribution.]

Gumbel (1953) contains others tables. In particular there are two relating to asymptotic distribution of range (see Section 16), and a table giving the probability density function in terms of the cumulative distribution function ($p = -F \log F$) to five decimal places for $F = 0.0001(0.0001)0.0100(0.001)0.999$.

Lieblein and Salzer (1957) have published a table of the expected value (to seven decimal places) of the mth largest among n independent random variables having the standard type 1 extreme value distribution (22.26), for

$$m = 1(1)\min(26, n), \quad n = 1(1)10(5)60(10)100.$$

Lieblein and Zelen (1956) gave the variances and covariances (also to seven decimal places) for sets of 2, 3, 4, 5, and 6 independent type 1 variables. [These values are also given by Lieblein (1962).] Mann (1968b) gave similar tables for the type 1 smallest value distribution for up to 25 variables.

These tables have been extended by White (1969), who gives (up to seven decimal places) expected values and variances of all order statistics for sample sizes 1(1)50(5)100. Extended tables of means, variances, and covariances of order statistics for sample sizes up to 30 have been provided by Balakrishnan and Chan (1992a, c).

Tables of coefficients for the best linear unbiased estimators of ξ and θ and the values of variances and covariance of these estimators have been presented by Balakrishnan and Chan (1992b, d) for the case of complete as well as Type-II censored samples for sample sizes up to 30. Mann

(1967, 1968a, b), and Mann, Schafer, and Singpurwalla (1974) have presented similar tables for the best linear invariant estimates of ξ and σ.

From (22.1) it follows that

$$-\log(-\log \Pr[X < x]) = \frac{x-\xi}{\theta}. \tag{22.61}$$

Hence, if the cumulative observed ***relative frequency*** F_x—equal to (number of observations less than or equal to x)/(total number of observations)—is calculated, and $-\log(-\log F_x)$ is plotted against x, an approximately straight-line relation should be obtained, with slope θ^{-1} and intersecting the horizontal (x) axis at $x = \xi$. In using graph paper with a vertical scale that gives $-\log(-\log F_x)$ directly, it is not necessary to refer to tables of logarithms. Such graph paper is sometimes called ***extreme value probability paper.*** It is also quite common to use such paper with the x-axis vertical, and for practical purposes it is sometimes convenient to have the $-\log(-\log F_x)$ marked not with F_x but with the "return period" $(1 - F_x)^{-1}$; see Gumbel (1949a) and Kimball (1960). Such a paper is called ***extreme probability paper.***

Tables of 500 random numbers (to three decimal places) representing values chosen at random from the standard type 1 distribution, and 500 each from three standard distributions of each of types 2 and 3 [$k^{-1} = 0.2, 0.5, 0.8$ in Eqs. (22.14) and (22.15)] have been given by Goldstein (1963).

Of course pseudorandom numbers from the standard type 1 distribution may be generated easily either through the inverse cdf method along with an efficient uniform random generator (see Chapter 26) or through the relationship with the exponential distribution (explained in Section 3) along with an efficient exponential random generator (see Chapter 19). Sibuya (1967) has discussed the latter. Landwehr, Matalas, and Wallis (1979) have advocated the use of the Lewis-Goodman-Miller algorithm for generating pseudorandom numbers from the uniform distribution for this specific purpose. These authors have also discussed a simulational algorithm for generating serially correlated Gumbel data. Let

$$z_i = \rho_z z_{i-1} + \sqrt{1-\rho_z^2}\,\delta_i$$

represent a Markov process, where p, denotes the first-order serial correlation of the z's and 6, is a standard normal variable independent of z_{i-1}. With δ_i's generated by the Box-Muller algorithm and z_i's determined by the equation above, the serially correlated Gumbel values X_i's may then be obtained as

$$X_i = \xi - \theta \log\{-\log \Phi(z_i)\}$$

where Φ is the standard normal cumulative distribution function.

8 CHARACTERIZATIONS

As mentioned earlier in Section 2, X has a type 1 extreme value distribution if and only if e^X has a **Weibull** distribution, and $e^{X/\theta}$ has an exponential distribution, and $\exp\{(X-\xi)/\theta\}$ has a ***standard*** exponential distribution. It is clear that some characterization theorems for exponential distributions may also be used for type 1 extreme value distributions, simply by applying them to $e^{X/\theta}$, or $\exp\{(X-\xi)/\theta\}$. Dubey (1966) characterizes this distribution by the property that $Y_n = \min(X_1, X_2, \cdots, X_n)$ is a type 1 random variable if and only if $X_1, X_2, \cdots, X_n$ are independent identically distributed type 1 random variables.

Sethuraman (1965) has obtained characterizations of all three types of extreme value distributions, in terms of "complete confounding" of random variables. If X and Y are independent and the distributions of Z, Z given Z = X, and Z given Z = Y are the same [e.g., Z might be equal to $\min(X, Y)$ as in the cases described in **Sethuraman** (1965)], they are said to ***completely confound*** each other with respect to the third. Sethuraman showed that if all pairs from the variables X, Y, and Z completely confound each other with respect to the third and if Y, Z have the same distributions as $a_1X + b_1, a_2X + b_2$, respectively [with $(a_1, b_1) \not\equiv (a_2, b_2)$], then the distribution of X is one of the three extreme value (minimum) distributions (provided we limit ourselves to the cases when $\Pr[X > Y] > 0; \Pr[Y > X] > 0$, etc.). The type of distribution depends on the values of a_1, a_2, b_1, b_2.

Gompertz (1825) derived a probability model for human mortality. He assumed ***the average exhaustion of a man's power to avoid death to be such that at the end of equal infinitely small intervals of time he lost equal portions of his remaining power to oppose destruction which he had at the commencement of these intervals***. From this hypothesis Gompertz (1825) deduced the force of mortality or the hazard function as

$$r(x) = Bc^x, \qquad x \geq 0, B > 0, c \geq 1,$$

which, when solved as a differential equation, readily yields the survival function as

$$1 - F(x) = e^{-B(c^x - 1)/\log c}, \qquad x \geq 0. \tag{22.62}$$

It may be readily seen that (22.62) is a truncated form of the type 1 distribution, and it includes the exponential distribution as a special case when $c = 1$. Then, just as the **memoryless** property

$$\Pr[X \geq x + y | X \geq x] = \Pr[X \geq y] \qquad \text{for all } x, y \geq 0 \tag{22.63}$$

characterizes the exponential distribution (see Chapter 19, Section 8), **Kaminsky** (1982) has characterized the Gompertz distribution in (22.62) through the condition

$$\Pr(X \geq x + y | X \geq x) = \{\Pr(X \geq y)\}^{h(x)}, \qquad x, y \geq 0, \quad (22.64)$$

and the requirement that the function $h(\cdot)$ must take the form $h(x) = c^x$ for some $c \geq 1$.

As one would expect, there are a number of characterizations of the type *1* distribution in the framework of extreme value theory. The most celebrated one is that the type *1* distribution is the only max-stable probability distribution function with the entire real line as its support; for example, see Theorem *1.4.1* in Leadbetter, Lindgren, and Rootzén (1983). In addition to the characterizations of the type *1* distribution itself, there are several characterization results available for the maximal domain of attraction of the type *1* distribution; de Haan (1970) will serve as a good source of information on this as well as characterizations for type *2* and type 3 distributions.

In Section *6* we have discussed the linear-drift Gumbel record model. We mentioned that under this model the random variables M_n and I_n are statistically independent for each n. Ballerini (1987) has proved this to be a characterization of the type *1* extreme value distribution; that is, M_n and I_n are independent for each n and for every c > 0 if and only if the X_i's are type *1* extreme value random variables.

Tikhov (1991) has characterized the extreme value distributions by the limiting information quantity associated with the maximum likelihood estimator based on a multiply censored sample.

9 METHODS OF INFERENCE

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from the type *1* extreme value distribution in (22.25). Then, as Downton (1966) has shown, the Cramér-Rao lower bounds of variances of unbiased estimators of ξ and θ are given by

$$\{1 + 6(1-\gamma)^2\pi^{-2}\}\theta^2 n^{-1} = 1.10867\theta^2 n^{-1},$$

$$6\pi^{-2}\theta^2 n^{-1} = 0.60793\theta^2 n^{-1}, \qquad (22.65)$$

respectively.

As has already been mentioned on several occasions in this chapter, as well as in Chapter *21*, if Z has a Weibull distribution with probability density function

$$p_Z(z) = \frac{c}{\beta}\left(\frac{z-\xi_0}{\beta}\right)^{c-1} e^{-[(z-\xi_0)/\beta]^c}, \qquad z \geq \xi_0, \qquad (22.66)$$

then $\log(Z - \xi_0)$ has a type 1 extreme value distribution. Consequently, if ξ_0 is known, the methods of estimation discussed in this section for the type 1 extreme value distribution can also be used for estimating the parameters β and c of the Weibull distribution (22.66). Conversely, as discussed in Section 4, some methods of estimating β and c of the Weibull distribution, when ξ_0 is known, can also be used for estimating the parameters ξ and θ of the type 1 extreme value distribution.

9.1 Moment Estimation

Let $\bar{X}$ and S denote the sample mean and the sample standard deviation. Then, using Eqs. (22.29) and (22.31), we simply obtain the moment estimates of θ and ξ as

$$\tilde{\theta} = \frac{\sqrt{6}}{\pi} S \quad \text{and} \quad \tilde{\xi} = \bar{X} - \gamma\tilde{\theta}. \tag{22.67}$$

Tiago de Oliveira (1963) has shown that

$$\operatorname{var}(\tilde{\xi}) \simeq \frac{\theta^2}{n}\left\{\frac{\pi^2}{6} + \frac{\gamma^2}{4}(\beta_2 - 1) - \frac{\pi}{\sqrt{6}}\gamma\sqrt{\beta_1}\right\} \tag{22.68}$$

and that

$$\operatorname{var}(\tilde{\theta}) \simeq \frac{\theta^2}{4n}(\beta_2 - 1), \tag{22.69}$$

where β_1 and β_2 are the coefficients of skewness and kurtosis as given in (22.32). Upon substituting for their values, we get

$$\operatorname{var}(\tilde{\xi}) \simeq \frac{1.1678\theta^2}{n} \text{ and } \operatorname{Var}(\tilde{\theta}) \simeq \frac{1.1\theta^2}{n}. \tag{22.70}$$

Tiago de Oliveira (1963) has also discussed the joint distribution of $\bar{X}$ and S.

A comparison of the variance formulas in (22.70) with the Cramér-Rao lower bounds in (22.65) readily reveals that the moment estimator $\tilde{\xi}$ has about 95% efficiency while the moment estimator $\tilde{\theta}$ has only about 55% efficiency. The estimators $\tilde{\xi}$ and $\tilde{\theta}$ are both $\sqrt{n}$-consistent; that is, $\sqrt{n}(\tilde{\xi} - \xi)$ and $\sqrt{n}(\tilde{\theta} - \theta)$ are bounded in probability.

Tiago de Oliveira (1963) has shown that the joint asymptotic distribution of $\tilde{\xi}$ and $\tilde{\theta}$ is bivariate normal with mean vector $(\xi, \theta)'$, variances as given in

(22.70), and the correlation coefficient as

$$\rho_{\hat{\xi},\hat{\theta}} = \frac{\pi^2\left[\sqrt{\beta_1} - 3\gamma(\beta_2 - 1)/2\pi\right]/6}{\left[\left\{\pi^2/6 + \gamma^2(\beta_2 - 1)/4 - \pi\left(\gamma\sqrt{\beta_1}\right)/\sqrt{6}\right\}(\beta_2 - 1)\right]^{1/2}}$$

$$\simeq 0.123. \tag{22.71}$$

By making use of this asymptotic result, asymptotic confidence regions for (ξ, θ) can be constructed.

Christopeit (1994) recently showed that the method of moments provides consistent estimates of the parameters of extreme value distributions, and used the estimation of the distribution of earthquake magnitudes in the middle Rhein region for illustration.

9.2 Simple Linear Estimation

Upon noting that the likelihood equations for ξ and 8 do not admit explicit solutions and hence need to be solved by numerical iterative methods, Kimball (1956) suggested a simple modification to the equation for 8 (based on the equation for ξ) that makes it easier to solve the resulting equation. The equation for 8 given by

$$\hat{\theta} = \bar{X} - \frac{\sum_{i=1}^{n} X_i\, e^{-X_i/\hat{\theta}}}{\sum_{i=1}^{n} e^{-X_i/\hat{\theta}}}, \tag{22.72}$$

used in conjunction with the equation for ξ given by

$$\hat{\xi} = -\hat{\theta}\log\left\{\frac{1}{n}\sum_{i=1}^{n} e^{-X_i/\hat{\theta}}\right\}, \tag{22.73}$$

can be rewritten as

$$\begin{aligned}\hat{\theta} &= \bar{X} - \frac{1}{n}\sum_{i=1}^{n} X_i\, e^{-(X_i-\hat{\xi})/\hat{\theta}} \\ &= \bar{X} + \frac{1}{n}\sum_{i=1}^{n} X_i \log \hat{F}_X(X_i),\end{aligned} \tag{22.74}$$

where $\hat{F}_X(X_i)$ is the estimated cumulative distribution function. By replacing log $\hat{F}_X(X_i')$ in (22.74) with the expected value of log $\hat{F}_X(X_i')$, Kimball (1956) derived a simplified linear estimator for 8 as

$$\hat{\theta}^* = \bar{X} + \frac{1}{n}\sum_{i=1}^{n} X_i'\left(\sum_{j=i}^{n}\frac{1}{j}\right) \tag{22.75}$$

which may be further approximated as

$$\hat{\theta}^* \simeq \bar{X} + \sum_{i=1}^{n} X_i' \log\left(\frac{i-\frac{1}{2}}{n+\frac{1}{2}}\right). \tag{22.76}$$

The estimator in (22.75) or in (22.76) is a linear function of the order statistics, and hence its bias and mean square error can be determined easily from means, variances and covariances of order statistics in Tables 22.3 and 22.4. Since the linear estimator in (22.76) is biased, Kimball (1956) presented a table of corrective multipliers to make it unbiased; from the table it appears that for n ≥ 10 the estimator

$$\frac{\hat{\theta}^*}{1 + 2.3n^{-1}} \tag{22.77}$$

is very nearly unbiased. Further a simplified linear estimator of ξ may then be obtained as

$$\text{Estimator of } \xi = \bar{X} - \gamma \times (\text{Estimator of } \theta). \tag{22.78}$$

Due to the linearity of the estimator of 8, it is only natural to compare it with the best linear unbiased estimator of 8 and with its approximations proposed by Blom (1958) and Weiss (1961).

Downton (1966) carried out a number of comparisons of this nature. He actually discussed the type 1 distribution appropriate to minima, with cumulative distribution function $1 - e^{-e^{(x-\xi)/\theta}}$, but his results also apply to the type 1 distribution in (22.1) (with some simple changes). His results are all in terms of efficiencies, that is, ratios of the values given by (22.65) to corresponding variances for the estimators in question. For each estimator of 8, the parameter ξ was estimated from (22.78). Tables 22.5 and 22.6, taken from Downton (1966), give efficiencies for various estimators of ξ and 8.

For the small values of n considered, the asymptotic formulas used in the calculations may not be accurate, yet the tables probably give a good idea of relative efficiency and the performance of different estimators considered. It can be seen from Table 22.5 that the location parameter ξ can be estimated with quite good accuracy using simple linear functions of order statistics; however, it may also be noted from Table 22.6 that the situation is rather unsatisfactory should one use such simple linear functions of order statistics to estimate the scale parameter 8.

Table 22.5 Efficiencies of linear unbiased estimators of ξ for the extreme value distribution

n	2	3	4	5	6	∞
Best linear	84.05	91.73	94.45	95.82	96.65	100.00
Blom's approximation	84.05	91.72	94.37	95.68	96.45	100.00
Weiss's approximation	84.05	91.73	94.41	95.74	96.53	—
Kimball's approximation	84.05	91.71	94.45	95.82	96.63	—

Note: Efficiencies are expressed in percentages.

For the case of a Type-I1 right-censored sample $X_1', X_2', \cdots, X_{n-s}'$ from the type 1 extreme value distribution for minima with cdf $F_X(x) = 1 - e^{-e^{(x-\xi)/\theta}}$, Bain (1972) suggested a simple unbiased linear estimator for the scale parameter 8. This estimator was subsequently modified by Engelhardt and Bain (1973) to the form

$$\hat{\hat{\theta}} = \frac{1}{nk_{n-s,n}} \sum_{i=1}^{n-s} |X_r' - X_i'|, \tag{22.79}$$

where

$$k_{n-s,n} = \frac{1}{n} \sum_{i=1}^{n-s} E|Y_r' - Y_i'|, \tag{22.80}$$

$Y_i' = (X_i' - \xi)/\theta$ being the order statistics from the standard type 1 extreme

Table 22.6 Efficiencies of linear unbiased estimators of θ for the extreme value distribution

n	2	3	4	5	6	∞
Best linear	42.70	58.79	67.46	72.96	76.78	100.00
Blom's approximation	42.70	57.47	65.39	70.47	74.07	100.00
Weiss's approximation	42.70	58.00	66.09	71.04	74.47	—
Kimball's approximation	42.70	57.32	65.04	69.88	73.25	—

Note: Efficiencies are expressed in percentages.

value distribution for minima, and

$$
\begin{aligned}
r &= n - s && \text{for } n - s \le 0.9n, \\
r &= n && \text{for } n - s = n,\ n \le 15, \\
r &= n - 1 && \text{for } n - s = n,\ 16 \le n \le 24, \\
r &= [0.892n] + 1 && \text{for } n - s = n,\ n \ge 25.
\end{aligned}
$$

By making use of the tables of means of order statistics referred to in Section 5, Bain (1972) determined exact values of $k_{n-s,n}$ for n = 5, 15, 20, 30, 60, and 100 and n infinite and $(n-s)/n = 0.1(0.1)0.9$ for integer $n - s$. Engelhardt and Bain (1973) gave exact values of $k_{n,n}$ for $n = 2(1)35(5)100$, n = 39, 49, and 59 and infinite n. Mann and Fertig (1975) also presented exact values of $k_{n-s,n}$ for $n = 25(5)60$ and $(n-s)/n = 0.1(0.1)1.0$ for integer $n - s$. [It needs to be mentioned that the values of $k_{n,n}$ given by Mann and Fertig (1975) are slightly different from those given by Engelhardt and Bain (1973) for n > 40 as the choice of r used by the former is different.]

Since θ is a scale parameter and $\hat{\hat{\theta}}$ is an unbiased estimator of θ, improvement is possible in terms of minimum mean-square-error estimator (see Section 9.3 for more details). The improvement in efficiency becomes considerable when the censoring is heavy in the sample. As Bain (1972) noted that for $(n-s)/n$ about at most 0.5, $\operatorname{var}(\hat{\hat{\theta}}) \simeq \theta^2/(nk_{n-s,n})$ and consequently

$$\frac{\hat{\hat{\theta}}}{1 + 1/(nk_{n-s,n})} = \frac{1}{1 + nk_{n-s,n}} \sum_{i=1}^{n-s} (X'_{n-s} - X'_i) \tag{22.81}$$

has a smaller mean square error than $\hat{\hat{\theta}}$ when $(n-s)/n \le 0.5$. On these grounds, an estimator that has been used in general is

$$\frac{\hat{\hat{\theta}}}{1 + l_{n-s,n}} = \frac{1}{nk_{n-s,n}(1 + l_{n-s,n})} \sum_{i=1}^{n-s} |X'_r - X'_i| \tag{22.82}$$

which has mean square error $\theta^2 l_{n-s,n}/(1 + l_{n-s,n})$; here, $l_{n-s,n} = \operatorname{var}(\hat{\hat{\theta}}/\theta)$. Values of $l_{n-s,n}$ have been tabulated by Engelhardt and Bain (1973) and Mann and Fertig (1975). From the tables of Bain (1972) and Engelhardt and Bain (1973), it is clear that the estimator $\hat{\hat{\theta}}$ in (22.79) is highly efficient; for example, when $(n-s)/n \le 0.7$, the asymptotic efficiency of $\hat{\hat{\theta}}$ relative to the Cramér-Rao lower bound is at least 97.7%.

The estimator $\hat{\hat{\theta}}$ in (22.79) may also be used to produce a simple linear unbiased estimator for ξ, through the moment equation

$$X'_r = E[X'_r] = \xi + \theta E[Y'_r], \tag{22.83}$$

as

$$\hat{\hat{\xi}} = X'_r - E[Y'_r]\hat{\hat{\theta}}. \qquad (22.84)$$

Using the estimators $\hat{\hat{\theta}}$ and $\hat{\hat{\xi}}$ in Eqs. (22.79) and (22.84), respectively, a simple linear unbiased estimator for the pth quantile ξ_p can be derived as

$$\hat{\hat{\xi}}_p = \hat{\hat{\xi}} + \hat{\hat{\theta}} \log(-\log(1-p)), \qquad 0 < p < 1. \qquad (22.85)$$

Confidence intervals for the parameters ξ and θ based on the linear unbiased estimators $\hat{\hat{\xi}}$ and $\hat{\hat{\theta}}$ have also been discussed. Bain (1972) suggested approximating the distribution of $2nk_{n-s,n}\hat{\theta}/\theta$ by a central chi-square distribution with $2nk_{n-s,n}$ degrees of freedom when $(n-s)/n$ is at most 0.5 and n at least 20. But Mann and Fertig (1975) have shown that for $n \geq 20$, $2(\hat{\theta}/\theta)/l_{n-s,n}$ is approximately distributed as chi-square with $2/l_{n-s,n}$ degrees of freedom. Interestingly this approximate result holds for all values of $(n-s)/n$ in $(0, 1]$. This approximation arose from an observation of van Montfort (1970) that the statistics

$$Z_i = \frac{X'_{i+1} - X'_i}{\{E[Y'_{i+1}] - E[Y'_i]\}\theta}, \qquad i = 1, 2, \ldots, n-1,$$

all have approximately an exponential distribution with mean exactly 1, variance approximately 1, and covariance almost zero [see also Pyke (1965)].

As aptly pointed out by Mann and Fertig (1975), since for $n - s \leq 0.90n$,

$$\begin{aligned}\frac{\hat{\theta}}{\theta} &= \frac{1}{nk_{n-s,n}} \sum_{i=1}^{n-s} \frac{X'_{n-s} - X'_i}{\theta} \\ &= \frac{1}{2nk_{n-s,n}} \sum_{i=1}^{n-s-1} \{E[Y'_{i+1}] - E[Y'_i]\} 2Z_i\end{aligned}$$

is approximately a sum of weighted independent chi-square variables, various approximations discussed in Chapter 18 for this distribution are useful in developing approximate inference for θ.

9.3 Best Linear Unbiased (Invariant) Estimation

Let $X'_{r+1} \leq X'_{r+2} \leq \cdots \leq X'_{n-s}$ be the available doubly Type-II censored sample from a sample of size n where the smallest r and the largest s

observations have been censored. Let us denote

$$\boldsymbol{X} = (X'_{r+1}, X'_{r+2}, \cdots, X'_{n-s})^T,$$

$$\mathbf{1} = (1, 1, \cdots, 1)^T_{1\times(n-r-s)},$$

$$\boldsymbol{\mu} = (E[Y'_{r+1}], E[Y'_{r+2}], \cdots, E[Y'_{n-s}])^T,$$

and

$$\boldsymbol{\Sigma} = \big((\mathrm{cov}(Y'_i, Y'_j))\big), \qquad r + 1 \le i, j \le n - s,$$

where $E[Y'_i]$ and $\mathrm{cov}(Y'_i, Y'_j)$ are as derived in Section 5. Then, by minimizing the generalized variance

$$(\boldsymbol{X} - \xi\mathbf{1} - \theta\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \xi\mathbf{1} - \theta\boldsymbol{\mu}),$$

we derive the best linear unbiased estimators (BLUEs) of ξ and θ as [see Balakrishnan and Cohen (1991, pp. 80–81)]

$$\xi^* = \left\{\frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\mathbf{1}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}\right\}\boldsymbol{X}$$
$$= \sum_{i=r+1}^{n-s} a_i X'_i \qquad (22.86)$$

and

$$\theta^* = \left\{\frac{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} - \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}\right\}\boldsymbol{X}$$
$$= \sum_{i=r+1}^{n-s} b_i X'_i. \qquad (22.87)$$

Further, the variances and covariance of these estimators are given by

$$\mathrm{var}(\xi^*) = \frac{\theta^2(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}$$
$$= \theta^2 V_1, \qquad (22.88)$$

$$\mathrm{var}(\theta^*) = \frac{\theta^2(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}$$
$$= \theta^2 V_2, \qquad (22.89)$$

and

$$\operatorname{cov}(\xi^*, \theta^*) = -\frac{\theta^2(\mu^T\Sigma^{-1}\mathbf{1})}{(\mu^T\Sigma^{-1}\mu)(\mathbf{1}^T\Sigma^{-1}\mathbf{1}) - (\mu^T\Sigma^{-1}\mathbf{1})^2}$$

$$= \theta^2 V_3. \tag{22.90}$$

Lieblein (1962) presented tables of coefficients a_i and b_i in (22.86) and (22.87), and the variances and covariance in (22.88)–(22.901, for samples sizes up to 6. These tables were extended by White (1964) for sample sizes up to 20, by Mann (1967) for sample sizes up to 25, and by Balakrishnan and Chan (1992b, d) for sample sizes up to 30 (for the case of complete as well as censored samples).

In Table 22.7, for example, the coefficients a_i and b_i are presented for n = 2(1)10 for the case of complete samples (i.e., r = s = 0). The corresponding values of the variance and covariance factors (V_1, V_2, and V_3) are presented in Table 22.8.

Table 22.7 Coefficients for the BLUEs of ξ and θ for complete samples

n	i	a_i	b_i	n	i	a_i	b_i
2	1	0.91637	−0.72135	8	1	0.27354	−0.39419
2	2	0.08363	0.72135	8	2	0.18943	− 0.07577
3	1	0.65632	−0.63054	8	3	0.15020	0.01112
3	2	0.25571	0.25582	8	4	0.12117	0.05893
3	3	0.08797	0.37473	8	5	0.09714	0.08716
4	1	0.51100	−0.55862	8	6	0.07590	0.10273
4	2	0.26394	0.08590	8	7	0.05613	0.10807
4	3	0.15368	0.22392	8	8	0.03649	0.10194
4	4	0.07138	0.24880	9	1	0.24554	−0.36924
5	1	0.41893	−0.50313	9	2	0.17488	−0.08520
5	2	0.24628	0.00653	9	3	0.14179	−0.00649
5	3	0.16761	0.13045	9	4	0.11736	0.03798
5	4	0.10882	0.18166	9	5	0.09722	0.06557
5	5	0.05835	0.18448	9	6	0.07957	0.08265
6	1	0.35545	−0.45927	9	7	0.06340	0.09197
6	2	0.22549	−0.03599	9	8	0.04796	0.09437
6	3	0.16562	0.07320	9	9	0.03229	0.08839
6	4	0.12105	0.12672	10	1	0.22287	− 0.34783
6	5	0.08352	0.14953	10	2	0.16231	−0.09116
6	6	0.04887	0.14581	10	3	0.13385	−0.01921
7	1	0.30901	−0.42370	10	4	0.11287	0.02218
7	2	0.20626	−0.06070	10	5	0.09564	0.04867
7	3	0.15859	0.03619	10	6	0.08062	0.06606
7	4	0.12322	0.08734	10	7	0.06699	0.07702
7	5	0.09375	0.11487	10	8	0.05419	0.08277
7	6	0.06733	0.12586	10	9	0.04175	0.08355
7	7	0.04184	0.12014	10	10	0.02893	0.07794

Table 22.8 Values of V_1, V_2, and V_3 for the BLUEs of ξ and θ for complete samples

n	V_1	V_2	V_3
2	0.65955	0.71186	−0.06432
3	0.40286	0.34471	0.02477
4	0.29346	0.22528	0.03469
5	0.23140	0.16665	0.03399
6	0.19117	0.13196	0.03137
7	0.16293	0.10910	0.02860
8	0.14198	0.09292	0.02608
9	0.12582	0.08088	0.02388
10	0.11297	0.07157	0.02198

Hassanein (1964) discussed the use of nearly best linear unbiased estimators and presented tables of coefficients of order statistics from censored samples for $n = 2(1)10(5)25$.

Observing that these estimators are minimum variance estimators in the class of all linear unbiased estimators, Mann (1969) considered the larger class of all linear estimators and derived improved estimators by minimizing the mean square error. Specifically, by considering the best linear unbiased estimators θ^* and $\eta^* = c_1\xi^* + c_2\theta^*$ of the parameters θ and $\eta = c_1\xi + c_2\theta$ and their respective variances $\theta^2 V_2$ and $\theta^2 V_4$ (where $V_4 = c_1^2 V_1 + c_2^2 V_2 + 2c_1c_2V_3$) and covariance $\theta^2 V_5$ (where $V_5 = c_1 V_3 + c_2 V_2$), Mann *(1969)* showed that the unique minimum-mean-square-error linear estimators of θ and η are given by

$$\theta^{**} = \frac{\theta^*}{1 + V_2} \quad \text{and} \quad \eta^{**} = \eta^* - \left(\frac{V_5}{1 + V_2}\right)\theta^*. \tag{22.91}$$

The mean square errors for these estimators are

$$\theta^2\left(\frac{V_2}{1 + V_2}\right) \quad \text{and} \quad \theta^2\left\{V_4 - \frac{V_5^2}{1 + V_2}\right\}, \tag{22.92}$$

respectively. These estimators are termed the best linear invariant estimators (BLIEs) by Mann *(1969)*. They become particularly useful when either the sample size is very small or there is a great deal of censoring in the sample. Of course the best linear invariant estimator of ξ may be derived from *(22.91)* by setting $c_1 = 1$ and $c_2 = 0$; similarly the best linear invariant estimator of the pth quantile ξ_p may be derived from *(22.91)* by setting $c_1 = 1$ and $c_2 = -\log(-\log p)$.

Denoting the best linear invariant estimators of ξ and 8 by

$$\xi^{**} = \sum_{i=r+1}^{n-s} a_i^* X_i' \quad \text{and} \quad 8^{**} = \sum_{i=r+1}^{n-s} b_i^* X_i', \tag{22.93}$$

and their respective mean square errors by

$$\text{MSE}(\xi^{**}) = \theta^2 W_1 \quad \text{and} \quad \text{MSE}(\theta^{**}) = \theta^2 W_2, \tag{22.94}$$

Mann (1967a, b) and Mann, Schafer, and Singpurwalla (1974) have presented tables for various sample sizes and different levels of censoring.

In Table 22.9, for example, the coefficients a_i^* and b_i^* are presented for n = 2(1)10 for the case of complete samples (i.e., $r = s = 0$). The corresponding values of the mean square error factors (W_1 and W_2) are presented in Table 22.10. A comparison of the entries in Tables 22.8 and 22.10 readily

Table 22.9 Coefficientsfor the BLIEs of ξ and θ for complete samples

n	i	a_i^*	b_i^*	n	i	a_i^*	b_i^*
2	1	0.88927	−0.42138	8	1	0.28294	−0.36068
2	2	0.11073	0.42138	8	2	0.19124	−0.06933
3	1	0.66794	−0.46890	8	3	0.14993	0.01018
3	2	0.25100	0.19024	8	4	0.11977	0.05392
3	3	0.08106	0.27867	8	5	0.09506	0.07975
4	1	0.52681	−0.45591	8	6	0.07345	0.09399
4	2	0.26151	0.07011	8	7	0.05355	0.09889
4	3	0.14734	0.18275	8	8	0.03405	0.09327
4	4	0.06434	0.20305	9	1	0.25370	−0.34161
5	1	0.43359	−0.43126	9	2	0.17676	−0.07883
5	2	0.24609	0.00560	9	3	0.14193	−0.00600
5	3	0.16381	0.11182	9	4	0.11652	0.03514
5	4	0.10353	0.15571	9	5	0.09577	0.06067
5	5	0.05298	0.15813	9	6	0.07774	0.07647
6	1	0.36818	−0.40573	9	7	0.06137	0.08508
6	2	0.22649	−0.03180	9	8	0.04587	0.08731
6	3	0.16359	0.06467	9	9	0.03034	0.08178
6	4	0.11754	0.11195	10	1	0.23000	−0.32460
6	5	0.07938	0.13210	10	2	0.16418	−0.08507
6	6	0.04483	0.12881	10	3	0.13424	−0.01793
7	1	0.31993	−0.38202	10	4	0.11241	0.02070
7	2	0.20783	−0.05472	10	5	0.09464	0.04542
7	3	0.15766	0.03263	10	6	0.07926	0.06165
7	4	0.12097	0.07875	10	7	0.06541	0.07188
7	5	0.09079	0.10357	10	8	0.05250	0.07724
7	6	0.06409	0.11348	10	9	0.04003	0.07797
7	7	0.03874	0.10832	10	10	0.02733	0.07273

Table 22.10 Values of W_1 and W_2 for the BLIEs of ξ and θ for complete samples

n	W_1	W_2
2	0.65713	0.41584
3	0.40241	0.25635
4	0.29248	0.18386
5	0.23040	0.14284
6	0.19030	0.11658
7	0.16219	0.09836
8	0.14136	0.08502
9	0.12530	0.07482
10	0.11252	0.06679

reveals that while there is only a slight improvement in the estimation of ξ, there is a significant gain in using the BLIE of θ particularly when n is small. McCool (1965) discussed the construction of good linear unbiased estimates from the best linear estimates in the case of small sample sizes.

9.4 Asymptotic Best Linear Unbiased Estimation

Johns and Lieberman (1966) tabulated approximate weights for obtaining BLIEs of the parameters ξ and θ based on the first $n - s$ order statistics of samples of size n for n = 10, 15, 20, 30, 50, and 100 and four values of s for each n. Johns and Lieberman (1966) also presented formulas for determining weights for the asymptotic optimal linear estimates in the case of Type-II censored samples. Of course, as mentioned earlier in Section 9.3, exact tables of weights for the BLIEs have been presented by Mann (1967a, b) for sample sizes up to 25 and $s = 0(1)n - 2$.

Optimal linear estimation of the parameters ξ and θ based on k selected order statistics, using the theory of Ogawa (1951, 1952), has also been discussed by a number of authors. Suppose that $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < 1$ is the spacing that needs to be determined optimally, and let $\lambda_0 = 0$ and $\lambda_{k+1} = 1$. $X'_{n_i:n}$ is termed the sample quantile of order λ_i, where $n_i = [n\lambda_i] + 1$. Then it can be shown that the asymptotic variances and covariance of the BLUEs, $\tilde{\xi}^*$ and $\tilde{\theta}^*$, based on the k selected sample quantiles are given by

$$\operatorname{var}(\tilde{\xi}^*) = \frac{\theta^2}{n} \cdot \frac{K_{22}}{K_{11}K_{22} - K_{12}^2}, \tag{22.95}$$

$$\operatorname{var}(\tilde{\theta}^*) = \frac{\theta^2}{n} \cdot \frac{K_{11}}{K_{11}K_{22} - K_{12}^2}, \tag{22.96}$$

and

$$\operatorname{cov}(\tilde{\xi}^*, \tilde{\theta}^*) = -\frac{\theta^2}{n} \cdot \frac{K_{12}}{K_{11}K_{22} - K_{12}^2} \tag{22.97}$$

In the equations above

$$K_{11} = \sum_{i=1}^{k+1} \frac{\{p_Y(G_i) - p_Y(G_{i-1})\}^2}{\lambda_i - \lambda_{i-1}}, \tag{22.98}$$

$$K_{12} = \sum_{i=1}^{k+1} \frac{\{p_Y(G_i) - p_Y(G_{i-1})\}\{G_i p_Y(G_i) - G_{i-1} p_Y(G_{i-1})\}}{\lambda_i - \lambda_{i-1}}, \tag{22.99}$$

and

$$K_{22} = \sum_{i=1}^{k+1} \frac{\{G_i p_Y(G_i) - G_{i-1} p_Y(G_{i-1})\}^2}{\lambda_i - \lambda_{i-1}}, \tag{22.100}$$

where $G_i = F_Y^{-1}(\lambda_i)$ and

$$p_Y(G_0) = G_0 p_Y(G_0) = p_Y(G_{k+1}) = G_{k+1} p_Y(G_{k+1}) = 0.$$

Appropriate functions involving K_{11}, K_{22}, and K_{12} need to be optimized, subject to the constraint $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < 1$ in order to determine the k optimal quantiles for the asymptotic best linear unbiased estimation of the parameters ξ and θ. Numerical results for this problem have been provided by Hassanein (1965,1968,1969,1972) and Chan and Kabir (1969), while optimal t-tests based on these estimators have been discussed by Chan and Mead (1971a, b), and Chan, Cheng, and Mead (1972). Similar estimation of the αth quantile of the distribution, given by $X_\alpha = \xi - \theta \log(-\log \alpha)$ for $0 < \alpha < 1$, based on k optimally selected order statistics has been discussed in great detail by Hassanein, Saleh, and Brown (1984,1986) and Hassanein and Saleh (1992).

For example, the optimal spacing $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ that maximizes K_{11} in (22.98) is presented in Table 22.11 for k = 1(1)10. These values then give the optimal sample quantiles to be used in a sample of size n for the asymptotic best linear unbiased estimator of ξ (when θ is known) since its variance in this case is given by

$$\operatorname{var}(\tilde{\xi}^*) = \frac{\theta^2}{nK_{11}}. \tag{22.101}$$

More elaborate tables may be found in the papers mentioned above. Tests of hypotheses concerning the equality of ξ_i's from l extreme value populations, based on these asymptotic best linear unbiased estimators, have been discussed by Hassanein and Saleh (1992).

Table 22.11 Optimal spacing for the asymptotic best linear unbiased estimator of ξ (when θ is known) for k = 1(1)10

k	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
1	0.2032									
2	0.0734	0.3615								
3	0.0345	0.1701	0.4705							
4	0.0190	0.0933	0.2581	0.5486						
5	0.0115	0.0566	0.1566	0.3329	0.6069					
6	0.0075	0.0369	0.1021	0.2171	0.3958	0.6521				
7	0.0052	0.0254	0.0703	0.1494	0.2723	0.4487	0.6880			
8	0.0037	0.0182	0.0504	0.1071	0.1953	0.3218	0.4935	0.7173		
9	0.0027	0.0135	0.0374	0.0794	0.1448	0.2386	0.3659	0.5319	0.7415	
10	0.0021	0.0103	0.0285	0.0605	0.1103	0.1818	0.2788	0.4052	0.5650	0.7619

9.5 Linear Estimation with Polynomial Coefficients

Based on a complete ordered sample $X_1', X_2', \cdots, X_n'$ from the type 1 extreme value distribution (22.25), Downton (1966) considered estimators of the form

$$\xi_* = \sum_{k=0}^{p} (k+1)\alpha_k \sum_{i=1}^{n} \frac{(i-1)^{(k)} X_i'}{n^{(k+1)}} \tag{22.102}$$

and

$$\theta_* = \sum_{k=0}^{p} (k+1)\beta_k \sum_{i=1}^{n} \frac{(i-1)^{(k)} X_i'}{n^{(k+1)}}, \tag{22.103}$$

where

$$\begin{aligned} m^{(r)} &= 1 && \text{if } r = 0 \\ &= m(m-1)\cdots(m-r+1) && \text{if } r = 1, 2, \ldots . \end{aligned}$$

Let us denote

$$\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \cdots, \alpha_p)^T,$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T,$$

$$\mathbf{1} = (1, 1, \cdots, 1)^T_{1\times(p+1)},$$

$$\boldsymbol{\mu} = \left(E[Y_{1:1}'], E[Y_{2:2}'], \cdots, E\left[Y_{p+1:p+1}'\right]\right)^T,$$

and

$$\boldsymbol{\Sigma} = ((\Sigma_{k,l}))_{(p+1)\times(p+1)},$$

where

$$\Sigma_{k,l} = \text{cov}\left((k+1)\sum_{i=1}^{n}\frac{(i-1)^{(k)}Y_i'}{n^{(k+1)}}, (l+1)\sum_{i=1}^{n}\frac{(i-1)^{(l)}Y_i'}{n^{(l+1)}}\right),$$

$$0 \le k, l \le p. \quad (22.104)$$

Then, by using least-squares theory, Downton (1966) derived the coefficients for the best linear unbiased estimators with polynomial coefficients ξ_* and θ_* in (22.102) and (22.103) as

$$[\boldsymbol{\alpha}\ \boldsymbol{\beta}] = \boldsymbol{\Sigma}^{-1}[\mathbf{1}\quad \boldsymbol{\mu}]\left[\begin{bmatrix}\mathbf{1}^T\\ \boldsymbol{\mu}^T\end{bmatrix}\boldsymbol{\Sigma}^{-1}[\mathbf{1}\ \boldsymbol{\mu}]\right]^{-1} \quad (22.105)$$

and the variance-covariance matrix of ξ_* and θ_* as

$$\begin{bmatrix}\text{var}(\xi_*) & \text{cov}(\xi_*, \theta_*)\\ & \text{var}(\theta_*)\end{bmatrix} = \theta^2\left[\begin{bmatrix}\mathbf{1}^T\\ \boldsymbol{\mu}^T\end{bmatrix}\boldsymbol{\Sigma}^{-1}[\mathbf{1}\ \boldsymbol{\mu}]\right]^{-1}; \quad (22.106)$$

for details, see Balakrishnan and Cohen (1991, pp. 109–113).

Downton (1966) examined the efficiency of these estimators, and compared their performance with many other estimators. For example, the efficiencies of the linear coefficients estimator and the quadratic coefficients estimator for ξ and 8 are presented in Tables 22.12 and 22.13 (which may be compared with the entries in Tables 22.5 and 22.6).

Furthermore, as with other methods of estimation, the linear estimators ξ_* and θ_* may be used to estimate the parameter $c_1\xi + c_2\theta$ by $c_1\xi_* + c_2\theta_*$, which may be shown to be the best linear unbiased estimator with polynomial coefficients of the parameter of interest. Of special interest in this case is the pth percentile or quantile of type 1 extreme value distribution in (22.25), and it is given by Eq. (22.34). Naturally then, the best linear unbiased estimator

Table 22.12 Efficiencies of linear unbiased estimators of ξ with linear coefficients and quadratic coefficients (%)

n	2	3	4	5	6	∞
Linear coefficient	84.05	91.18	93.83	95.21	96.07	99.63
Quadratic coefficient	84.05	91.73	94.42	95.79	96.60	99.87

Table 22.13 Efficiencies of linear unbiased estimators of θ with linear coefficients and quadratic coefficients (%)

n	2	3	4	5	6	∞
Linear coefficient	42.70	54.56	60.13	63.37	65.48	75.55
Quadratic coefficient	42.70	58.78	67.14	72.26	75.71	93.64

with polynomial coefficients of the pth percentile of the distribution is given by

$$\xi_{p_*} = \xi_* - \theta_* \log(-\log p), \qquad 0 < p < 1. \tag{22.107}$$

The relative efficiencies of the estimator (22.107) to the Cramér-Rao lower bound were determined by Downton (1966).

9.6 Maximum Likelihood Estimation

Based on a random sample $X_1, X_2, \cdots, X_n$, the maximum likelihood estimators $\hat{\xi}$ and $\hat{\theta}$ satisfy the equations

$$\sum_{i=1}^{n} e^{-(X_i-\hat{\xi})/\hat{\theta}} = n \tag{22.108}$$

and

$$\sum_{i=1}^{n} (X_i - \hat{\xi})\{1 - e^{-(X_i-\hat{\xi})/\hat{\theta}}\} = n\hat{\theta}. \tag{22.109}$$

The asymptotic variances of $\hat{\xi}$ and $\hat{\theta}$ are given by the Cramér-Rao lower bounds in (22.65). The asymptotic correlation coefficient between $\hat{\xi}$ and $\hat{\theta}$ is

$$\left\{1 + \frac{\pi^2}{6(1-\gamma)^2}\right\}^{-1/2} \simeq 0.313 \tag{22.110}$$

Equation (22.108) can be rewritten as

$$\hat{\xi} = -\hat{\theta} \log\left(\frac{1}{n} \sum_{i=1}^{n} e^{-X_i/\hat{\theta}}\right); \tag{22.111}$$

this, when used in Eq. (22.109), yields the following equation for $\hat{\theta}$:

$$\hat{\theta} = \bar{X} - \frac{\sum_{i=1}^{n} X_i e^{-X_i/\hat{\theta}}}{\sum_{i=1}^{n} e^{-X_i/\hat{\theta}}}. \tag{22.112}$$

It is necessary to solve (22.112) by an iterative method for $\hat{\theta}$; Eq. (22.111) then will give $\hat{\xi}$. If $\hat{\theta}$ is large compared to X_i's, then the rhs of (22.112) is approximately

$$\bar{X}\left\{1 - \frac{n-1}{n} \cdot \frac{S^2}{\hat{\theta}\bar{X}}\right\} \tag{22.113}$$

This will provide an approximate solution to (22.112) which can sometimes be used as an initial guess for the iterative method used to solve Eq. (22.112).

The asymptotic confidence interval at significance level α is given by

$$\left(\frac{\hat{\xi}-\xi}{\theta}\right)^2 - 2(1-\gamma)\left(\frac{\hat{\xi}-\xi}{\theta}\right)\left(\frac{\hat{\theta}-\theta}{\theta}\right) + \left\{\frac{\pi^2}{6} + (1-\gamma)^2\right\}\left(\frac{\hat{\theta}-\theta}{\theta}\right)^2 \le -\frac{2}{n}\log\alpha;$$

that is,

$$\left(\frac{\hat{\xi}-\xi}{\theta}\right)^2 - 0.84556\left(\frac{\hat{\xi}-\xi}{\theta}\right)\left(\frac{\hat{\theta}-\theta}{\theta}\right) + 1.82367\left(\frac{\hat{\theta}-\theta}{\theta}\right)^2 \le -\frac{2}{n}\log\alpha.$$

These are ellipses in the (ξ, θ) plane. For the estimator

$$\hat{\xi}_p = \hat{\xi} - \log(-\log p)\hat{\theta}$$

of the pth percentile of the distribution, the asymptotic variance is given by

$$\frac{\theta^2}{n}\left[1 + \frac{6}{\pi^2}\{1 - \gamma - \log(-\log p)\}^2\right].$$

Tiago de Oliveira (1972) has shown that the best asymptotic point predictor of the maximum of (the next) m observations is

$$\hat{\xi} + (\gamma + \log m)\hat{\theta}$$

and its asymptotic variance is

$$\frac{\theta^2}{n}\left[1 + \frac{6}{\pi^2}\{1 + \log m\}^2\right]$$

If the scale parameter 8 is known, then the maximum likelihood estimator of ξ is obtained from (22.108) to be

$$\hat{\xi}_{|\theta} = -\theta\log\left\{\frac{1}{n}\sum_{i=1}^{n} e^{-X_i/\theta}\right\} \tag{22.114}$$

This estimator is not unbiased for ξ. Kimball (1956) has in fact shown that (when 8 is known)

$$E\left[\hat{\xi}_{|\theta}\right] = \xi + \theta\left\{\gamma + \log n - 1 - \frac{1}{2} - \cdots - \frac{1}{n-1}\right\} \qquad (22.115)$$

and

$$\operatorname{var}\left(\hat{\xi}_{|\theta}\right) = \theta^2\left\{\frac{\pi^2}{6} - 1^2 - \frac{1}{2^2} - \cdots - \frac{1}{(n-1)^2}\right\}. \qquad (22.116)$$

While $\hat{\xi}_{|\theta}$ is a biased estimator of $\xi, e^{-\hat{\xi}_{|\theta}/\theta}$ is an unbiased estimator of $e^{-\xi/\theta}$. This is so because $e^{-X/\theta}$ has an exponential distribution with expected value $e^{-\xi/\theta}$. Consequently confidence intervals for this quantity and also for ξ (when 8 is known) can be constructed using methods discussed in Chapter 19, Section 7.

Suppose that the available sample is a doubly Type-II censored sample $X'_{r+1}, X'_{r+2}, \cdots, X'_{n-s}$. Then the log-likelihood function based on this censored sample is

$$\log L = \log n! - \log r! - \log s! - \sum_{i=r+1}^{n-s} Y'_i - \sum_{i=r+1}^{n-s} e^{-Y'_i}$$
$$- (n-r-s)\log\theta + r\log F_Y(Y'_{r+1}) + s\log\{1 - F_Y(Y'_{n-s})\}, \qquad (22.117)$$

where $Y'_i = (X'_i - \xi)/\theta$ are the order statistics from the standard type 1 extreme value distribution with density (22.26) and $F_Y(y)$ its corresponding cdf. From (22.117) we obtain the likelihood equations for ξ and θ to be

$$\frac{\partial \log L}{\partial \xi} = \frac{1}{\theta}\left[(n-r-s) - \sum_{i=r+1}^{n-s} e^{-Y'_i} - r\frac{p_Y(Y'_{r+1})}{F_Y(Y'_{r+1})} + s\frac{p_Y(Y'_{n-s})}{1 - F_Y(Y'_{n-s})}\right]$$
$$= 0 \qquad (22.118)$$

and

$$\frac{\partial \log L}{\partial \theta} = \frac{1}{\theta}\left[\sum_{i=r+1}^{n-s} Y'_i - \sum_{i=r+1}^{n-s} Y'_i e^{-Y'_i} - (n-r-s)\right.$$
$$\left. - rY'_{r+1}\frac{p_Y(Y'_{r+1})}{F_Y(Y'_{r+1})} + sY'_{n-s}\frac{p_Y(Y'_{n-s})}{1 - F_Y(Y'_{n-s})}\right]$$
$$= 0. \qquad (22.119)$$

Harter and Moore (1968a) and Harter (1970) have discussed the numerical solution of the above likelihood equations. The maximum likelihood estimation of ξ, when θ is known, based on right-censored data had been discussed earlier by Harter and Moore (1967). The asymptotic variance-covariance matrix of the maximum likelihood estimates, $\hat{\xi}$ and $\hat{\theta}$, determined from Eqs. (22.118) and (22.119) is given by [Harter (1970, pp. 127–128)]

$$\frac{\theta^2}{n}\begin{bmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{bmatrix}, \tag{22.120}$$

where $((V_{ij}))$ is the inverse of the matrix $((V^{ij}))$ with

$$V^{11} = 1 - q_1 - q_2 + q_1 \log q_1 - (1 - q_2)\log(1 - q_2),$$

$$\begin{aligned} V^{22} = {}& -(1 - q_1 - q_2) - 2\{\Gamma'(1; -\log q_1) - \Gamma'(1; -\log(1 - q_2))\} \\ & -\Gamma''(2; -\log q_1) - \Gamma''(2; -\log(1 - q_2)) + 2\{\Gamma'(2; -\log q_1) \\ & -\Gamma'(2; -\log(1 - q_2))\} - q_1 \log q_1 \log(-\log q_1) \\ & \times\{2 + \log(-\log q_1)\} + (1 - q_2)\log(1 - q_2)\log\{-\log(1 - q_2)\} \\ & \times\left[2 + \log\{-\log(1 - q_2)\} + \log(1 - q_2)\frac{\log\{-\log(1 - q_2)\}}{q_2}\right], \end{aligned}$$

and

$$\begin{aligned} V^{12} &= V^{21} \\ &= -\Gamma'(2; -\log q_1) + \Gamma'(2; -\log(1 - q_2)) + q_1 \log(q_1)\log(-\log q_1) \\ &\quad -(1 - q_2)\log(1 - q_2)\log\{-\log(1 - q_2)\} \\ &\quad -\left(\frac{1}{q_2} - 1\right)\log^2(1 - q_2)\log\{-\log(1 - q_2)\}. \end{aligned}$$

In the equations above $q_1 = r/n$, $q_2 = s/n$, $\Gamma(p; a) = \int_0^a e^{-t} t^{p-1}\,dt$, $\Gamma'(p; a) = (d/du)\Gamma(u; a)|_{u=p}$, and $\Gamma''(p; a) = (d^2/du^2)\Gamma(u; a)|_{u=p}$. Harter (1970), for example, has tabulated the values of V_{11}, V_{12}, and V_{22} for $q_1 = 0.0(0.1)0.9$ and $q_2 = 0.0(0.1)(0.9 - q_1)$.

Phien (1991) has discussed further the maximum likelihood estimation of the parameters ξ and θ based on censored samples. Escobar and Meeker (1986) have discussed the determination of the elements of the Fisher information matrix (V^{ij}'s) based on censored data. Phien carried out an extensive simulation study and observed the following concerning the effects of Type-I censoring on the estimation of parameters and quantiles of the Gumbel distribution using the maximum likelihood method: (1) light censoring on the right may be useful in reducing the bias in estimating the parameters, while left and double censoring are useful for a wider range of censoring levels; (2) the bias in estimating the parameters and quantiles is very small; (3) for complete samples the MLE of ξ overestimates ξ, while the MLE of θ underestimates θ slightly; and (4) censoring introduces an increase in the variances of the estimates.

Phien (1991) has also discussed the maximum likelihood estimation of the parameters based on doubly Type-I censored data. Specifically, for the distribution

$$F_X(x) = e^{-e^{-(x-\xi)/\theta}}$$

with X_l and X_r as the left- and right-censoring time points and with r lowest and s largest observations censored, the likelihood function is proportional to

$$\{F_X(X_l)\}^r \prod_{i=r+1}^{n-s} p_X(X_i)\{1 - F_X(X_r)\}^s.$$

Note in this case that r and s are random variables while X, and X_r are fixed. The log-likelihood function is

$$\log L = \text{const.} - (n - r - s)\log\theta - \sum_{i=r+1}^{n-s}\{Y_i + e^{-Y_i}\} - rd + s\log q,$$

where

$$d = e^{-Y_l},$$

$$q = 1 - e^{-e^{-Y_r}},$$

$$Y_l = \frac{X_l - \xi}{\theta},$$

$$Y_r = \frac{X_r - \xi}{\theta},$$

$$Y_i = \frac{X_i - \xi}{\theta}.$$

The maximum likelihood estimators of ξ and 8 satisfy the equations

$$\frac{\partial \log L}{\partial \theta} = -\frac{G}{\theta} = 0 \quad \text{and} \quad \frac{\partial \log L}{\partial \xi} = -\frac{H}{\theta} = 0,$$

where

$$G = P + P_l + P_r \quad \text{and} \quad H = Q + Q_l + Q_r$$

with

$$P = n - r - s - \sum_{i=r+1}^{n-s} Y_i + \sum_{i=r+1}^{n-s} Y_i e^{-Y_i}, Q = -(n-r-s) + \sum_{i=r+1}^{n-s} e^{-Y_i},$$

$$P_l = r d Y_l,$$

$$Q_l = r d,$$

$$P_r = \frac{s e^{Y_r}(1-q) Y_r}{q},$$

$$Q_r = \frac{s e^{-Y_r}(1-q)}{q}$$

Phien (1991) recommended solving these equations using Newton's procedure.

Posner (1965), when applying the extreme value theory to error-free communication, estimated the parameters ξ and 8 for the complete sample case by the maximum likelihood theory and justified it on the basis of its asymptotic properties. By pointing out that the asymptotic theory need not be valid for Posner's sample size ($n = 30$), Gumbel and Mustafi (1966) showed that in fact a modified method of moments gives better results for Posner's data.

An alternative approach was taken by Balakrishnan and Varadan (1991), who approximated the likelihood equations by using appropriate linear functions and derived approximate maximum likelihood estimators of ξ and 0. They derived these estimators for the type 1 extreme value distribution for the minimum and we present their estimators in the same form for convenience. [The estimators for the type 1 extreme value distribution for the maximum in (22.25) can be obtained simply by interchanging r and s and replacing ξ by $-\xi$ and X_i' by $-X_{n-i+1}'$.] The likelihood equations for ξ and θ in this case are

$$\frac{\partial \log L}{\partial \xi} = -\frac{1}{\theta}\left[r\frac{p_Y(Y_{r+1}')}{F_Y(Y_{r+1}')} - s\frac{p_Y(Y_{n-s}')}{1-F_Y(Y_{n-s}')} + \sum_{i=r+1}^{n-s}\frac{p_Y'(Y_i')}{p_Y(Y_i')}\right]$$
$$= 0 \qquad (22.121)$$

and

$$\frac{\partial \log L}{\partial \theta} = -\frac{1}{\theta}\left[n - r - s + rY'_{r+1}\frac{p_Y(Y'_{r+1})}{F_Y(Y'_{r+1})}\right.$$

$$\left. -sY'_{n-s}\frac{p_Y(Y'_{n-s})}{1 - F_Y(Y'_{n-s})} + \sum_{i=r+1}^{n-s} Y'_i \frac{p'_Y(Y'_i)}{p_Y(Y'_i)}\right] = 0, \qquad (22.122)$$

where $Y'_i = (X'_i - \xi)/\theta$, $p_Y(y) = e^y e^{-e^y}$, and $F_Y(y) = 1 - e^{-e^y}$. Upon expanding the three functions in *(22.121)* and *(22.122)* in a Taylor series around the point $F^{-1}(p_i) = \log(-\log q_i)$ (with $p_i = 1 - q_i = i/(n+1)$), we get the approximate expressions

$$\frac{p_Y(Y'_{r+1})}{F_Y(Y'_{r+1})} \simeq \gamma - \delta Y'_{r+1}, \qquad \frac{p'_Y(Y'_i)}{p_Y(Y'_i)} \simeq \alpha_i - \beta_i Y'_i,$$

$$\frac{p_Y(Y'_{n-s})}{1 - F_Y(Y'_{n-s})} \simeq 1 - \alpha_{n-s} + \beta_{n-s}Y'_{n-s} \qquad (22.123)$$

where

$$\gamma = -\frac{q_{r+1}}{p_{r+1}}\log q_{r+1}\{1 - \log(-\log q_{r+1})\}$$

$$+ \frac{q_{r+1}}{p_{r+1}^2}(\log q_{r+1})^2 \log(-\log q_{r+1}),$$

$$\delta = \frac{q_{r+1}}{p_{r+1}}\log q_{r+1}\left\{1 + \frac{1}{p_{r+1}}\log q_{r+1}\right\},$$

$$\alpha_i = 1 + \log q_i\{1 - \log(-\log q_i)\},$$

$$\beta_i = -\log q_i.$$

By making use of the above approximate expressions in *(22.121)* and *(22.122)* and solving the resulting equations, Balakrishnan and Varadan *(1991)* derived the approximate maximum likelihood estimators of ξ and θ to be

$$\hat{\hat{\xi}} = A - \hat{\hat{\theta}}B \quad \text{and} \quad \hat{\hat{\theta}} = \frac{-C + \sqrt{C^2 + 4AD}}{2(n - r - s)}, \qquad (22.124)$$

where

$$A = \frac{r\,\delta X'_{r+1} + s\beta_{n-s}X'_{n-s} + \sum_{i=r+1}^{n-s}\beta_i X'_i}{m},$$

$$B = \frac{r\gamma - s(1-\alpha_{n-s}) + \sum_{i=r+1}^{n-s}\alpha_i}{m},$$

$$C = r\gamma X'_{r+1} - s(1-\alpha_{n-s})X'_{n-s} + \sum_{i=r+1}^{n-s}\alpha_i X'_i - mAB,$$

$$D = r\,\delta X'^2_{r+1} + s\beta_{n-s}X'^2_{n-s} + \sum_{i=r+1}^{n-s}\beta_i X'^2_i - mA^2,$$

$$m = r\,\delta + s\beta_{n-s} + \sum_{i=r+1}^{n-s}\beta_i.$$

Through a simulation study Balakrishnan and Varadan (1991) have displayed that the above estimators are as efficient as the maximum likelihood estimators, best linear unbiased estimators, and best linear invariant estimators even for samples of size as small as 10. For example, values of bias and mean square error for various estimators of ξ and θ are presented in Table 22.14 for $n = 10$ and 20, $r = 0$, and some choices of s. Estimators of this form have been seen earlier in Chapters 13 and 14.

Estimators of this form based on multiply Type-II censored samples have been discussed by Balakrishnan, Gupta, and Panchapakesan (1992) and Fei, Kong, and Tang (1994).

Table 22.14 Comparison of bias and mean square error of various estimators of ξ and θ for $n = 10$ and 20 and right censoring ($r = 0$)

	$n = 10$				$n = 20$				
$s =$	0	1	2	3	0	1	2	3	4
Bias$(\hat{\hat{\xi}})/\theta$	−0.085	−0.089	−0.103	−0.125	−0.042	−0.042	−0.043	−0.046	−0.049
MSE$(\hat{\hat{\xi}})/\theta^2$	0.123	0.129	0.143	0.171	0.058	0.059	0.061	0.063	0.066
Bias$(\hat{\xi})/\theta$	−0.04	−0.05	−0.08	−0.11	−0.02	—	−0.02	—	−0.04
MSE$(\hat{\xi})/\theta^2$	0.114	0.122	0.137	0.166	0.056	—	0.060	—	0.066
var$(\xi^*)/\theta^2$	0.113	0.120	0.134	0.162	0.056	—	0.059	—	0.065
MSE$(\xi^{**})/\theta^2$	0.113	0.120	0.134	0.161	0.056	—	0.059	—	0.065
Bias$(\hat{\hat{\theta}})/\theta$	−0.066	−0.073	−0.085	−0.100	−0.033	−0.033	−0.035	−0.038	−0.041
MSE$(\hat{\hat{\theta}})/\theta^2$	0.067	0.082	0.098	0.116	0.032	0.036	0.040	0.044	0.048
Bias$(\hat{\theta})/\theta$	−0.07	−0.08	−0.10	−0.12	−0.04	—	−0.04	—	−0.05
MSE$(\hat{\theta})/\theta^2$	0.063	0.077	0.094	0.113	0.033	—	0.042	—	0.050
var$(\theta^*)/\theta^2$	0.072	0.088	0.107	0.132	0.033	—	0.041	—	0.050
MSE$(\theta^{**})/\theta^2$	0.067	0.081	0.097	0.117	0.032	—	0.039	—	0.047

Note: (ξ^*, θ^*) are the best linear unbiased estimators, (ξ^{**}, θ^{**}) are the best linear invariant estimators, $(\hat{\xi}, \hat{\theta})$ are the maximum likelihood estimators, $(\hat{\hat{\xi}}, \hat{\hat{\theta}})$ are the approximate maximum likelihood estimators.

9.7 Conditional Method

The conditional method of inference for location and scale parameters, first suggested by Fisher *(1934)* and discussed in detail by Lawless **(1982)**, has been used effectively for the type *1* extreme value distribution by Lawless *(1973,1978)* and Viveros and Balakrishnan **(1994)**. These developments are described for the type 1 extreme value distribution for minimum with cumulative distribution function $1 - e^{-e^{(x-\xi)/\theta}}$

Suppose that $X_1' \leq X_2' \leq \ldots \leq X_{n-s}'$ is the available Type-II right-censored sample. Then the joint density function of X = $(X_1', X_2', \cdots, X_{n-s}')$ is

$$p_X(x;\xi,\theta) = \frac{n!}{s!\theta^{n-s}} \prod_{i=1}^{n-s} p_Y\left(\frac{x_i' - \xi}{\theta}\right)\left\{1 - F_Y\left(\frac{x_i' - \xi}{\theta}\right)\right\}^s, \quad (22.125)$$

where $F_Y(\cdot)$ and $p_Y(\cdot)$ are the cdf and pdf of the standard form of the type *1* extreme value distribution for minimum given by

$$F_Y(y) = 1 - e^{-e^y} \quad \text{and} \quad p_Y(y) = e^y e^{-e^y}. \qquad (22.126)$$

Then the joint density in *(22.125)* preserves the location-scale structure that may be seen easily by noting from *(22.125)* that the standardized variables, $(X_1' - \xi)/\theta, \ldots, (X_{n-s}' - \xi)/\theta$, have a joint distribution functionally independent of ξ and θ. Suppose that $\hat{\xi}$ and $\hat{\theta}$ are the maximum likelihood estimates of ξ and θ (or some equivariant estimators like BLUEs or BLIEs) which jointly maximize the likelihood of (ξ, θ) that is proportional to *(22.125)*. Then, Z, = $(\hat{\xi} - \xi)/\hat{\theta}$ and $Z_2 = \hat{\theta}/\theta$ are pivotal quantities in the sense that their joint density involves neither ξ nor *8*. With $A_i = (X_i' - \hat{\xi})/\hat{\theta}$ $(i = 1, 2, \cdots, n - s)$, $A = (A,, A,, \cdots, A_{n-s})$ forms an ancillary statistic, and inferences for [and *8* may be based on the joint distribution of Z, and Z_2 conditional on the observed value a of A.

Noting that $(x_i' - \hat{\xi})/\hat{\theta} = a_i z_2 + z_1 z_2$, the joint density of Z_1 and Z_2, conditionally on the observed value a, can be obtained from *(22.125)* as

$$\begin{aligned} p(z_1, z_2|a) &= C(a) z_2^{n-s-1} \prod_{i=1}^{n-s} p_Y(a_i z_2 + z_1 z_2)\{1 - F_Y(a_i z_2 + z_1 z_2)\}^s \\ &= C(a) z_2^{n-s-1} e^{(n-s)z_1 z_2 + s_a z_2} \\ &\quad \times e^{-\sum_{i=1}^{n-s} e^{a_i z_2 + z_1 z_2} - s e^{a_{n-s} z_2 + z_1 z_2}}, \\ &\qquad -\infty < z_1 < \infty, 0 \leq z_2 < \infty, \quad (22.127) \end{aligned}$$

where $C(a)$ is the normalizing constant, and $s_a = \sum_{i=1}^{n-s} a_i$. Using *(22.1271,* Lawless *(1973,1978)* used algebraic manipulations and numerical integration techniques to determine the marginal conditional densities $p(z_1|a)$ and $p(z_2|a)$ that can be utilized to make individual inferences on the parameters ξ and θ.

Conditional inferences for other parameters such as the pth quantile (X_p) of the distribution can also be developed from Eq. (22.127). For example, with the maximum likelihood estimates of X_p being $\hat{X}_p = \hat{\xi} + \hat{\theta} F_Y^{-1}(p)$, one can use the pivotal quantity $Z_3 = (X_p - \hat{\xi})/\hat{\theta} = F_Y^{-1}(p)/Z_2 - Z_1$ to develop inference for X_p. Upon transformation, one may obtain the joint conditional density function of Z_2 and Z_3 from (22.127) from which the marginal conditional density function of Z_3 may be obtained by integration which then will enable one to make inference regarding the pth quantile X_p. Lawless (1973,1978) has noted that tolerance limits, confidence limits for reliability, and prediction intervals can all be similarly handled using the conditional method.

Viveros and Balakrishnan (1994) have developed a similar conditional method of inference based on a Type-II progressively censored data under which scheme one or more surviving items may be removed from the life-test (or, progressively censored) at the time of each failure occurring prior to the termination of the experiment. The familiar complete sample case or the Type-II right-censored sample case (discussed earlier) are special cases of this general scheme.

9.8 Method of Probability-Weighted Moments

Landwehr, Matalas, and Wallis (1979) proposed a simple method of estimation of the parameters ξ and 8 based on the probability-weighted moments

$$M_{(k)} = E\left[X\{1 - F(X)\}^k\right], \qquad k = 0, 1, 2, \dots .$$

An unbiased estimator of $M_{(k)}$ is given by

$$\hat{M}_{(k)} = \frac{1}{n} \sum_{i=1}^{n} X_i' \frac{\binom{n-i}{k}}{\binom{n-1}{k}}, \qquad k = 0, 1, 2, \dots .$$

Then, by making use of the explicit expressions of $M_{(0)}$ and $M_{(1)}$ and equating them to the sample estimators $\hat{M}_{(0)}$ and $\hat{M}_{(1)}$ and solving for the parameters ξ and θ, Landwehr, Matalas, and Wallis (1979) derived the probability-weighted moments estimators to be

$$\hat{8} = \frac{M_{(0)} - 2\hat{M}_{(1)}}{\log 2} \quad \text{and} \quad \hat{\xi} = \hat{M}_{(0)} - \gamma\hat{\theta}.$$

They then compared the performance of these estimators with the moment estimators (Section 9.1) and the maximum likelihood estimators (Section 9.6), in terms of bias and mean square error. Their extensive simulation study indicated that this method of estimation is simple and also highly efficient (in

Table 22.15 Bias, mean square error, and relative efficiency of the moment estimators, PWM estimators, and ML estimators of θ and ξ based on a complete sample of size *n*

		e			ξ		
Method	n	Bias	MSE	Relative Efficiency	Bias	MSE	Relative Efficiency
M	5	0.18	0.37	0.83	−0.10	0.49	0.97
PWM		0.15	0.34	1	−0.08	0.49	1
ML		0.00	0.44	0.74	0.01	0.48	1.05
M	9	0.11	0.30	0.74	−0.06	0.36	0.96
PWM		0.09	0.26	1	−0.04	0.36	1
ML		0.00	0.21	0.76	0.00	0.36	1.03
M	19	0.05	0.22	0.66	−0.03	0.25	0.97
PWM		0.04	0.18	1	−0.02	0.24	1
ML		0.00	0.21	0.76	0.00	0.24	1.02
M	29	0.04	0.18	0.63	−0.02	0.20	0.96
PWM		0.03	0.15	1	−0.01	0.20	1
ML		0.00	0.17	0.77	0.00	0.20	1.00
M	49	0.02	0.14	0.60	−0.01	0.15	0.96
PWM		0.02	0.11	1	0.00	0.15	1
ML		0.00	0.13	0.77	0.00	0.15	1.00
M	99	0.01	0.10	0.57	−0.01	0.11	0.96
PWM		0.01	0.08	1	0.00	0.11	1
ML		0.00	0.09	0.76	0.00	0.11	1.00

terms of efficiency relative to the maximum likelihood estimates). Values of bias, mean square error, and relative efficiency, taken from Landwehr, Matalas, and Wallis (1979), are presented in Table 22.15 for some selected values of n.

These authors also compared (through simulations) the performance of the three methods of estimation of pth quantile (at p = 0.001, 0.01, 0.02, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.98, 0.99, 0.999) based on samples of sizes n = 5, 9, 19, 29, 49, 99, 999.

9.9 "Block-Type" Estimation

Weissman (1978), Huesler and Schuepbach (1986), and Huesler and Tiago de Oliveira (1988), among others, studied the following "block-type" estimation procedure. Suppose that the given observations are X_{ij} for i = 1, 2, ⋯, n and j = 1, 2, ..., k (k may be viewed as the number of years or blocks and n is the number of observations made per year or block). Let $Y_j = \max\{X_{ij}, i \le n\}$. Assume that X_{ij}'s are such that for sufficiently large n, Y_j's have approximately the Gumbel distribution

$$\Pr[Y_j \le y] = e^{-e^{-(y-\xi_n)/\theta}}$$

For the case when θ is known (e.g., $\theta = 1$, without loss of any generality), that is, when X_{ij}'s are i.i.d. with

$$\Pr[X_{ij} \le x] = e^{-e^{-x+\xi}}$$

and

$$\Pr[Y_j \le y] = e^{-e^{-y+\xi+\log n}}$$

where $\xi_n = \xi + \log n$. Huesler and Tiago de Oliveira (1988) estimate ξ from $Y_1, Y_2, \cdots, Y_k$ by using the MLE:

$$\hat{\xi}_A = \hat{\xi}_n - \log n = -\log\left\{\frac{1}{k}\sum_{i=1}^{k} e^{-Y_i}\right\} - \log n$$

with

$$E[\hat{\xi}_A] = \xi + \frac{1}{2k} + \frac{1}{12k^2} + O(k^{-4})$$

and

$$\mathrm{MSE}(\hat{\xi}_A) = \frac{1}{k} + \frac{3}{4k^2} + \frac{1}{4k^3} + O(k^{-4}).$$

Further the distribution of $\sqrt{k}(\hat{\xi}_A - \xi)$ tends to the standard normal as $k \to \infty$.

Weissman (1978) proposes estimation based on the k largest observations of all $N = nk$ values of X_{ij}. Let them be denoted by

$$Z_{1:N} \ge Z_{2:N} \ge \cdots \ge Z_{k:N}.$$

Then Weissman's estimator of ξ is

$$\hat{\xi}_B = -\log\left[\frac{1}{k}\left\{\sum_{i=1}^{k} e^{-Z_{i:N}} + (N-k)\, e^{-Z_{k:N}}\right\}\right]$$

which is just the MLE for the extreme value parameter ξ based on a censored sample. Yet another estimator suggested by Weissman (1978) is

$$\hat{\xi}_C = Z_{k:N} - \log n$$

based on asymptotic properties (as $n \to \infty$). Huesler and Tiago de Oliveira (1988) have noted that all three estimators have the same asymptotic distribution (as $n \to \infty$) and also that $n(\hat{\xi}_B - \hat{\xi}_C) = O_p(1)$.

In the two-parameter case we have correspondingly

$$\hat{\xi}_A = \hat{\xi}_n - \hat{\theta}\log n \quad \text{and} \quad \hat{\theta}_A = \hat{\theta},$$

where $\hat{\xi}_n$ and $\hat{\theta}$ are the MLEs of ξ_n and 8. Here

$$\Pr[X_{ij} \le x] = e^{-e^{-(x-\xi)/\theta}}$$

and

$$\Pr[Y_j \le y] = e^{-e^{-(y-\xi-\theta\log n)/\theta}}$$
$$= e^{-e^{-(y-\xi_n)/\theta}}$$

with $\xi_n = \xi + \theta \log n$. The correlation between $\hat{\xi}_A$ and $\hat{\theta}_A$ approaches -1 as $n \to \infty$.

Estimation based on the k largest values $Z_{1:N} \ge \cdots \ge Z_{k:N}$, for fixed n and k, corresponds to the maximum likelihood estimators $\hat{\xi}_B$ and $\hat{\theta}_B$ based on a left-censored sample, Asymptotically

$$\hat{\theta}_B = \bar{Z}_k - Z_{k:N} + O_p\left(\frac{\log n}{n}\right),$$

$$\hat{\xi}_B = Z_{k:N} - \hat{\theta}_B \log n + O_p\left(\frac{1}{n}\right),$$

where $\bar{Z}_k = (1/k)\sum_{i=1}^{k} Z_{i:n}$. The correlation between $\hat{\xi}_B$ and $\hat{\theta}_B$ also converges slowly to -1 as $n \to \infty$.

Huesler and Tiago de Oliveira (1988) have shown that the Cramér-Rao efficiency of $(\hat{\xi}_B, \hat{\theta}_B)$ with respect to $(\hat{\xi}_A, \hat{\theta}_A)$ defined by

$$\text{eff}(B, A) \equiv \frac{\det(\Sigma_A)}{\det(\Sigma_B)},$$

where Σ_A and Σ_B are the asymptotic variance-covariance matrices of $(\hat{\xi}_A, \hat{\theta}_A)$ and $(\hat{\xi}_B, \hat{\theta}_B)$, respectively, is given by

$$\text{eff}(B, A) \to \frac{6}{\pi^2} = 0.6079 \qquad \text{as } k \to \infty.$$

More delicate comparisons have revealed that method A is not always more efficient than method B. Huesler and Tiago de Oliveira (1988) have presented a data set for which method B has higher efficiency. These authors have concluded that for the estimation of the pth quantile of the annual maximum, method A is better than method B for $p \ge 0.9$ and that method A is definitely superior when $k \ge 15$.

9.10 A Survey of Other Developments

The details on various methods of inference presented in the last nine subsections are by no means complete. Numerous other papers have appeared dealing with different aspects of inference relating to the extreme value distribution such as proposing new and simplified methods, making finer improvements over the existing methods, dealing with numerical algorithms for the estimation, discussing accelerated life-tests and extreme value regression, and so on. These results are equally important, and they are listed and described briefly below.

Engelhardt (1975) and Engelhardt and Bain (1977) have provided further discussions on the simplified estimators of the parameters and associated inferential procedures. Singh (1975) has discussed admissibility of some estimators. Meeker and Nelson (1975) have proposed and examined optimum accelerated life-tests; see also Nelson and Meeker (1978). Lawless and Mann (1976) considered tests for the homogeneity of scale parameters (θ_i's) in k samples from extreme value populations. While Smith (1977) discussed the interval estimation of parameters, Durrant (1978) constructed a nomogram for confidence limits on quantiles of the normal distribution and discussed its usefulness for the extreme value distribution. A preliminary test of significance was considered by Tsujitani, Ohta, and Kase (1979). Ashour and El-Adl (1980) examined the Bayesian estimation of the parameters.

Cheng and Iles (1983,1988) discussed confidence bands for the cumulative distribution functions. Schuepbach and Huesler (1983) proposed some simple estimators for the parameters ξ and 8 based on censored samples. Bootstrap confidence intervals for the parameters when the available sample is progressively censored have been discussed by Robinson (1983). Some graphical methods of estimating the parameters were put forward by Stone and Rosen (1984). Keating (1984) has commented on the estimation of percentiles and the reliability function. In an interesting article Smith and Weissman (1985) discussed the maximum likelihood estimation of the lower tail of the distribution. A comparison of confidence intervals derived by different methods was carried out by Chao and Hwang (1986). Welsh's (1986) discussion on the use of the empirical distribution and characteristic function to estimate the parameters includes the extreme value distribution as one of the cases. Singh (1987) estimated the parameters of the type 1 extreme value distribution from the joint distribution of m extremes. A weighted least-squares method of estimation was considered by Öztürk (1987). While Schneider and Weissfeld (1989) discussed the interval estimation of parameters based on censored data, Ahmed (1989) considered the problem of selecting the extreme value population with the smallest ξ_i.

Achcar (1991) presented another reparametrization for the extreme value distribution. Hooda, Singh, and Singh (1991) discussed the estimation of the Gumbel distribution parameters from doubly censored samples. Comparisons of approximate confidence intervals for the extreme value simple linear regression model under time censoring (or Type-I censoring) were made by Doganoksoy and Schmee (1991). Abdelhafez and Thomas (1991) discussed bootstrap confidence bands for the extreme value regression models with randomly censored data.

10 TOLERANCE LIMITS AND INTERVALS

Based on a complete sample or Type-I1 censored sample observed from the distribution, the lower a tolerance limit for proportion $1-\gamma$ is $\hat{\xi}+k_L\hat{\theta}$ satisfying the equation

$$\Pr\left[\Pr\left[X \geq \hat{\xi}+k_L\hat{\theta}\right] \geq 1-\gamma\right]=\alpha; \qquad (22.128)$$

similarly the upper a tolerance limit for proportion $1-\gamma$ is $\hat{\xi}+k_U\hat{\theta}$ satisfying the equation

$$\Pr\left[\Pr\left[X \leq \hat{\xi}+k_U\hat{\theta}\right] \geq 1-\gamma\right]=\alpha. \qquad (22.129)$$

The constants k_L and k, are referred to as the lower and upper tolerance factors, respectively.

In the case of the type 1 extreme value distribution for the minima with cumulative distribution function

$$F_X(x)=1-e^{-e^{(x-\xi)/\theta}},$$

equations *(22.128)* and *(22.129)* become

$$\Pr\left[\frac{\hat{\xi}-\xi}{\theta}+k_L\frac{\hat{\theta}}{\theta} \leq \log[-\log(1-\gamma)]\right]=\alpha \qquad (22.130)$$

and

$$\Pr\left[\frac{\hat{\xi}-\xi}{\theta}+k_U\frac{\hat{\theta}}{\theta} \geq \log(-\log\gamma)\right]=\alpha, \qquad (22.131)$$

respectively. Upon rewriting Eqs. *(22.130)* and *(22.131)* as

$$\Pr\left[\frac{\theta}{\hat{\theta}}\log[-\log(1-\gamma)]-\frac{\hat{\xi}-\xi}{\hat{\theta}} \geq k_L\right]=\alpha \qquad (22.132)$$

and

$$\Pr\left[\frac{\theta}{\hat{\theta}}\log(-\log\gamma)-\frac{\hat{\xi}-\xi}{\hat{\theta}} \leq k_U\right]=\alpha, \qquad (22.133)$$

we observe that k_L and k, are the upper and lower $100\alpha\%$ points of the distributions of the pivotal quantities

$$P_1 = \frac{\theta}{\hat{\theta}} \log[-\log(1-\gamma)] - \frac{\hat{\xi} - \xi}{\hat{\theta}},$$

$$P_2 = \frac{\theta}{\hat{\theta}} \log(-\log \gamma) - \frac{\hat{\xi} - \xi}{\hat{\theta}}, \qquad (22.134)$$

respectively. As the distributions of these two pivotal quantities are not derivable explicitly, their percentage points need to be determined either through Monte Carlo simulations or by some approximations.

Mann and Fertig (1973) used the best linear invariant estimators to prepare tables of tolerance factors for Type-II right-censored samples when $n = 3(1)25$ and $n - s = 3(1)n$, where s is the number of largest observations censored in the sample; see also Mann, Schafer, and Singpurwalla (1974). While Thoman, Bain, and Antle (1970) presented tables that can be used to determine tolerance bounds for complete samples up to size n = 120, Billman, Antle, and Bain (1972) gave tables which can be used to determine tolerance bounds for samples of sizes $n = 40(20)120$ with 50% or 75% of the largest observations censored. Johns and Lieberman (1966) presented extensive tables that can be used to get tolerance bounds for sample sizes $n = 10$, 15, 20, 30, 50, and 100 with Type-II right censoring at four values of s (number of observations censored) for each n. By making use of the efficient simplified linear estimator presented by Bain (1972) (described in Section 9.2), Mann, Schafer, and Singpurwalla (1974) derived approximate tolerance bounds based on a F-approximation. This F-approximation is quite effective and hence deserves a special mention.

By using Bain's simplified linear estimators $\hat{\hat{\theta}}$ of θ and the related linear estimator $\hat{\hat{\xi}}$ of ξ, Mann, Schafer, and Singpurwalla (1974, p. 249) have shown that an approximate lower $100\alpha\%$ confidence bound on the quantile X_γ is

$$\hat{\hat{\xi}} + \hat{\hat{\theta}} \left\{ -\frac{C_{n-s,n}}{l_{n-s,n}}(1 - F_{1-\alpha}) + F_{1-\alpha} \log(-\log \gamma) \right\}, \qquad (22.135)$$

where $B_{n-s,n}$, $C_{n-s,n}$ and $l_{n-s,}$ are constants depending on s and n, and $F,_{-\alpha}$ is the upper $1 - \alpha$ percentage point of an F distribution with degrees of freedom

$$d_1 = \frac{2\{\log(-\log \gamma) + C_{n-s,n}/l_{n-s,n}\}^2}{\{B_{n-s,n} - C_{n-s,n}^2/l_{n-s,n}\}},$$

$$d_2 = \frac{2}{l_{n-s.n}}. \qquad (22.136)$$

As demonstrated by Mann, Schafer, and Singpurwalla (1974, p. 250), this F-approximation can also be used with the best linear unbiased estimators ξ^* and θ^* (see Section 9.3); in fact they have indicated that the approximation turns out to be good even in the case of moderate sample sizes with heavy censoring. Values of the constants $B_{n-s,n}$, $C_{n-s,n}$, and $l_{n-s,n}$, which depend on the means, variances, and covariances of order statistics from the type 1 extreme value distribution for the minimum, have been tabulated by Mann, Schafer, and Singpurwalla (1974) for some choices of n and s.

An alternate F-approximation was proposed by Lawless (1975) for the lower a confidence bound on the quntile X_γ. His approximation is based on the fact that, at least when the censoring in the sample is fairly heavy, the estimators $\hat{\hat{\xi}}$ and $\hat{\hat{\theta}}$ are almost the same as the maximum likelihood estimators $\hat{\xi}$ and $\hat{\theta}$; specifically, we have

$$\frac{\hat{\hat{\theta}}}{1 + l_{n-s,n}} \simeq \hat{\theta} \quad \text{and} \quad \hat{\hat{\xi}} - \frac{C_{n-s,n}}{1 + l_{n-s,n}}\hat{\hat{\theta}} \simeq \hat{\xi}. \tag{22.137}$$

These are exactly the same linear transformations, described in Section 9.3, that transform the BLUEs to BLIEs. Using Eq. (22.137) and the maximum likelihood estimates $\hat{\xi}$ and $\hat{\theta}$, Lawless (1975) derived an approximate lower α confidence bound on the quantile X_γ as

$$\hat{\xi} + \hat{\theta}\left\{C_{n-s,n} + (1 + l_{n-s,n})\left[-\frac{C_{n-s,n}}{l_{n-s,n}}(1 - F_{1-\alpha}) + F_{1-\alpha}\log(-\log\gamma)\right]\right\}. \tag{22.138}$$

This F-approximation is quite accurate over a wide range of situations.

Lawless noted that the quantity

$$Z_\gamma = \frac{1}{\hat{\theta}}\left\{\theta\log(-\log\gamma) - (\hat{\xi} - \xi)\right\} \tag{22.139}$$

is a pivotal quantity, since $Z_\gamma = \{\log(-\log\gamma)/Z_2\} - Z_1$, where $Z_1 = (\hat{\xi} - \xi)/\hat{\theta}$ and $Z_2 = \hat{\theta}/\theta$ are pivotal quantities, discussed earlier, that can be used to construct tolerance bounds. For example,

$$\Pr\left[Z_\gamma \ge z_{\gamma,\alpha}\right] = \alpha \Rightarrow \Pr\left[z_{\gamma,\alpha}\hat{\theta} + \hat{\xi} \le X_\gamma\right] = \alpha, \tag{22.140}$$

Table 22.16 Comparison of exact and Fapproximation tolerance bounds

		y = 0.95		y = 0.90	
n	$n-s$	$z_{\gamma,0.95}$	F-Approximation	$z_{\gamma,0.95}$	F-Approximation
60	54	−3.76	−3.79	−2.88	−2.91
60	42	−3.85	−3.88	−2.93	−2.96
60	30	−3.99	−4.01	−3.00	−3.03
60	18	−4.19	−4.23	−3.08	−3.13
60	6	−4.69	−4.83	−3.09	−3.38
40	36	−4.01	−4.02	−3.09	−3.06
40	28	−4.12	−4.16	−3.13	−3.17
40	20	−4.34	−4.35	−3.26	−3.28
40	12	−4.68	−4.72	−3.40	−3.46
40	8	−5.02	−5.11	−3.49	−3.60
40	4	−5.96	−5.99	−3.53	−3.74
25	20	−4.50	−4.52	−3.44	−3.47
25	10	−5.22	−5.28	−3.83	−3.89
25	5	−6.54	−6.62	−4.33	−4.47

and hence $z_{\gamma,\alpha}\hat{\theta} + \hat{\xi}$ becomes a lower a confidence bound on X_γ. The percentage points of the distribution of Z_γ in (22.139) therefore yield upper tolerance limits [see the pivotal quantity P_2 in Eq. (22.134)]. In Table 22.16, taken from Lawless (1975), a comparison of the exact tolerance bounds determined from the distribution of Z_γ in (22.139) with $\alpha = 0.95$ and the corresponding F-approximations are presented.

Mann and Fertig (1977) discussed the correction for small-sample bias in Hassanein's (1972) asymptotic best linear unbiased estimators of ξ and θ based on k optimally selected quantiles (see Section 9.4). They presented tables of these bias-correction factors for complete samples of sizes n = 20(1)40. These tables will not only allow one to obtain estimates based on the specified sets of order statistics that are best linear unbiased estimates or best linear invariant estimates, but can also be utilized to determine approximate confidence bounds on X_γ and the related tolerance limits using approximation ideas discussed above.

Through the conditional method of inference discussed in detail in Section 9.7, Lawless (1975) has shown that the conditional tail probability of the distribution of Z_γ in (22.139) is given by

$$\Pr\left[Z_\gamma \geq z \,|\, \boldsymbol{a}\right] = (n-s-1)!\,C_{n-s}(\boldsymbol{a})$$

$$\times \int_0^\infty \frac{t^{n-s-2}\, e^{t\sum_{i=1}^{n-s} a_i}\, \Gamma_{h(t,z)}(n-s)}{\Gamma(n-s)\left\{\sum_{i=1}^{n-s} e^{a_i t} + s\, e^{a_{n-s} t}\right\}^{n-s}}\, dt, \qquad (22.141)$$

where $\Gamma_b(p)$ is the incomplete gamma function

$$\Gamma_b(p) = \int_0^b e^{-x} x^{p-1}\, dt, \qquad 0 \le b < \infty,$$

$$h(t, z) = -\log \gamma \cdot e^{-tz} \left\{ \sum_{i=1}^{n-s} e^{a_i t} + s e^{a_{n-s} t} \right\}. \tag{22.142}$$

The integral in (22.141) needs to be evaluated numerically. The normalizing constant $C_{n-s}(\boldsymbol{a})$ is determined numerically by using the condition that $\Pr[Z_\gamma \ge -\infty | \boldsymbol{a}] = 1$ in which case $h(t, z) = \infty$ and consequently $\Gamma_{h(t,z)}(n - s) = \Gamma_\infty(n - s) = \Gamma(n - s)$. Thus we get

$$C_{n-s}(\boldsymbol{a}) = \left[(n - s - 1)! \int_0^\infty \frac{t^{n-s-2} e^{t \sum_{i=1}^{n-s} a_i}}{\left\{ \sum_{i=1}^{n-s} e^{a_i t} + s e^{a_{n-s} t} \right\}^{n-s}} dt \right]^{-1}. \tag{22.143}$$

Once the percentage points of Z_γ are determined from (22.141) by numerical methods, tolerance limits can be obtained as explained earlier.

Gerisch, Struck, and Wilke (1991) took a completely different direction and discussed the determination of one-sided tolerance limit factors for the exact extreme value distributions from a normal parent distribution. They justified the need for these factors based on the grounds that one-sided tolerance limits for the asymptotic extreme value distributions cannot be regarded as sufficient approximations of one-sided tolerance limits for the corresponding exact extreme value distributions.

11 PREDICTION LIMITS AND INTERVALS

Suppose that $\hat{\xi}$ and $\hat{\theta}$ are the maximum likelihood estimators of ξ and θ based on a sample of size n from the type 1 extreme value distribution for the maximum (discussed in Section 9.6). Suppose that Z is an independent observation, to be made from the same distribution. Then, as Antle and Rademaker (1972) showed, the construction of prediction intervals for Z is based on the pivotal quantity

$$T_1 = \frac{Z - \hat{\xi}}{\hat{\theta}}. \tag{22.144}$$

Antle and Rademaker presented a table of percentage points, $t_{1,\gamma}$, of the distribution of T_1 for selected values of n and γ, and they determined these values, appearing in Table 22.17, by Monte Carlo simulations.

The irregular progression of values in Table 22.17 (especially for $n = 100$) is

Table 22.17 Percentage points $t_{1,\gamma}$ of the distribution of T_1 in (22.144)

n	γ 0.90	0.95	0.975	0.98	0.99	0.995
10	2.64	3.59	4.51	4.88	6.00	
20	2.41	3.24	4.04	4.26	5.12	6.18
30	2.33	3.06	3.89	4.14	4.90	5.86
40	2.30	3.00	3.81	4.04	4.79	5.68
50	2.29	2.98	3.79	3.99	4.69	5.56
60	2.26	2.97	3.74	3.99	4.70	5.53
70	2.26	2.98	3.72	3.94	4.66	5.46
100	2.24	2.96	3.66	3.90	4.68	5.38
∞	2.25	2.97	3.68	3.90	4.60	5.30

presumably due to sampling variation. Using the values of $t_{1,\gamma}$ presented above, an upper $100\gamma\%$ prediction limit for Z can be determined as $\hat{\xi} + \hat{\theta}t_{1,\gamma}$.

Engelhardt and Bain (1979), on the other hand, used their simplified linear estimators of ξ and θ described in Section 22.9.2 to construct prediction intervals for Z_1' in a future sample of size m from the type 1 extreme value distribution for the minimum, based on a Type-II right-censored sample of size $n - s$. With $\hat{\hat{\xi}}$ and $\hat{\hat{\theta}}$ denoting the simplified linear estimators of ξ and θ based on the right-censored sample of size $n - s$, Engelhardt and Bain (1979) considered the pivotal quantity

$$T_2 = \frac{\hat{\hat{\xi}} - Z_1'}{\hat{\hat{\theta}}}. \tag{22.145}$$

With $t_{2,\gamma}$ being the γth quantile of the distribution of T_2, it is readily seen that $\hat{\hat{\xi}} - t_{2,\gamma}\hat{\hat{\theta}}$ becomes a lower $100\gamma\%$ prediction limit for Z_1'. Engelhardt and Bain also developed an efficient approximation for $t_{2,\gamma}$ as follows. Upon writing

$$\Pr[T_2 < t] = \Pr\left[W(t) < \frac{Z_1' - \xi}{\theta}\right] \tag{22.146}$$

where $W(t) = (\hat{\hat{\xi}} - \xi)/\theta - t\hat{\hat{\theta}}/\theta$, they used the approximation [see Engelhardt and Bain (1977)]

$$W(t) \to \log\left\{\frac{k\chi^2(l)}{l}\right\}, \tag{22.147}$$

where $k(t)$ and $l(t)$ are chosen so that both sides of (22.147) have the same mean and variance. With

$$\nu = \operatorname{var}(W(t)) = \operatorname{var}\left(\frac{\hat{\hat{\xi}} - \xi}{\theta}\right) + 2t^2 \operatorname{var}\left(\frac{\hat{\hat{\theta}}}{\theta}\right) - 2t \operatorname{cov}\left(\frac{\hat{\hat{\xi}} - \xi}{\theta}, \frac{\hat{\hat{\theta}}}{\theta}\right),$$

they derived convenient approximations for l and k as

$$l = (8\nu + 12)/(\nu^2 + 6\nu) \qquad \text{(not depending on } t) \tag{22.148}$$

and

$$k = \exp\left\{-t + \frac{15l^2 + 5l + 6}{15l^3 + 6l}\right\} \tag{22.149}$$

Since $(Z_1' - \xi)/\theta \to \log(\chi^2(2)/2m)$ independently of $W(t)$, we get the approximation [using (22.146) and (22.147)]

$$\Pr(T_2 < t) \doteq \Pr[mk < F(2, l)] \tag{22.150}$$

where $F(2, l)$ denotes a central F-distribution with $(2, l)$ degrees of freedom; (22.150), when used with the exact expression $F_{1-\gamma}(2, l) = (l/2)(\gamma^{-2/l} - 1)$, yields a simple approximation for $t_{2,\gamma}$ as the value of t such that

$$\gamma = \left(1 + \frac{2mk}{l}\right)^{-l/2}. \tag{22.151}$$

Prediction intervals/limits for Z_j' $(2 \le j \le m)$ have been developed by Engelhardt and Bain (1979), based on the pivotal quantity

$$T_3 = \frac{\hat{\hat{\xi}} - Z_j'}{\hat{\hat{\theta}}}. \tag{22.152}$$

Since

$$\exp\left(\frac{Z_j' - \xi}{\theta}\right) \to \sum_{i=1}^{j} \frac{\chi_i^2(2)}{2(m - i + 1)} \tag{22.153}$$

(independently of $W(t)$), where $\chi_1^2(2), \dots, \chi_j^2(2)$ are independent, the distribution of the linear combination of chi-square variables can be closely approximated by the form (see Patnaik's approximation in Chapter 18)

$$\exp\left(\frac{Z_j' - \xi}{\theta}\right) \xrightarrow{\text{approx.}} \frac{c\chi^2(\nu)}{\nu}, \tag{22.154}$$

where

$$c = \sum_{i=1}^{j} \frac{1}{m-i+1} \simeq \log(m+0.5) - \log(m-j+0.5),$$

$$\nu = 2\frac{\{\sum_{i=1}^{j} 1/(m-i+1)\}^2}{\sum_{i=1}^{j} 1/(m-i+1)^2}$$

$$\simeq \frac{\{\log(m+0.5) - \log(m-j+0.5)\}^2}{\dfrac{1}{m-j+0.5} - \dfrac{1}{m+0.5}}$$

Then a lower $100\gamma\%$ prediction limit for Z'_j is given by $\hat{\xi} - t_{3,\gamma}\hat{\hat{\theta}}$, where $t_{3,\gamma}$ is approximated by the value t such that $F_{1-\gamma}(\nu, l) = k/c$.

Fertig, Meyer, and Mann *(1980)* used the best linear invariant estimates ξ^{**} and θ^{**}, described in Section *22.9.3,* for the prediction of Z'_1 in a future sample of size m based on the pivotal quantity

$$S_1 = \frac{\xi^{**} - Z'_1}{\theta^{**}}. \tag{22.155}$$

With $s_{1,\gamma}$ denoting the 100γth percentile of the distribution of S_1 in (22.155), the $100\gamma\%$ lower prediction bound for Z'_1 is given by $\xi^{**} - s_{1,\gamma}\theta^{**}$. *A* $100\gamma\%$ upper prediction bound for Z'_1 may also be obtained by replacing $s_{1,\gamma}$ by $s_{1,1-\gamma}$. Using Monte Carlo simulations, Fertig, Meyer, and Mann (1980) determined the values of $s_{1,\gamma}$ for different choices of n, n $- s$, and γ when m = *1.* Selected values from their table are presented in Table *22.18.*

Mann, Schafer, and Singpurwalla *(1974)* suggested an F-approximation for the distribution of the statistic S, in *(22.155)* to be used only for large future sample sizes and moderate levels of confidence. Mann (1976) discussed conditions under which this approximation is sufficiently precise for use. Mann and Saunders *(1969)* presented solutions to two special cases when the given samples are of sizes two and three. Fertig, Meyer, and Mann (1980), by employing a procedure that is an extension of the one used by Fertig and Mann *(1978)* to approximate the distribution of the studentized extreme value statistics, have suggested an alternate F-approximation and examined its accuracy.

Engelhardt and Bain *(1982)* provided further discussions on the prediction problem and derived in particular two simpler approximations for the percentage points of the statistic T_2 in *(22.145).* The approximation for t,,, from Eq. *(22.151)* needs to be determined by numerical iterative methods. For this reason Engelhardt and Bain *(1982)* presented the following two simpler

Table 22.18 Distribution percentiles of $S_s = (\xi^{**} - Z_1')/\theta^{**}$ for Type-II right-censored samples of size $n - s$ from a sample of size n

n	$n-s$	γ 0.02	0.05	0.10	0.25	0.40	0.50	0.60	0.75	0.90	0.95	0.98
5	3	−9.67	−5.20	−3.04	−0.97	−0.09	0.37	0.85	1.85	4.37	6.74	11.68
	5	−2.68	−1.77	−1.17	−0.39	0.14	0.51	0.92	1.74	3.39	4.78	6.99
10	3	−15.94	−8.87	−5.27	−1.91	−0.56	0.02	0.55	1.43	3.22	5.25	8.89
	5	−4.41	−2.91	−1.88	−0.68	−0.04	0.35	0.74	1.52	3.03	4.30	6.38
	10	−1.76	−1.32	−0.96	−0.36	0.12	0.43	0.79	1.48	2.69	3.68	5.02
15	3	−21.17	−11.36	−6.80	−2.60	−0.97	−0.26	0.33	1.20	2.81	4.43	7.59
	5	−5.72	−3.62	−2.36	−0.92	−0.17	0.24	0.64	1.35	2.76	4.00	5.96
	10	−2.16	−1.56	−1.10	−0.41	0.06	0.39	0.73	1.39	2.63	3.59	5.00
	15	−1.62	−1.25	−0.92	−0.34	0.10	0.42	0.73	1.37	2.52	3.41	4.53

approximations:

$$\Pr[T_2 < t] \simeq e^{-m e^{-t+\nu(t)/(2g)}}, \tag{22.156}$$

where

$$g = 1 + \frac{5 + \frac{1}{2}\log m}{n - s},$$

$$\Pr[T_2 < t] \simeq e^{-m e^{-t}}. \tag{22.157}$$

In fact the approximations (22.150) and (22.156) both converge to (22.157) as $n \to \infty$ with $(n-s)/n \to p > 0$. The advantage of the approximations (22.156) and (22.157) is that they can be solved explicitly for quantiles $t_{2,\gamma}$. For example, upon equating the right-hand side of (22.156) to γ and solving the resulting quadratic equation

$$-t + \frac{\nu(t)}{2g} = \log\left(-\frac{1}{m}\log\gamma\right),$$

we obtain an explicit approximation for $t_{2,\gamma}$ as

$$t_{2,\gamma} \simeq (A+B) - \left\{(A+B)^2 - C + 2A\log\left(-\frac{1}{m}\log\gamma\right)\right\}^{1/2}, \tag{22.158}$$

where

$$A = \frac{g}{\operatorname{var}\left(\hat{\hat{\theta}}/\theta\right)},$$

$$B = \frac{\operatorname{cov}\left(\hat{\hat{\xi}}/\theta, \hat{\hat{\theta}}/\theta\right)}{\operatorname{var}\left(\hat{\hat{\theta}}/\theta\right)},$$

$$C = \frac{\operatorname{var}\left(\hat{\hat{\xi}}/\theta\right)}{\operatorname{var}\left(\hat{\hat{\theta}}/\theta\right)}.$$

The limiting approximation (22.157) also readily yields an explicit approximation for $t_{2,\gamma}$ as

$$t_{2,\gamma} \simeq -\log\left(-\frac{1}{m}\log\gamma\right) \tag{22.159}$$

Engelhardt and Bain (1982) have examined the accuracy of all these approximations.

Pandey and Upadhyay (1986) discussed approximate prediction limits for the Weibull distribution, which may be transformed to the type 1 extreme value model for the minimum through the usual logarithmic transformation, based on preliminary test estimator. Abdelhafez and Thomas (1990) derived approximate prediction limits for the Weibull and extreme value regression models.

12 OUTLIERS AND ROBUSTNESS

For the type 1 extreme value distribution for minimum, Mann (1982) proposed three statistics to test for k upper outliers in the sample. These three test statistics are given by

$$V = \frac{\theta_n^{**}}{\theta_{n-k}^{**}}, \tag{22.160}$$

$$Q = \frac{X_n' - X_{n-k}'}{\theta_{n-k}^{**}}, \tag{22.161}$$

and

$$W = \frac{X_{n-k+1}' - X_{n-k}'}{\theta_{n-k}^{**}} \tag{22.162}$$

where θ_n^{**} and θ_{n-k}^{**}, are the best linear invariant estimators of **8** (see Section ***9.3*** for details) based on the complete sample of size n and on the smallest n − k order statistics, respectively. Since the exact null distributions of these test statistics are intractable, Mann (1982) determined the critical values by Monte Carlo simulations and presented some tables. Further, through an empirical power study, Mann demonstrated that her statistic W in (22.162) provides a powerful test in detecting upper outliers from a labeled slippage location-shift model.

Fung and Paul (1985) carried out an extensive empirical study to examine the performance of several outlier detection procedures. In addition to the preceding three test statistics, these authors also considered the following five test statistics:

$$G = S_{n-k}^2 / S_n^2, \tag{22.163}$$

where S_{n-k}^2 is the sum of squared deviations of the smallest n − k order statistics and S_n^2 is the sum of squared deviations of all n observations,

$$R_1 = \frac{X'_n - X'_{n-k}}{X'_n - X'_1}, \tag{22.164}$$

$$R_2 = \frac{X'_n - X'_{n-k}}{X'_n - X'_2}, \tag{22.165}$$

$$R_3 = \frac{X'_n - X'_{n-k}}{X'_n - X'_3}, \tag{22.166}$$

and

$$L = \frac{\sum_{i=n-k-1}^{n-1} (X'_{i+1} - X'_i)/(E[Y'_{i+1}] - E[Y'_i])}{\sum_{i=1}^{n-1} (X'_{i+1} - X'_i)/(E[Y'_{i+1}] - E[Y'_i])}. \tag{22.167}$$

Fung and Paul (1985) have also considered the counterparts of these five tests (obtained by changing X'_i to X'_{n-i+1}) for testing for k lower outliers in the sample. Fung and Paul have also presented critical values for all these tests determined through Monte Carlo simulations. They then compared the performance of the test statistics in terms of their sizes and powers in detecting k = 1, 2, and 3 outliers. For the upper outliers they used all eight test statistics, while only the last five test statistics were used for the lower outliers.

In their empirical power study Fung and Paul (1985) used a contamination outlier model with location shift as well as with location and scale shift [instead of the labeled slippage model considered by Mann (1982)]. Under this contamination outlier model, Mann's W test performed very poorly as compared to the other tests. The test procedure based on the L statistic in

(22.167) performed the best for the contamination model, while the counterpart of G in (22.163) performed well for testing the lower outliers. The test statistic R, in (22.164) and its counterpart also provide useful tests in general.

By using joint distributions of order statistics, Paul and Fung (1986) presented explicit formulas for the calculation of critical values of the test statistics R_1, R_2, and R_3 (and their counterparts for testing for lower outliers) for testing for $k = 1$ and 2 outliers.

13 PROBABILITY PLOTS, MODIFICATIONS, AND MODEL VALIDITY

Due to the prominence and significance of the extreme value distributions, considerable work has been done with regard to testing whether an extreme value distribution is appropriate for the data at hand. In this section a brief description of these investigations is presented. The book by D'Agostino and Stephens (1986) provides an elaborate account of various goodness-of-fit tests developed for the extreme value distributions.

One of the easiest goodness-of-fit tests is the "correlation coefficient" test for the type 1 extreme value distribution. This test is based on the product-moment correlation between the sample order statistics and their expected values. Since $E[X_i'] = \xi + \theta E[Y_i']$, one may as well use the correlation between the sample order statistics X_i' and the expected values of standard order statistics $E[Y_i']$ for the type 1 extreme value distribution. Naturally large values (close to 1) of this correlation will support the assumption of the type 1 extreme value distribution for the data at hand. Smith and Bain (1976) discussed this test and presented tables of critical points; tables were also provided by these authors for the case when the available sample is Type-II censored. A more extensive table of points for $n(1 - R^2)$, where R is the sample correlation coefficient, has been provided by Stephens (1986). Stephens's choice of the statistic $n(1 - R^2)$ permits easy interpolation in the tables. Further his tables also facilitate the test even in case of doubly Type-II censored samples. Kinnison (1989) discussed the same correlation test for the type 1 extreme value distribution and presented tables of smoothed values of the percentage points of r (in the case of complete samples) when $n = 5(5)30(10)100, 200$. Kinnison used the approximation

$$E[Y_i'] \doteq -\log\{-\log(i/(n+1))\}$$

in the plot and the resulting calculation of the correlation coefficient. As pointed out by Lockhart and Spinelli (1990), use of the exact values of $E[Y_i']$ or even Blom's approximation $E[Y_i'] \doteq -\log\{-\log[(i - 0.25)/(n + 0.25)]\}$ may result in an increase in the power of the test. However, as aptly mentioned by Lockhart and Spinelli, even though the correlation test is simple to use and has an intuitive appeal, its power properties are undesirable. As a matter of fact McLaren and Lockhart (1987) have shown that the correlation test has asymptotic efficiency equal to 0 relative to standard tests such as Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests.

Stephens (1977) presented goodness-of-fit tests based on empirical distribution function statistics W^2, U^2, and A^2 given by

$$W^2 = \sum_i \left\{ F_X(X_i') - \frac{2i-1}{2n} \right\}^2 + \frac{1}{12n}, \tag{22.168}$$

$$U^2 = W^2 - n\left\{ \frac{1}{n} \sum_i F_X(X_i') - \frac{1}{2} \right\}^2, \tag{22.169}$$

and

$$A^2 = -\frac{1}{n} \sum_i (2i-1)\left[\log F_X(X_i') + \log\{1 - F_X(X_{n-i+1}')\}\right] - n. \tag{22.170}$$

Stephens discussed the asymptotic percentage points of these three statistics for the three cases when one or both of the parameters ξ and θ need to be estimated from the data (using the MLEs). Stephens (1977) also suggested slight modifications of these statistics in order to enable the usage of the asymptotic percentage points in case of small sample sizes; these are presented in Table 22.19.

Table 22.19 Percentage points for modified statistics W^2, U^2, and A^2

Statistic	Case[a]	Modification	Upper Tail Percentage Points, α 0.75	0.90	0.95	0.975	0.99
W^2	0	$(W^2 - 0.4/n + 0.6/n^2)(1.0 + 1.0/n)$	—	0.347	0.461	0.581	0.743
	1	$W^2(1 + 0.16/n)$	0.116	0.175	0.222	0.271	0.338
	2	None	0.186	0.320	0.431	0.547	0.705
	3	$W^2(1 + 0.2/\sqrt{n})$	0.073	0.102	0.124	0.146	0.175
U^2	0	$(U^2 - 0.1/n + 0.1/n^2)(1.0 + 0.8/n)$	—	0.152	0.187	0.221	0.267
	1	$U^2(1 + 0.16/n)$	0.090	0.129	0.159	0.189	0.230
	2	$U^2(1 + 0.15/\sqrt{2n})$	0.086	0.123	0.152	0.181	0.220
	3	$U^2(1 + 0.2/\sqrt{n})$	0.070	0.097	0.117	0.138	0.165
A^2	0	None	—	1.933	2.492	3.070	3.857
	1	$A^2(1 + 0.3/n)$	0.736	1.062	1.321	1.591	1.959
	2	None	1.060	1.725	2.277	2.854	3.640
	3	$A^2(1 + 0.2/\sqrt{n})$	0.474	0.637	0.757	0.877	1.038

[a] In case 0, both ξ and θ are known; in case 1, ξ is unknown, while θ is known; in case 2, ξ is known, while θ is unknown; in case 3, both ξ and θ are unknown.

Along similar lines Chandra, Singpurwalla, and Stephens (1981) considered the **Kolmogorov-Smirnov** statistics D^+, D^-, and D and the Kuiper statistic V given by

$$D^+ = \max_i \left\{ \frac{i}{n} - F_X(X_i') \right\}, \tag{22.171}$$

$$D^- = \max_i \left\{ F_X(X_i') - \frac{i-1}{n} \right\}, \tag{22.172}$$

$$D = \max(D^+, D^-), \tag{22.173}$$

and

$$V = D^+ + D^-. \tag{22.174}$$

They determined some percentage points of these statistics for the three cases when one or both of the parameters ξ and θ need to be estimated from the data (using the **MLEs).** Percentage points of the four statistics in the case when both ξ and θ are unknown, taken from Chandra, Singpurwalla, and Stephens **(1981),** are presented in Table 22.20 for n = 10, 20, 50, and ∞.

Table 22.20 Percentage points of the statistics $\sqrt{n}\,D^+$, $\sqrt{n}\,D^-$, $\sqrt{n}\,D$, and $\sqrt{n}\,V$ when both ξ and θ are unknown

		Upper Tail Significance Level α			
Statistics	n	0.10	0.05	0.025	0.01
$\sqrt{n}\,D^+$	10	0.685	0.755	0.842	0.897
	20	0.710	0.780	0.859	0.926
	50	0.727	0.796	0.870	0.940
	∞	0.733	0.808	0.877	0.957
$\sqrt{n}\,D^-$	10	0.700	0.766	0.814	0.892
	20	0.715	0.785	0.843	0.926
	50	0.724	0.796	0.860	0.944
	∞	0.733	0.808	0.877	0.957
$\sqrt{n}\,D$	10	0.760	0.819	0.880	0.944
	20	0.779	0.843	0.907	0.973
	50	0.790	0.856	0.922	0.988
	∞	0.803	0.874	0.939	1.007
$\sqrt{n}\,V$	10	1.287	1.381	1.459	1.535
	20	1.323	1.428	1.509	1.600
	50	1.344	1.453	1.538	1.639
	∞	1.372	1.477	1.557	1.671

Probability plots are often used to aid assessment of the validity of a statistical distribution; in fact the correlation test is based on such a plot. Unfortunately, owing to the unequal variances of the plotted points, interpretation of the plots is difficult. The stabilized probability plot proposed by Michael *(1983)* is to plot

$$S_i = \frac{2}{\pi}\sin^{-1}\left\{F_X\left(\frac{X_i' - \xi}{\theta}\right)\right\}^{1/2} \tag{22.175}$$

with respect to

$$r_i = \frac{2}{\pi}\sin^{-1}\left(\frac{i - 0.5}{n}\right)^{1/2}.$$

In this way the unequal variance problem can be avoided, since S_i in *(22.175)* have approximately equal variance, as the asymptotic variance of $\sqrt{n}\,S_i$ is $(1/\pi^2)$ independent of q when $n \to \infty$ and $i/n \to q$. *A* goodness-of-fit statistic that arises naturally from the stabilized probability plot is

$$D_{SP} = \max_i |r_i - S_i|. \tag{22.176}$$

Kimber (1985) presented critical values for the statistic D_{SP} in *(22.176)* for some selected choices of n, and these are presented in Table *22.21.*

By starting with a doubly Type-II censored sample $X'_{r+1}, X'_{r+2}, \cdots, X'_{n-s}$ from the type *1* extreme value distribution for the minimum, Lockhart,

Table 22.21 Critical values for the statistic D_{SP}

	α				
n	0.50	0.25	0.10	0.05	0.01
3	0.085	0.109	0.137	0.154	0.167
4	0.096	0.119	0.144	0.167	0.209
5	0.097	0.122	0.148	0.167	0.201
6	0.098	0.124	0.148	0.165	0.201
8	0.096	0.119	0.142	0.157	0.186
10	0.094	0.115	0.136	0.150	0.176
14	0.088	0.107	0.127	0.139	0.163
20	0.082	0.098	0.116	0.127	0.149
30	0.073	0.087	0.103	0.113	0.134
40	0.066	0.079	0.093	0.103	0.122
60	0.059	0.069	0.081	0.089	0.107
80	0.052	0.062	0.072	0.080	0.096
100	0.047	0.056	0.066	0.073	0.089

O'Reilly, and Stephens (1986b) considered three tests based on the normalized spacings

$$Z_i = \frac{X'_{i+1} - X'_i}{E[Y'_{i+1}] - E[Y'_i]}, \qquad i = r+1, r+2, \dots, n-s-1, \quad (22.177)$$

where Y'_i denotes the order statistic from the standard distribution. One may use the exact values of $E[Y'_{i+1}] - E[Y'_i]$ tabulated by Mann, Scheuer, and Fertig (1973) for n = 3(1)25 and Blom's approximation for larger sample sizes. With

$$Z_i^* = \frac{\sum_{j=r+1}^{i} Z_j}{\sum_{j=r+1}^{i+1} Z_j}, \qquad i = r+1, \dots, n-s-2, \qquad (22.178)$$

Lockhart, O'Reilly, and Stephens (1986b) focused on the Anderson-Darling statistic

$$A^2 = -(n-r-s-2) - \frac{1}{n-r-s-2}\left[\sum_{i=r+1}^{n-s-2}(2i-1)\{\log Z_i^* + \log(1 - Z_{n-s-1-i}^*)\}\right], \qquad (22.179)$$

and compared its performance with the S-statistic introduced by Mann, Scheuer, and Fertig (1973) [see also Mann, Fertig, and Scheuer (1971)] and $\bar{Z}^*$ statistic introduced by Tiku and Singh (1981); here

$$T = 1 - Z_t^*,$$

where

$$t = \begin{cases} r + \dfrac{n-r-s}{2} & \text{if } n-r-s \text{ is even} \\ r + \dfrac{n-r-s-1}{2} & \text{if } n-r-s \text{ is odd,} \end{cases} \qquad (22.180)$$

and

$$\bar{Z}^* = \frac{1}{n-r-s-2}\sum_{i=r+1}^{n-s-2} Z_i^*. \qquad (22.181)$$

Through their comparative study Lockhart, O'Reilly, and Stephens (1986b) recommend overall the A^2 test, and they also mentioned that while the $\bar{Z}^*$ test gives good power in many situations, it may also be inconsistent [also see

Lockhart, O'Reilly, and Stephens (1986a) for a general discussion on tests based on normalized spacings].

Tsujitani, Ohta, and Kase (1980) proposed a test based on the sample entropy, presented its critical points for some sample sizes determined through Monte Carlo simulations, and showed that it has desirable power properties compared with some of the tests mentioned above. Öztürk (1986) considered the Shapiro-Wilk W test and presented some percentage points determined through Monte Carlo simulations. A major difficulty of using the W test is the requirement of the variance-covariance matrix of order statistics. To overcome this difficulty, Öztürk (1986) used an approximation for it obtained from the generalized lambda distribution. A modification of the W statistic has been considered by Öztürk and Korukoğlu (1988) in which the test statistic has been obtained as the ratio of two linear estimators of the parameter 8. These authors have determined percentage points of this statistic through Monte Carlo simulations and have also displayed by means of an empirical comparative study that this test possesses good power properties.

By using Kimball's simplified linear estimators $\hat{\xi}$ and $\hat{\theta}$ of ξ and 8 (see Section 9.2), Aly and Shayib (1992) proposed the statistic

$$M_n = -\sum_{i=1}^{n}\left\{\left(\frac{X_i' - \hat{\xi}}{\hat{\theta}}\right) - \log\left[-\log\left(1 - \frac{i}{n+1}\right)\right]\right\}^2 \times\left(1 - \frac{i}{n+1}\right)\log\left(1 - \frac{i}{n+1}\right) \tag{22.182}$$

for testing the validity of the type 1 extreme value distribution for the minimum. They determined the critical points of M_n for some selected sample sizes through Monte Carlo simulations. These values are presented in Table 22.22. Aly and Shayib (1992) also compared the power of this test with some other tests including the A^2 test in (22.179) discussed by Lockhart, O'Reilly, and Stephens (1986b). From this brief power study it seems that the M_n test outperforms the A^2 test for skewed alternatives (like log-Weibull and log-chi-square); however, in the case of symmetric alternatives (like normal and logistic), the A^2 test seems to be considerably better than the M_n test.

Tiago de Oliveira (1981) discussed the statistical choice among the different extreme value models. Vogel (1986) discussed further on the probability plot and the associated correlation coefficient test. Cohen (1986,1988) presented detailed discussions on the large-sample theory for fitting extreme value distributions to maxima. Mann and Fertig (1975) proposed a goodness-of-fit test for the two-parameter Weibull (or the type 1 extreme value distribution for the maximum) against a three-parameter Weibull alternative (see Chapter 21). Aitkin and Clayton (1980) discussed the fitting of extreme

Table 22.22 Critical values for the statistic M_n

	α				α		
n	0.10	0.05	0.01	n	0.10	0.05	0.01
6	1.857	2.803	7.814	19	0.892	1.081	1.499
7	1.577	2.108	4.513	20	0.851	1.011	1.388
8	1.418	1.827	2.991	25	0.803	0.944	1.273
9	1.282	1.586	2.436	30	0.763	0.902	1.251
10	1.176	1.419	2.166	35	0.743	0.866	1.189
11	1.109	1.350	1.953	40	0.723	0.855	1.186
12	1.053	1.260	1.810	45	0.698	0.832	1.098
13	1.002	1.194	1.675	50	0.681	0.806	1.085
14	0.969	1.162	1.666	60	0.648	0.769	1.071
15	0.956	1.142	1.571	70	0.627	0.745	1.038
16	0.935	1.108	1.528	80	0.619	0.722	1.030
17	0.912	1.065	1.514	90	0.588	0.717	1.031
18	0.869	1.044	1.453	100	0.599	0.716	0.998

value distributions to complex censored survival data using the GLIM software.

14 APPLICATIONS

From the very definition of the extreme value distributions, it is clear that these distributions will play a vital role in numerous applied problems. As mentioned earlier in Sections 1 and 2, Gumbel played a pioneering role during the 1940s and 1950s in bringing out several interesting applications for the extreme value data and developing sound statistical methodology to analyze such data. To give a good idea about the variety of applications that have emerged over the years and the order in which these applications have developed, we describe below these applied papers in a chronological order.

The first paper that described an application of the extreme values in flood flows was by Fuller (1914). Griffith (1920) brought out an application while discussing the phenomena of rupture and flow in solids. Next Gumbel (1937a, b) used the extreme value distribution to model radioactive emissions and human lifetimes. The use of the distribution to model the phenomenon of rupture in solids was discussed by Weibull (1939). In this area Weibull effectively advocated the use of reversed type 3 distributions which have now become well-known as ***Weibull distributions*** and have been discussed in great length in Chapter 21.

Gumbel (1941) applied the distribution to analyzing data on. flood flows, and in subsequent work he continued his discussion on the plotting of flood discharges, estimation of flood levels, and forecast of floods [Gumbel (1944, 1945, 1949a)]. Frenkel and Kontorova (1943) used the distribution to study the brittle strength of crystals. The application to study earthquake magnitudes was pointed out by Nordquist (1945). While discussing factors influencing self-purification and their relation to pollution abatement, Velz (1947) used the distribution to model the microorganism survival times. Epstein (1948) applied the theory of extreme values to problems involving fracture data. The role of the extreme value theory in the study of the dielectric strength of paper capacitors was highlighted by Epstein and Brooks (1948). Rantz and Riggs (1949) illustrated an application while analyzing the magnitude and frequency of floods in the Columbia River Basin measured during a U.S. Geological Survey. An interesting new application of the extreme value theory to gust-load problems was brought out by Press (1949). The extreme value distribution was used by Potter (1949) to study rainfall data and to develop normalcy tests of precipitation and facilitate frequency studies of runoff on small watersheds. Weibull (1949) stressed the role of extreme value distributions to represent fatigue failures in solids but, in doing so, advocated once again the use of the Weibull distribution in place of the type 1 extreme value distribution.

The so-called Gumbel method has been applied successfully to both regular-type events (e.g., temperature and vapor pressure) and irregular-type events (e.g., rainfall and wind) but with some deficiencies arising from the asymptotic approximation, as noted by Jenkinson (1955). Thom emphasized how the sparse sampling in time of extreme events obscured much of the information in a rainfall process. He showed how the parameters of a Poisson process could be identified with the annual recurrence rates of hourly rainfalls above certain selected base values. Methods of analysis of extreme hydrological events have changed little since the publication of Gumbel (1941) on asymptotic theory dealing with flood discharges by streams. Assumptions of the theory are that the frequency distribution of extremes within successive intervals remains constant and that observed extremes may be taken as being independent samples from a homogeneous population.

Gumbel (1954,1958) presented consolidated accounts of the statistical theory of extreme values and several practical applications. These works may be studied in conjunction with his later works [Gumbel (1962a, b)] to gain a deeper and better knowledge of extreme value distributions. Thom (1954) applied the distribution while discussing the frequency of maximum wind speeds. In an interesting paper Aziz (1955) applied the extreme value theory to an analysis of maximum pit depth data for aluminum. Kimball (1955) ably explained several practical applications of the theory of extreme values and also described some aspects of the statistical problems associated with them. Jenkinson (1955) applied the extreme value distribution to model the annual maximum or minimum values of some meteorological elements. Lieblein and Zelen (1956) carried out an extensive study relating to inference based on the extreme value distribution and applied their methods to investigate the fatigue life of deep-grove ball bearings. Eldredge (1957) discussed an analysis of corrosion pitting by extreme value statistics and applied it to oil well tubing caliper surveys. King (1959) summarized developments on extreme value theory and explained their implications to reliability analysis. Canfield (1975) and Canfield and Borgman (1975), while discussing various possible statistical distributions as models of time to failure for reliability applications, recommended highly the usage of the type 1 extreme value distribution.

As mentioned earlier in Section 3, Clough and Kotz (1965) gave interesting interpretations for the conditions (22.13)–(22.15) and as a result presented some special queueing model applications for the extreme value distributions. Posner (1965) detailed an application of the extreme value theory to communication engineering; see also the comments by Gumbel and Mustafi (1966) on the paper by Posner. In a series of reports Simiu and Filliben (1975,1976) and Simiu, Bietry, and Filliben (1978) used the extreme value distributions extensively in the statistical analysis of extreme winds.

Shen, Bryson, and Ochoa (1980) applied the distributions for predictions of flood. Watabe and Kitagawa (1980) demonstrated an application while discussing the expectancy of maximum earthquake motions in Japan. While Okubo and Narita (1980) followed the lines of Simiu and Filliben (1975,1976) and used the extreme value distribution to model the data on extreme winds in Japan, Wantz and Sinclair (1981) carried out a similar analysis on the distribution of extreme winds in the Bonneville power administration service area. Metcalfe and Mawdsley (1981) applied extreme value distribution to estimate extreme low flows for pumped storage reservoir designs. The use of the distribution in regional flood frequency estimation and network design was demonstrated by Greis and Wood (1981). Roldan-Canas, Garcia-Guzman, and Losada-Villasante (1982) constructed a stochastic extreme value model for wind occurrence. An application of the extreme value distribution in rainfall analysis was illustrated by Rasheed, Aldabagh, and Ramamoorthy (1983). Henery (1984) presented an interesting application of the extreme value model in predicting the results of horse races. While Pericchi and Rodriguez-Iturbe (1985) used the extreme value distribution in a statistical analysis of floods, Burton and Makropoulos (1985) applied it in an analysis of seismic risk of circum-Pacific earthquakes. The last authors specifically used the extreme values from the type 1 extreme value distribution and their relationship with strain energy release.

A two-component extreme value distribution was proposed by Rossi, Fiorentino, and Versace (1986) for flood frequency analysis; also see the comments on this paper by Beran, Hosking, and Arnell (1986) and Rossi's (1986) subsequent reply. Smith (1987), Jain and Singh (1987), and Ahmad, Sinclair, and Spurr (1988), all provided further discussions on the application of the type 1 extreme value distribution for flood frequency analysis. Achcar, Bolfarine, and Pericchi (1987) discussed the advantages of transforming a survival data to a type 1 extreme value distribution form and then analyzing it. Nissan (1988) demonstrated an interesting application of the type 1 distribution in estimating insurance premiums. The role of statistics of extremes in climatological problems was discussed in great detail by Buishand (1989).

Cockrum, Larson, and Taylor (1990) and Taylor (1991) applied the extreme value distributions in modeling and simulation studies involving product flammability testing. Wiggins (1991) displayed an application in stock markets. A mixture of extreme value distributions was used by Fahmi and Abbasi (1991) to study earthquake magnitudes in Iraq and conterminous regions. Tawn (1992) discussed the estimation of probabilities of extreme sea levels, while Hall (1992) discussed further on flood frequency analysis. Bai, Jakeman, and McAleer (1992) demonstrated an interesting application of the extreme value distribution in predicting the upper percentiles that are of great interest in environmental quality data.

Hopke and Paatero (1993) discussed the extreme value estimation in the study of airborne particles and specifically in the estimation of the size distribution of the aerosol and some related environmental problems. Kanda (1993) considered an empirical extreme value distribution to model maximum load intensities of the earthquake ground motion, the wind speed, and the live load in supermarkets. Goka (1993) applied the extreme value distribution to model accelerated life-test data to tantalum capacitors for space use and to on-orbit data of single event phenomenon of memory integrated circuits in the space radiation environment. Rajan (1993) stressed on the importance of the extreme value theory by providing experimental examples where significant deviations from the average microstructure exist in pertinent materials physics. Some of these examples include the deviations from classical Mullins–von Neumann law for two-dimensional grain growth, the changes occurring in the extreme values of grain size distributions associated with significant changes in materials properties, and the role of extreme values of pore size distributions in synthetic membranes. Scarf and Laycock (1993) and Shibata (1993) have demonstrated some applications of extreme value theory in corrosion engineering. Applications of extreme values in insurance have been illustrated by Teugels and Beirlant (1993).

In addition many more problems and data sets for which the extreme value distributions have been used for the analysis may be seen in the applied books and volumes listed among the References.

15. GENERALIZED EXTREME VALUE DISTRIBUTIONS

The cumulative distribution function of the generalized extreme value distributions is given by

$$F_X(x) = \begin{cases} e^{-\{1-\gamma((x-\xi)/\theta)\}^{1/\gamma}}, & -\infty < x \le \xi + \theta/\gamma \quad \text{when } \gamma > 0, \\ & \xi + \theta/\gamma \le x < \infty \quad \text{when } \gamma < 0, \\ e^{-e^{-(x-\xi)/\theta}}, & -\infty < x < \infty \quad \text{when } \gamma = 0. \end{cases} \tag{22.183}$$

As mentioned already in Section *2*, the distribution above includes the type *2* distribution in **Eq.** (22.2) when $\gamma > 0$, the type *3* distribution in **Eq.** *(22.3)* when $\gamma < 0$, and the type *1* distribution in **Eq.** (22.1) when $\gamma = 0$. The distribution is referred to as the ***uon Mises type extreme value distribution*** or the ***uon Mises–Jenkinson type distribution***. Senkinson (1955) used this generalized distribution to analyze annual maximum or minimum values of certain meteorological elements. The density function corresponding to (22.183) is

$$p_X(x) = \begin{cases} e^{-\{1-\gamma((x-\xi)/\theta)\}^{1/\gamma}} \cdot \dfrac{1}{\theta}\left\{1 - \gamma\left(\dfrac{x-\xi}{\theta}\right)\right\}^{(1/\gamma)-1}, & \\ \qquad -\infty < x \le \xi + \dfrac{8}{\gamma} & \text{when } \gamma > 0, \\ \qquad \xi + \dfrac{8}{\gamma} \le x < \infty & \text{when } \gamma < 0, \\ e^{-e^{-(x-\xi)/\theta}} \cdot \dfrac{1}{\theta} e^{-(x-\xi)/\theta}, & \\ \qquad -\infty < x < \infty & \text{when } \gamma = 0. \end{cases} \tag{22.184}$$

The standard form of the generalized extreme value distributions has cdf

$$F_Y(y) = \begin{cases} e^{-(1-\gamma y)^{1/\gamma}}, & -\infty < y \le 1/\gamma \quad \text{when } \gamma > 0, \\ & 1/\gamma \le y < \infty \quad \text{when } \gamma < 0, \\ e^{-e^{-y}}, & -\infty < y < \infty \quad \text{when } \gamma = 0, \end{cases} \tag{22.185}$$

and pdf

$$p_Y(y) = \begin{cases} e^{-(1-\gamma y)^{1/\gamma}}(1-\gamma y)^{(1/\gamma)-1}, & -\infty < y \le 1/\gamma \quad \text{when } \gamma > 0, \\ & 1/\gamma \le y < \infty \quad \text{when } \gamma < 0, \\ e^{-e^{-y}} e^{-y}, & -\infty < y < \infty \quad \text{when } \gamma = 0. \end{cases} \tag{22.186}$$

Maritz and Munro (1967) studied order statistics from this generalized extreme value distribution, and presented tables of means of order statistics from sample sizes *5* to 10 for the choices of the shape parameter $\gamma = -0.10(0.05)0.40$. These authors have also discussed the estimation of all three parameters ξ, θ, and γ by the use of order statistics.

From Eqs. (22.185) and (22.186), we observe the characterizing differential equation

$$(1-\gamma y)p_Y(y) = -F_Y(y)\log F_Y(y). \qquad (22.187)$$

Balakrishnan, Chan, and Ahsanullah (1993) have exploited the differential equation (22.187) in order to establish several recurrence relations satisfied by the single and the product moments of lower record values. Specifically, let $Y_{L(1)} \equiv Y_1, Y_{L(2)}, \dots$ be the lower record values arising from the sequence $\{Y_i\}$ of i.i.d. random variables with generalized extreme value distribution (22.185). Then, by proceeding on lines similar to those explained in Section 6 and using the differential equation in (22.187), Balakrishnan, Chan, and Ahsanullah (1993) established the following relationships:

$$E\left[Y_{L(n+1)}^{r+1}\right] = \left\{1+\frac{\gamma(r+1)}{n}\right\}E\left[Y_{L(n)}^{r+1}\right] - \frac{r+1}{n}E\left[Y_{L(n)}^{r}\right],$$
$$n=1,2,\dots,\; r=0,1,\dots. \qquad (22.188)$$

$$E\left[Y_{L(m)}^{r+1}Y_{L(m+1)}^{s}\right] = \frac{1}{m+\gamma(r+1)}\left\{(r+1)E\left[Y_{L(m)}^{r}Y_{L(m+1)}^{s}\right] + mE\left[Y_{L(m+1)}^{r+s+1}\right]\right\},$$
$$m=1,2,\dots,\; r,s=0,1,\dots. \qquad (22.189)$$

$$E\left[Y_{L(m)}^{r+1}Y_{L(n)}^{s}\right] = \frac{1}{m+\gamma(r+1)}\left\{(r+1)E\left[Y_{L(m)}^{r}Y_{L(n)}^{s}\right] + mE\left[Y_{L(m+1)}^{r+1}Y_{L(n)}^{s}\right]\right\},$$
$$1\le m\le n-2,\; r,s=0,1,\dots. \qquad (22.190)$$

$$E\left[Y_{L(m)}^{r}Y_{L(m+2)}^{s+1}\right] = \{1+\gamma(s+1)\}E\left[Y_{L(m)}^{r}Y_{L(m+1)}^{s+1}\right] + (s+1)E\left[Y_{L(m)}^{r}Y_{L(m+1)}^{s}\right]$$
$$+\, m\left\{E\left[Y_{L(m+1)}^{r+s+1}\right] - E\left[Y_{L(m+1)}^{r}Y_{L(m+2)}^{s+1}\right]\right\},$$
$$m=1,2,\dots,\; r,s=0,1,\dots. \qquad (22.191)$$

$$E\left[Y_{L(m)}^{r}Y_{L(n+1)}^{s+1}\right] = \frac{1}{n-m}\Big\{\{n-m+\gamma(s+1)\}E\left[Y_{L(m)}^{r}Y_{L(n)}^{s+1}\right]$$
$$-(s+1)E\left[Y_{L(m)}^{r}Y_{L(n)}^{s}\right]$$
$$+m\left\{E\left[Y_{L(m+1)}^{r}Y_{L(n)}^{s+1}\right] - E\left[Y_{L(m+1)}^{r}Y_{L(n+1)}^{s+1}\right]\right\}\Big\},$$
$$1\le m\le n-2,\; r,s=0,1,\dots. \qquad (22.192)$$

From these recurrence relations Balakrishnan, Chan, and Ahsanullah (1993) also deduced the results

$$E[Y_{L(n+1)}] = \left(1 + \frac{\gamma}{n}\right)E[Y_{L(n)}] - \frac{1}{n} \qquad \text{for } n \geq 1,$$

$$\operatorname{cov}(Y_{L(m)}, Y_{L(m+1)}) = \frac{m}{m+\gamma}\operatorname{var}(Y_{L(m+1)}) \qquad \text{for } m \geq 1,$$

$$\operatorname{cov}(Y_{L(m)}, Y_{L(n)}) = \frac{(n-1)^{(n-m)}}{(n-1+\gamma)^{(n-m)}}\operatorname{var}(Y_{L(n)}) \quad \text{for } 1 \leq m \leq n-2,$$

where

$$r^{(i)} = \begin{cases} 1 & \text{if } i = 0, \\ r(r-1)\cdots(r-i+1) & \text{if } i = 1, 2, \ldots . \end{cases}$$

Recurrence relations for product moments involving more than two record values have also been established by these authors. When the shape parameter $y \to 0$, the relations in (22.188)–(22.192) reduce to the corresponding results for the type 1 extreme value distribution presented in Section 6. Ahsanullah and Holland (1994) have discussed the estimation of the location and scale parameters of the generalized extreme value distribution (when y is known) based on the record values.

The maximum likelihood estimation of the parameters ξ, 8, and y have been discussed by a number of authors including Jenkinson (1969), Prescott and Walden (1980, 1983), Hosking (1985), and Macleod (1989). Based on a complete sample of size n from the generalized extreme value distribution (22.183), the Fisher expected information matrix is given by [Prescott and Walden (1980)]

$$E\left[-\frac{\partial^2 \log L}{\partial \xi^2}\right] = \frac{n}{\theta^2}p,$$

$$E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right] = \frac{n}{\theta^2\gamma^2}\{1 - 2\Gamma(2-\gamma) + p\},$$

$$E\left[-\frac{\partial^2 \log L}{\partial \gamma^2}\right] = \frac{n}{\gamma^2}\left\{\frac{\pi^2}{6} + \left(1 - 0.5772157 - \frac{1}{\gamma}\right)^2 + \frac{2q}{\gamma} + \frac{p}{\gamma^2}\right\},$$

$$E\left[-\frac{\partial^2 \log L}{\partial \xi\, \partial \theta}\right] = \frac{n}{\theta^2\gamma}\{p - \Gamma(2-\gamma)\},$$

$$E\left[-\frac{\partial^2 \log L}{\partial \xi\, \partial \gamma}\right] = -\frac{n}{\theta\gamma}\left(q + \frac{p}{\gamma}\right),$$

$$E\left[-\frac{\partial^2 \log L}{\partial \theta\, \partial \gamma}\right] = \frac{n}{\theta\gamma^2}\left[1 - 0.5772157 - \frac{\{1 - \Gamma(2-\gamma)\}}{Y} - q - \frac{p}{Y}\right],$$

where

$$= (1-\gamma)^2\Gamma(1-2\gamma) \quad \text{and} \quad q = \Gamma(2-\gamma)\left\{ \quad -\gamma) - \frac{1-\gamma}{\gamma}\right\}$$

The regularity conditions are satisfied when $\gamma < \frac{1}{2}$, and in this case the asymptotic variances and covariances of the maximum likelihood estimators are given by the elements of the inverse of the Fisher information matrix whose elements are as given above.

Hosking (1985) has presented a FORTRAN subroutine MLEGEV that facilitates the calculation of the maximum likelihood estimates of the parameters ξ, θ, and y (by the Newton-Raphson method) and the variance-covariance matrix of the estimated parameters (by the expressions given above). Macleod (1989) has noted that if the initial estimate for the shape parameter y is 0, then Hosking's algorithm will attempt to calculate 1.0/0.0, which will cause a failure on many compilers. Macleod has therefore suggested an adjustment that should be applied to Hosking's algorithm.

Hosking, Wallis, and Wood (1985) have discussed the method of probability-weighted moments (PWM) for the estimation of the parameters ξ, θ, and y. In this approach one considers the moments

$$\beta_r = E\left[X\{F(X)\}^r\right], \qquad r = 0,1,2,\dots, \tag{22.193}$$

and sets up the necessary number of moment equations by using the sample statistics

$$b_r = \frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)^{(r)}}{(n-1)^{(r)}}X_i', \qquad r = 0,1,2,\dots, \tag{22.194}$$

which are unbiased estimators of the moments β_r (see Section 9.8). One may instead use the simplified estimates

$$\hat{\beta}_r[p_{i,n}] = \frac{1}{n}\sum_{i=1}^{n} p_{i,n}^r X_i', \tag{22.195}$$

where $p_{i,n}$ is a plotting position [a distribution-free estimate of $F(X_i')$] that may be taken as

$$p_{i,n} = \frac{i-a}{n}, \qquad 0 < a < 1,$$

or

$$p_{i,n} = \frac{i-a}{n+1-2a}, \qquad -\frac{1}{2} < a < \frac{1}{2}.$$

For the generalized extreme value distribution, Hosking, Wallis, and Wood (1985) derived

$$\beta_r = \frac{1}{r+1}\left[\xi + \frac{\theta}{\gamma}\left\{1 - \frac{\Gamma(1+\gamma)}{(1+r)^{\gamma}}\right\}\right], \qquad \gamma > -1, \gamma \neq 0. \quad (22.196)$$

They used (22.196) to show that

$$\hat{\beta}_0 = \beta_0 = \xi + \frac{\theta}{\gamma}\{1 - \Gamma(1+\gamma)\}, \quad (22.197)$$

$$2\hat{\beta}_1 - \hat{\beta}_0 = 2\beta_1 - \beta_0 = \frac{\theta}{\gamma}\Gamma(1+\gamma)(1 - 2^{-\gamma}), \quad (22.198)$$

and

$$\frac{3\hat{\beta}_2 - \hat{\beta}_0}{2\hat{\beta}_1 - \hat{\beta}_0} = \frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1 - 3^{-\gamma}}{1 - 2^{-\gamma}}. \quad (22.199)$$

Since the exact solution for γ from Eq. (22.199) requires iterative methods, Hosking, Wallis, and Wood (1985) suggested the approximate estimator

$$\hat{\gamma} = 7.8590c + 2.9554c^2, \quad (22.200)$$

where

$$c = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{3\hat{\beta}_2 - P_0} - \frac{\log 2}{\log 3}.$$

Using the estimator $\hat{\gamma}$ in (22.2001, we readily obtain from Eqs. (22.198) and (22.197) the estimators of θ and ξ to be

$$\hat{\theta} = \frac{(2\hat{\beta}_1 - \hat{\beta}_0)\hat{\gamma}}{\Gamma(1+\hat{\gamma})(1 - 2^{-\hat{\gamma}})} \quad (22.201)$$

and

$$\hat{\xi} = \hat{\beta}_0 + \frac{\hat{\theta}}{\hat{\gamma}}\{\Gamma(1 ++) - 1). \quad (22.202)$$

Using standard arguments, Hosking, Wallis, and Wood (1985) have shown

Table 22.23 Elements of the asymptotic variance-covariance matrix of the PWM estimators of the parameters of the generalized extreme value distribution

γ	w_{11}	w_{12}	w_{13}	w_{22}	w_{23}	w_{33}
−0.4	1.6637	1.3355	1.1405	1.8461	1.1628	2.9092
−0.3	1.4153	0.8912	0.5640	1.2574	0.4442	1.4090
−0.2	1.3322	0.6727	0.3926	1.0013	0.2697	0.9139
−0.1	1.2915	0.5104	0.3245	0.8440	0.2240	0.6815
0.0	1.2686	0.3704	0.2992	0.7390	0.2247	0.5633
0.1	1.2551	0.2411	0.2966	0.6708	0.2447	0.5103
0.2	1.2474	0.1177	0.3081	0.6330	0.2728	0.5021
0.3	1.2438	−0.0023	0.3297	0.6223	0.3033	0.5294
0.4	1.2433	−0.1205	0.3592	0.6368	0.3329	0.5880

that the asymptotic variance-covariance matrix of $(\hat{\xi}\ \hat{\theta}\ \hat{\gamma})'$ is given by

$$\frac{1}{n}\begin{bmatrix} \theta^2 w_{11} & \theta^2 w_{12} & \theta w_{13} \\ & \theta^2 w_{22} & \theta w_{23} \\ & & w_{33} \end{bmatrix} \tag{22.203}$$

where the w's depend only on y. Values of these elements for different choices of the shape parameter y, taken from Hosking, Wallis, and Wood (1985), are presented in Table 22.23. The asymptotic efficiency of the individual PWM estimators and the overall efficiency (determined by determinants of the variance-covariance matrices) are presented in Figure 22.2 [taken from Hosking, Wallis, and Wood (1985)].

In defining partial probability-weighted moments, Wang (1990) discussed the estimation of the parameters of the generalized extreme value distribution based on censored samples. Prescott and Walden (1983) have discussed the maximum likelihood estimation of the parameters ξ, 8, and γ a doubly Type-II censored sample $X'_{r+1}, \cdots, X'_{n-s}$ (where the smallest r and the largest s observations are censored in a sample of size n) from the generalized extreme value distribution (22.183). They have also presented expressions for the asymptotic variance-covariance matrix of these MLEs.

Smith (1984) has discussed a choice probability characterization of generalized extreme value models. Testing whether the shape parameter γ is zero in the generalized extreme value distributions for the data at hand has been discussed by Hosking (1984). Some goodness-of-fit tests for the generalized extreme value distributions have been examined by Chowdhury, Stedinger, and Lu (1991). An excellent discussion on the models for exceedances over high thresholds by Davison and Smith (199) provides further insight into issues relating to these distributions. By giving a predictive likelihood that

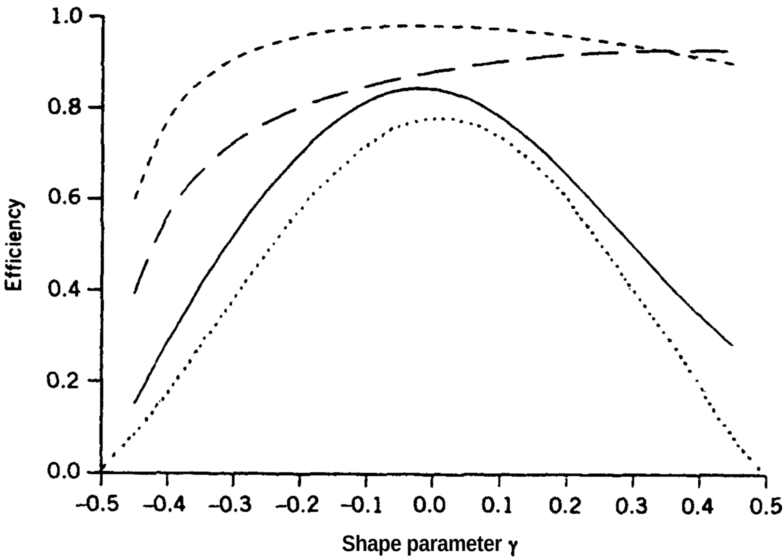


Figure 22.2 Asymptotic Efficiency of PWM Estimators of Parameters of the GEV Distribution: —$\hat{\gamma}$; – – –$\hat{\theta}$; ------$\hat{\xi}$; ...... overall efficiency (i.e., ratio of determinants of asymptotic covariance matrices of ML and PWM estimators).

approximates both Bayes and maximum likelihood predictive inference, Davison (1986) has applied it to the prediction of extremes using the generalized extreme value distribution.

As has already been pointed out in Section 2, the Gompertz (1825) distribution of lifetimes is a reparametrization of the type 1 extreme value distribution. This distribution gives good fit to data from clinical trials on older subjects and is also useful in the construction of life tables [Stephens (1977)]. The cdf has been given by various authors in different forms. Garg, Raja Rao, and Redmond (1970) have defined it in terms of the hazard rate (or the force of mortality) as

$$r(t) = Ke^{\alpha t}, \qquad t \geq 0$$

yielding the survival function

$$1 - F(t) = e^{-K(e^{\alpha t}-1)/\alpha}, \qquad t \geq 0 \tag{22.204}$$

and the probability density function

$$p(t) = Ke^{\alpha t}e^{-K(e^{\alpha t}-1)/\alpha}, \qquad t \geq 0.$$

Actually Gompertz (1825) defined a function (or transformation) $y(t) = K(e^{a'} - 1)/\alpha$ which transforms the random variable T into $y(T)$ which is exponentially distributed with mean 1.

Ahuja (1971) provided the classical definition in terms of the distribution function

$$F(t) = e^{-\rho e^{-t/\mu}}, \qquad -\infty < t < \infty, \tag{22.205}$$

which was earlier generalized by Ahuja and Nash (1967), by introducing an additional shape parameter 4, with the density function

$$p(t; \rho, \mu\, \phi) = \frac{1}{\mu\Gamma(\phi)}\left(\rho\, e^{-t/\mu}\right)^{\phi} e^{-\rho e^{-t/\mu}}, \qquad -\infty < t < \infty. \tag{22.206}$$

The cumulants of the distribution (22.205) are

$$\kappa_1 = \mu(\gamma + \log\rho), \quad \kappa_2 = \frac{\pi^2\mu^2}{6}, \quad \kappa_3 \simeq 2.404\mu^3,$$

ignoring the terms $e^{-\rho}$ [Revfeim (1984b)].

Garg, Raja Rao, and Redmond (1970) observed the following property of the Gompertz distribution (22.204). If the origin is shifted to the point (i.e., by writing $t' = t - t_0$ so that $t' \geq 0$), the density remains in the form

$$p(t') = K' e^{\alpha t'} e^{-K'(e^{\alpha t'} - 1)/\alpha}, \qquad t' \geq 0$$

with $K' = Ke^{\alpha t_0} = r(t_0)$, the hazard rate at t_0. Thus, truncating a Gompertz distribution at time t, and setting the origin at t, leaves the distribution unchanged except that the constant K changes to K'.

Garg, Raja Rao, and Redmond (1970) have also discussed the maximum likelihood estimation of the parameters based on censored samples and grouped data. For example, consider the time interval $[0, t_m)$ subdivided into m subintervals $[0, t_1), [t_1, t_2), \cdots, [t_{m-1}, t_m)$. Let

n = number of individuals in the sample,

d_i = observable number of individuals falling (dying) within the time interval $[t_{i-1}, t_i)$,

s_i = observable number of individuals surviving upto time t_i and lost or withdrawn from the followup

for $i = 1, 2, \ldots, m$. Then the log-likelihood function is given by

$$\log L = \text{const.} + D \log K + \alpha T - \frac{K}{a} Q(\alpha),$$

where

$$T = \sum_{i=1}^{m} d_i \tau_i,$$

$$D = \sum_{i=1}^{m} d_i,$$

$$Q(\alpha) = \sum_{i=1}^{m} \left\{ s_i(e^{\alpha t_i} - 1) + d_i(e^{\alpha \tau_i} - 1) \right\}$$

yielding the maximum likelihood estimator of K as

$$\hat{K} = \frac{D\hat{\alpha}}{Q(\hat{\alpha})}, \tag{22.207}$$

and the solution of the equation

$$T + \frac{D}{\alpha} - D\frac{Q'(\alpha)}{Q(\alpha)} = 0 \tag{22.208}$$

as the maximum likelihood estimator of a. An iterative solution to **(22.208)** can be achieved by Newton's method; the initial estimate a, may be selected as the least-squares estimate of a obtained by calculating the numerical value of the force of mortality, $r(t)$, for each t for the data and minimizing

$$\sum_t \{\log r(t) - \log K - at)'. \tag{22.209}$$

The maximum likelihood estimate of K may then be obtained from **(22.207)**. Numerical data based on an experiment to determine the effects of prolonged oral conception on mortality of mice conducted by Garg, Raja Rao, and Redmond **(1970)** showed that this distribution described quite well the mortality of the mice in each of the five treatment groups. Furthermore the fit was observed to improve substantially by the use of the maximum likelihood estimators as compared to the least-squares estimators.

Ahuja (1972) concentrated on the generalized Gompertz density **(22.206)** and showed that, given two independent random variables X and Y with respective generalized Gompertz densities

$$p_X(t_1; \rho_1, \mu, \phi) \quad \text{and} \quad p_Y(t_2; \rho_2, \mu, \theta), \tag{22.210}$$

then the conditional density function of X given Z = X − Y = z is a generalized Gompertz density

$$p\left(t;\rho_1+\rho_2 e^{z/\mu},\mu,\phi+\theta\right).$$

This property may be compared with damage models involving binomial and Poisson distributions, and also with normal distributions (see Chapters 4 and 13). Moreover the characteristic function corresponding to the generalized Gompertz density *(22.206)* is given by

$$\psi(u;\rho,\mu,\phi)=e^{i\mu u}\frac{\Gamma(\phi-i\mu u)}{\Gamma(\phi)}; \tag{22.211}$$

hence the characteristic function corresponding to the difference of two independent Gompertz variables with parameters (ρ_1,μ,ϕ) and (ρ_2,μ,θ) is given by

$$\psi_Z(u)=\left(\frac{\rho_1}{\rho_2}\right)^{i\mu u}\frac{\Gamma(\phi-i\mu u)\Gamma(\theta+ipu)}{\Gamma(\phi)\Gamma(\theta)} \tag{22.212}$$

The characteristic function of Z in *(22.212)* readily shows that Z is a generalized logistic random variable (see Chapter 23) with density function

$$p_Z(z;\rho,\mu,\phi,\theta)=\frac{1}{\mu B(\phi,\theta)}\cdot\frac{\left(\rho e^{-z/\mu}\right)^{\phi}}{\left(1+\rho e^{-z/\mu}\right)^{\phi+\theta}},\qquad -\infty<z<\infty. \tag{22.213}$$

Scarf (1992) has considered a four-parameter generalized extreme value distribution, and discussed the maximum likelihood estimation and the probability-weighted moment estimation of the parameters. Scarf has noted that in certain applications, data on extremes arise as paired observations (X_i,t_i), $i=1,2,\ldots,$ n, where X_i is observed at time t_i, independently of X, at time t,. One such application arises in metallic corrosion where X, is the depth of the largest pit penetration over a standard area of metal surface exposed to a corrosive environment for time t_i. In this situation Scarf *(1992)* has proposed a four-parameter form of the generalized extreme value distribution as

$$F_{X,t}(x)=e^{-\{1-\gamma[(xt^{-\beta}-\xi)/\theta]\}^{1/\gamma}},\qquad \gamma xt^{-\beta}<\xi\gamma+\theta,\qquad \theta,\beta>0. \tag{22.214}$$

Scarf has then discussed methods of estimation of the four parameters ξ, *8*, γ, and β.

16 OTHER RELATED DISTRIBUTIONS

There is clearly a close connection between the three types of extremal distributions. As seen in the last section, the standard type 1 extreme value distribution is a transitional limiting form between type 2 and type 3 (Weibull) distributions. Furthermore, as mentioned in Section 9 (and also in Chapter 21), a logarithmic transformation of a Weibull random variable results in a type 1 extreme value random variable. Also, if Y is a standard type 1 extreme value random variable with density (22.26), then e^{-Y} has a standard exponential distribution (as noted earlier in Section 4).

A rather unexpected relation holds between the logistic and type 1 distributions. If two independent random variables each have the same type 1 distribution, their difference has a logistic distribution [Gumbel (1961)]. Gumbel (1962c, d) has also studied the distribution of products and ratios of independent variables having extreme value distributions. Tables of the distribution of the "extremal quotient" [(greatest)/(−least), i.e., $X'_n/(-X'_1)$] have been published by Gumbel and Pickands (1967).

Limiting distributions of second, third, and so forth, greatest (or least) values may also be regarded as being related to extreme value distributions. Gumbel (1958) has shown that under the same conditions as those leading to the type 1 extreme value distribution, the limiting distribution of the rth greatest value $Y'_{n-r+1} = (X'_{n-r+1} - \xi)/\theta$ has the standard form of probability density function

$$p_{Y'_{n-r+1}}(y) = r^r[(r-1)!]^{-1}\exp[-ry - re^{-y}]. \qquad (22.215)$$

$100\alpha\%$ points of this distribution are given by Gumbel (1958) to 5 decimal places for

$r = 1(1)15(5)50,$
$\alpha = 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995.$

The moment-generating function of distribution (22.215) is

$$\frac{r^t\Gamma(r-t)}{\Gamma(r)}.$$

The cumulant-generating function is

$$t\log r + \log\Gamma(r-t) - \log\Gamma(r)$$

so the cumulants are

$$\kappa_1 = \log r - \psi(r)$$
$$\kappa_s = (-1)^r\psi^{(s-1)}(r), \qquad s \geq 2. \qquad (22.216)$$

The limiting distribution (22.215), which corresponds to a fixed value of r, should be distinguished from distributions obtained by allowing r to vary with n (usually in such a way that r/n is nearly constant) or by keeping r constant but varying the argument value. Borgman (1961), for example, has shown that if x_n be defined by $F_X(x_n) = 1 - w/n$, for given fixed w [where $F_X(x)$ is the cumulative distribution function of the population distribution], then

$$\lim_{n \to \infty} \Pr[X'_{n-r+1} \le x_n] = 1 - [(r-1)!]^{-1} \int_0^w t^{r-1} e^{-t}\, dt. \quad (22.217)$$

The right-hand side of (22.217) can also be written in terms of a χ^2 distribution, as $\Pr[\chi^2_{2r} > 2w]$.

The asymptotic distribution of *range* is naturally closely connected with extreme value distributions. If both the greatest and least values have limiting distributions of type 1, then [Gumbel (1947)] the limiting distribution of the range, R, is of form

$$\Pr[R \le r] = 2e^{-r/2} K_1(2e^{-r/2}), \qquad r > 0, \qquad (22.218)$$

with probability density function

$$p_R(r) = 2e^{-r} K_0(2e^{-r/2}), \qquad r > 0,$$

where K_0, K_1 are modified Bessel functions of the second kind of orders zero, one, respectively. Gumbel gives the values

$$E[R] = 2\gamma = 1.15443,$$

$$\text{Median } R = 0.92860,$$

$$\text{Modal } R = 0.50637.$$

Also

$$\operatorname{var}(R) = \frac{\pi^2}{3} = 3.2899.$$

In Gumbel (1949b) there are tables of $\Pr[R \le r]$ and $p_R(r)$ to seven decimal places for

$$r = -4.6(0.1) - 3.3(0.05)11.00(0.5)20.0,$$

and of percentile points R_α to four decimal places for

$$\alpha = 0.0002(0.0001)0.0010(0.001)0.010(0.01)0.95(0.001)0.998$$

and to three decimal places for

$$\alpha = 0.001, 0.999(0.0001)0.9999.$$

Some forms of ***generalized*** and ***compound*** type 1 extreme value distributions have been constructed by Dubey (1969). He generalizes the distribution by introducing an extra parameter τ, defining the cumulative distribution function by the equation

$$\Pr[X \le x] = \exp\left[-\tau\theta \exp\left\{-\frac{x-\xi}{\theta}\right\}\right]. \qquad (22.219)$$

However, since

$$\tau\theta \exp\left\{-\frac{x-\xi}{\theta}\right\} = \exp\left\{-\frac{x-\xi'}{\theta}\right\}$$

with $\xi' = \xi + \theta \log \tau\theta$, it can be seen that X still has an ordinary type 1 distribution. This generalized distribution is, however, introduced only as an intermediate step in the construction of a ***compound*** type 1 extreme value distribution, which can be denoted formally as

$$\text{"Generalized" type 1 extreme value } (\xi, \theta, \tau) \bigwedge_{\tau} \text{Gamma } (p, \beta)$$

Here τ is supposed to have probability density function

$$p_\tau(t) = \frac{\beta^p}{\Gamma(p)} t^{p-1} e^{-\beta t}, \qquad t > 0; p > 0, \beta > 0.$$

The resulting compound distribution has cumulative distribution function

$$\begin{aligned}\Pr[X \le x] &= \left[\frac{\beta^p}{\Gamma(p)}\right] \int_0^\infty t^{p-1} \exp\left[-t\left\{\beta + \theta \exp\left(-\frac{x-\xi}{\theta}\right)\right\}\right] dt \\ &= \left[1 + \theta\beta^{-1} \exp\left\{-\frac{x-\xi}{\theta}\right\}\right]^{-p}.\end{aligned}$$

We may note that this distribution, different from the generalized logistic distribution introduced by Ahuja and Nash (1967), can also be regarded as a ***generalized logistic distribution.*** [See Hald (1952) and Chapter 23, Section 10.1 In fact this is termed a ***type I generalized logistic distribution*** in Chapter 23. By considering a cumulative distribution function

$$\Pr[X \le x] = 1 - \exp\left[-\tau\theta \exp\left\{\frac{x-\xi}{\theta}\right\}\right] \qquad (22.221)$$

and using a similar gamma compounding, Balakrishnan and Leung (1988a) derived the cumulative distribution function

$$\Pr[X \le x] = 1 - e^{-p(x-\xi)/\theta}\left[\theta\beta^{-1} + \exp\left\{-\frac{x-\xi}{\theta}\right\}\right]^{-p}. \quad (22.222)$$

This distribution has been termed a ***Type-II generalized logistic distribution*** in Chapter *23.* As mentioned there, the type I and Type-I1 generalized logistic distributions are related by a simple negation of the random variables.

Proceeding similarly, Balakrishnan and Leung (1988a) started with the exponential-gamma density function

$$p_X(x|\tau) = \exp\left[-\tau\exp\left\{-\frac{x-\xi}{\theta}\right\}\right]\exp\left\{-\frac{\kappa(x-\xi)}{\theta}\right\}\frac{\tau^\kappa}{\theta\Gamma(\kappa)},$$

$$-\infty < x < \infty, \kappa > 0, \theta > 0, \quad (22.223)$$

and compounded it with a gamma density function for τ to derive the density function

$$p_X(x) = \int_0^\infty e^{-te^{-(x-\xi)/\theta}} e^{-\kappa(x-\xi)/\theta}\frac{t^\kappa}{\theta\Gamma(\kappa)}\frac{\beta^p}{\Gamma(p)}t^{p-1}e^{-\beta t}\,dt$$

$$= \frac{\beta^p}{\theta\Gamma(p)\Gamma(\kappa)}e^{-\kappa(x-\xi)/\theta}\int_0^\infty e^{-t[\beta+e^{-(x-\xi)/\theta}]}t^{\kappa+p-1}\,dt$$

$$= \frac{1}{\theta B(\kappa,p)}\frac{[\beta^{-1}\exp\{-(x-\xi)/\theta\}]^\kappa}{[1+\beta^{-1}\exp\{-(x-\xi)/\theta\}]^{\kappa+p}},$$

$$-\infty < x < \infty, \kappa > 0, p > 0, \theta > 0. \quad (22.224)$$

The density function in (22.224) has been termed a ***type IV generalized logistic density*** in Chapter 23. For the special case when $p = \kappa$, the type ***IV*** generalized logistic density function in (22.224) becomes symmetric about $x = \xi$ and has been referred to as a ***Type-III generalized logistic density*** in Chapter 23.

The standard log-gamma density function

$$p_Y(y) = \frac{1}{\Gamma(\kappa)}e^{\kappa y - e^y}, \qquad -\infty < y < \infty, \kappa > 0 \qquad (22.225)$$

can be considered as a generalization of the standard type 1 extreme value density. Specifically, if Y has the density function in (22.225), for the case when $\kappa = 1$ the variable $-Y$ is distributed simply as a standard type 1 extreme value random variable. We may note that for integral values of κ, (22.225) is related to the density (22.215). The cumulative distribution function corresponding to the density (22.225) is

$$F_Y(y) = I_{e^y}(\kappa), \qquad -\infty < y < \infty, \kappa > 0, \tag{22.226}$$

where $I_t(\kappa)$ is the incomplete gamma function ratio

$$I_t(\kappa) = \int_0^t \frac{1}{\Gamma(\kappa)} e^{-z} z^{\kappa-1}\, dz, \qquad 0 < t < \infty, \kappa > 0.$$

For integral values of κ therefore we have (see Chapter 17)

$$1 - F_Y(y) = e^{-e^y} \sum_{i=0}^{\kappa-1} \frac{e^{iy}}{i!}, \qquad -\infty < y < \infty, \kappa = 1, 2, \dots. \tag{22.227}$$

The moment-generating function corresponding to the density (22.225) is

$$E[e^{tY}] = \frac{\Gamma(\kappa + t)}{\Gamma(\kappa)};$$

in particular, we have

$$E[Y] = \psi(\kappa) \quad \text{and} \quad \text{var}(Y) = \psi'(\kappa). \tag{22.228}$$

Since $\psi(\kappa) \sim \log \kappa$ and $\psi'(\kappa) \sim 1/\kappa$ for large κ, Prentice (1974) suggested a reparametrized log-gamma density function

$$p_Y^*(Y) = \frac{\kappa^{\kappa - 1/2}}{\Gamma(\kappa)} e^{\sqrt{\kappa} y - \kappa e^{y/\sqrt{\kappa}}}, \qquad -\infty < y < \infty, \kappa > 0 \tag{22.229}$$

which tends to the standard normal density function as $\kappa \to \infty$. By introducing a location parameter ξ and a scale parameter θ in the density (22.225), we obtain a three-parameter log-gamma density function as

$$p_X(x) = \frac{1}{\theta \Gamma(\kappa)} e^{\kappa(x-\xi)/\theta} e^{-e^{(x-\xi)/\theta}}, \qquad \kappa > 0, \theta > 0. \tag{22.230}$$

This is clearly a generalization of the type 1 extreme value density function (22.25). Lawless (1980,1982) has illustrated the usefulness of the three-parameter log-gamma density (22.230) as a life-test model and the maximum likelihood estimation of the parameters; also see Prentice (1974). Balakrishnan and Chan (1994a, b, c, d) have discussed order statistics from this distribution and also the best linear unbiased estimation, the asymptotic best linear unbiased estimation, and the maximum likelihood estimation of the parameters based on complete as well as Type-I1 censored samples. Young and Bakir (1987) have discussed the log-gamma regression model. Lawless (1980) and DiCiccio (1987) have discussed inferential procedures for a related generalized gamma distribution (see Chapter 17 for details). Mihram (1975) referred to this distribution as a generalized extreme value distribution and discussed some basic properties of the distribution (like the closure under linear transformation, shapes, etc.) and inferential methods for the parameters (like sufficiency, efficiency, etc.).

A two-component mixture of extreme value distributions with density function

$$p_X(x) = \frac{\alpha}{\theta} e^{-(x-\xi)/\theta} e^{-e^{-(x-\xi)/\theta}} + \frac{1-\alpha}{\theta^*} e^{-(x-\xi^*)/\theta^*} e^{-e^{-(x-\xi^*)/\theta^*}},$$
$$-\infty < x < \infty, 0 < \alpha < 1, \theta > 0, \quad \theta^* > 0, \quad (22.231)$$

and cumulative distribution function

$$F_X(x) = \alpha e^{-e^{-(x-\xi)/\theta}} + (1-\alpha) e^{-e^{-(x-\xi^*)/\theta^*}}, \qquad -\infty < x < \infty, \quad (22.232)$$

has also been used in some applied problems. The moment-generating function of this distribution is

$$M_X(t) = \alpha e^{t\xi}\Gamma(1 - \mathrm{Bt}) + (1 - \mathrm{a}) e^{t\xi^*}\Gamma(1 - \theta^* t), \qquad |t|\max(\theta, \theta^*) < 1. \quad (22.233)$$

In particular, the mean and variance are

$$E[X] = \{\alpha(\xi - \xi^*) + \xi^*\} + \gamma\{\alpha(\theta - \theta^*) + \theta^*\} \qquad (22.234)$$

and

$$\mathrm{var}(X) = \frac{\pi^2}{6}\{\alpha\theta^2 + (1-\alpha)\theta^{*2}\} + \alpha(1-\alpha)\{(\xi - \xi^*) + \gamma(\theta - \theta^*)\}^2. \quad (22.235)$$

Rossi, Fiorentino, and Versace (1986) have made use of this two-component extreme value distribution for flood frequency analysis; also see Beran, Hosking, and Arnell (1986) and Rossi (1986) for additional comments in this regard.

Revfeim (1984a) discussed an alternative parametric form of the type 1 extreme value distribution for the maximum and used it to derive an extended family of type 1 extreme value distributions. To be specific, let us suppose that events occur in a Poisson process of rate p. If the sizes of the events are distributed independently of occurrence and of each other with cdf $G(x)$, then the maximum sizes within unit time intervals have cdf

$$F(x) = \frac{e^{-\rho\{1-G(x)\}} - e^{-\rho}}{1 - e^{-\rho}}. \tag{22.236}$$

If p is large, then $e^{-\rho}$ is negligible. For the exponential distribution with $G(x) = 1 - e^{-x/\mu}$, (22.236) then gives (for large ρ)

$$F(x) = e^{-\rho e^{-x/\mu}}, \tag{22.237}$$

which is just a reparametrized form of the type 1 extreme value distribution for the maximum in (22.1) [Revfeim (1984b)]; see also Revfeim (1984c) and Revfeim and Hessell (1984). Next, choosing $G(x)$ to be the gamma distribution of integer order p with cdf

$$G(x) = 1 - e^{-x/\mu} \sum_{i=0}^{p-1} \frac{(x/\mu)^i}{i!},$$

Revfeim (1984a) derived from (22.236) an extended family of type 1 extreme value distributions with cdf

$$F(x) = \exp\left\{-\rho\, e^{-x/\mu} \sum_{i=0}^{p-1} \frac{(x/\mu)^i}{i!}\right\}. \tag{22.238}$$

For p = 1, (22.238) reduces to the type 1 distribution in (22.237). Moment properties of the distribution (22.238) for integral p > 1 have been discussed by Revfeim (1984a). Revfeim and Hessell (1984) have applied the distribution (22.238) to model extreme wind gusts. The distribution (22.238) was also derived by Zelenhasic (1970) in connection with river flow exceedances. The mean of this distribution is approximated by

$$E[X] \simeq \mu a(\log \rho + b)$$

where a and b are functions of the gamma shape parameter p. For the value of p = 8 (the most likely value for maximum wind gusts), a and b are 1.58 and 6.00, respectively. Similarly a and b are 1.31 and 3.55 when p = 4, and 1.13 and 1.82 when p = 2. (Note that for p = 1, a = 1 and b = γ.)

Maximum likelihood estimators of μ and p, when p is known, are given by

$$\hat{\mu} = \frac{\bar{X}}{p}\left(1 + \frac{\hat{S}_1}{\hat{S}_0}\right) \quad \text{and} \quad \hat{\rho} = \frac{1}{\hat{S}_0}, \tag{22.239}$$

where

$$S_0 = \sum_{i=0}^{p-1} \frac{Z_i}{i!},$$

$$S_1 = \frac{Z_p}{p!},$$

$$S_2 = \frac{Z_{p+1}}{p!}$$

with

$$Z_i = \frac{1}{n} \sum_{j=1}^{n} e^{-x_j/\mu} \left(\frac{x_j}{\mu} \right)^i.$$

$\hat{\mu}$ can be obtained iteratively by dividing the current value of μ by 1 + D, where

$$D = \frac{S_0 + S_1 - \bar{X} S_0/(p\mu)}{S_2 + \bar{X}[(S_0/p) - S_1]/\mu}.$$

The general formula for the kth raw moment of the distribution (22.228) is

$$E[X^k] = \mu^k \frac{p}{\Gamma(p)} \int_0^1 (-\log y)^{p+k-1} e^{-py \sum_{i=0}^{p-1} (-\log y)^i / i!} \, dy,$$

$$\text{with } y = e^{-x/\mu}. \qquad (22.240)$$

This is difficult to evaluate even numerically, due to the singularity at $y = 0$, and especially for large p and k.

REFERENCES

Abbasi, J. N. Al, and Fahmi, K. J. (1991). GEMPAK: A FORTRAN-77 program for calculating Gumbel's first, third, and mixture upper earthquake magnitude distributions employing maximum likelihood estimation, ***Computers & Geosciences*, *17*,** **271–290.**

Abdelhafez, M. E. M., and Thomas, D. R. (1990). Approximate prediction limits for the Weibull and extreme value regression models, ***Egyptian Statistical Journal,*** **34,408–419.**

Abdelhafez, M. E. M., and Thomas, D. R. (1991). Bootstrap confidence bands for the Weibull and extreme value regression models with randomly censored data, ***Egyptian Statistical Journal, 35,*** **95–109.**

Abramowitz, M., and Stegun, I. A. (eds.) (1965). ***Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,*** New York: Dover.

Achcar, J. A. (1991). A useful reparametrization for the extreme value distribution, ***Computational Statistics Quarterly, 6,*** 113–125.

Achcar, J. A., Bolfarine, H., and Pericchi, L. R. (1987). Transformation of survival data to an extreme value distribution, ***The Statistician,*** **36,** 229–234.

Ahmad, M. I., Sinclair, C. D., and Spurr, B. D. (1988). Assessment of flood frequency models using empirical distribution function statistics, ***Water Resources Research,*** **24,** 1323–1328.

Ahmed, E. (1989). On the probability of selecting extreme value populations with the smallest location parameters, ***The Statistician,*** **38,** 191–195.

Ahsanullah, M. (1990). Estimation of the parameters of the Gumbel distribution based on the *m* record values, ***Computational Statistics Quarterly, 5,*** 231–239.

Ahsanullah, M. (1991). Inference and prediction problems of the Gumbel distribution based on record values, ***Pakistan Journal of Statistics, Series B, 7,*** 53–62.

Ahsanullah, M., and Holland, B. (1994). On the use of record values to estimate the location and scale parameters of the generalized extreme value distribution, ***Sankhyii, Series B*** (to appear).

Ahuja, J. C. (1972). On certain properties of the generalized Gompertz distribution, ***Sankhyii,*** Series ***B, 34,*** 541–544.

Ahuja, J. C., and Nash, S. W. (1967). The generalized Gompertz-Verhulst family of distributions, ***Sankhyii, Series A,*** **29,** 141–156.

Aitkin, M., and Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM, ***Applied Statistics,*** **29,** 156–163.

Aly, E.-E. A. A., and Shayib, M. A. (1992). On some goodness-of-fit tests for the normal, logistic and extreme-value distributions, ***Communications in Statistics—Theory and Methods,*** **21,** 1297–1308.

Antle, C. E., and Rademaker, F. (1972). An upper confidence limit on the maximum of *m* future observations from a type I extreme value distribution, ***Biometrika,*** **59,** 475–477. Correction, **68,** 738.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). ***A First Course in Order Statistics,*** New York: Wiley.

Ashour, S. K., and El-Adl, Y. M. (1980). Bayesian estimation of the parameters of the extreme value distribution, ***Egyptian Statistical Journal,*** **24 / 2,** 140–152.

Aziz, P. M. (1955). Application of the statistical theory of extreme values to the analysis of maximum pit depth data for aluminum, ***Corrosion,*** **12,** 495–506.

Bai, J., Jakeman, A. J., and McAleer, M. (1992). On the use of extreme value distributions for predicting the upper percentiles of environmental quality data, ***Mathematics and Computers in Simulation,*** **33,** 483–488.

Bain, L. J. (1972). Inferences based on censored sampling from the Weibull or extreme-value distribution, ***Technometrics,*** **14,** 693–702.

Bain, L. J., and Antle, C. E. (1967). Estimation of parameters in the Weibull distribution, ***Technometrics,*** **9,** 621–627.

Bain, L. J., and Engelhardt, M. (1991). ***Statistical Analysis of Reliability and Life-testing Models.*** New York: Dekker.

Balakrishnan, N., Ahsanullah, M., and Chan, P. S. (1992). Relations for single and product moments of record values from Gumbel distribution, ***Statistics & Probability Letters,*** **15,** 223–227.

Balakrishnan, N., Balasubramanian, K., and Panchapakesan, S. (1994). δ-exceedance records, ***Submitted for publication.***

Balakrishnan, N., and Chan, P. S. (1992a). Order statistics from extreme value distribution, I: Tables of means, variances and covariances, ***Communications in Statistics—Simulation and Computation, 21,*** 1199–1217.

Balakrishnan, N., and Chan, P. S. (1992b). Order statistics from extreme value distribution, II: Best linear unbiased estimates and some other uses, ***Communications in Statistics—Simulation and Computation,*** **21,** 1219–1246.

Balakrishnan, N., and Chan, P. S. (1992c). Extended tables of means, variances and covariances of order statistics from the extreme value distribution for sample sizes up to 30, ***Report,*** Department of Mathematics and Statistics, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Chan, P. S. (1992d). Extended tables of best linear unbiased estimates from complete and Type-I1 censored samples from the extreme value distribution for sample sizes up to 30, ***Report,*** Department of Mathematics and Statistics, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Chan, P. S. (1994a). Log-gamma order statistics and linear estimation of parameters, ***Computational Statistics & Data Analysis*** (to appear).

Balakrishnan, N., and Chan, P. S. (1994b). Asymptotic best linear unbiased estimation for log-gamma distribution, ***Sankhyā, Series*** B (to appear).

Balakrishnan, N., and Chan, P. S. (1994c). Maximum likelihood estimation for the log-gamma distribution under Type-I1 censored samples and associated inference, In ***Recent Advances in Life-testing and Reliability*** (ed., *N.* Balakrishnan), Boca Raton, FL: CRC Press, pp. 409–437.

Balakrishnan, N., and Chan, P. S. (1994d). Maximum likelihood estimation for the three-parameter log-gamma distribution, In ***Recent Advances in Life-testing and Reliability*** (ed., *N.* Balakrishnan), Boca Raton, FL: CRC Press, pp. 439–453.

Balakrishnan, N., Chan, P. S., and Ahsanullah, M. (1993). Recurrence relations for moments of record values from generalized extreme value distribution, ***Communications in Statistics—Theory and Methods, 22,*** 1471–1482.

Balakrishnan, N., and Cohen, A. C. (1991). ***Order Statistics and Inference: Estimation Methods,*** San Diego, CA: Academic Press.

Balakrishnan, N., Gupta, S. S., and Panchapakesan, S. (1992). Estimation of the location and scale parameters of the extreme value distribution based on multiply Type-I1 censored samples, ***Technical Report,*** Department of Statistics, Purdue University, West Lafayette, IN.

Balakrishnan, N., and Leung, M. Y. (1988). Order statistics from the Type I generalized logistic distribution, ***Communications in Statistics—Simulation and Computation,*** **17,** 25–50.

Balakrishnan, N., and Varadan, J. (1991). Approximate MLEs for the location and scale parameters of the extreme value distribution with censoring, ***IEEE Transactions on Reliability, 40,*** 146–151.

Ballerini, R. (1987). Another characterization of the type I extreme value distribution, ***Statistics & Probability Letters,*** **5,** 83–85.

Ballerini, R., and Resnick, S. I. (1985). Records from improving populations, ***Journal of Applied Probability*, 22**, 487–502.

Ballerini, R., and Resnick, S. I. (1987a). Records in the presence of a linear trend, ***Advances in Applied Probability*, 19**, 801–828.

Ballerini, R., and Resnick, S. I. (1987b). Embedding sequences of successive maxima in extremal processes, with applications, ***Journal of Applied Probability*, 24**, 827–837.

Beran, M., Hosking, J. R. M., and Arnell, N. (1986). Comment on "Two-component extreme value distribution for flood frequency analysis" by Fabio Rossi, Mauro Fiorentino, Pasquale Versace, ***Water Resources Research*, 22**, 263–266.

Berman, S. M. (1962). Limiting distribution of the maximum term in sequences of dependent random variables, ***Annals of Mathematical Statistics*, 33**, 894–908.

Billman, B. R., Antle, C. E., and Bain, L. J. (1972). Statistical inference from censored Weibull samples, ***Technometrics*, 14**, 831–840.

Black, C. M., Durham, S. D., and Padgett, W. J. (1990). Parameter estimation for a new distribution for the strength of brittle fibers: A simulation study, ***Communications in Statistics—Simulation and Computation*, 19**, 809–825.

Blom, G. (1958). ***Statistical Estimates and Transformed Beta-Variables***, Stockholm: Almquist and Wiksell.

Borgman, L. E. (1961). The frequency distribution of near extreme, ***Journal of Geophysical Research*, 66**, 3295–3307.

Bortkiewicz, L., von (1922). Variationsbreite und mittlerer Fehler, ***Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 21**, 3–11.

Buishand, T. A. (1989). Statistics of extremes in climatology, ***Statktica Neerlandica*, 43**, 1–30.

Burton, P. W., and Makropoulos, K. C. (1985). Seismic risk of circum-Pacific earthquakes: II. Extreme values using Gumbel's third distribution and the relationship with strain energy release, ***Pure and Applied Geophysics*, 123**, 849–866.

Canfield, R. V. (1975). The Type-I extreme-value distribution in reliability, ***IEEE Transactions on Reliability*, 24**, 229–236.

Canfield, R. V., and Borgman, L. E. (1975). Some distributions of time to failure for reliability applications, ***Technometrics*, 17**, 263–268.

Castillo, E. (1988). ***Extreme Value Theory in Engineering***, San Diego, CA: Academic Press.

Chan, L. K., Cheng, S. W. H., and Mead, E. R. (1972). An optimum t-test for the scale parameter of an extreme-value distribution, ***Naval Research Logistics Quarterly*, 19**, 715–723.

Chan, L. K., and Kabir, A. B. M. L. (1969). Optimum quantiles for the linear estimation of the parameters of the extreme-value distribution in complete and censored samples, ***Naval Research Logistics Quarterly*, 16**, 381–404.

Chan, L. K., and Mead, E. R. (1971a). Linear estimation of the parameters of the extreme-value distribution based on suitably chosen order statistics, ***IEEE Transactions on Reliability*, R-20**, 74–83.

Chan, L. K., and Mead, E. R. (1971b). Tables to facilitate calculation of an asymptotically optimal t-test for equality of location parameters of a certain extreme-value distributions, ***IEEE Transactions on Reliability*, R-20**, 235–243.

Chandra, M., Singpurwalla, N. D., and Stephens, M. A. **(1981).** Kolmogorov statistics for tests of fit for the extreme-value and Weibull distribution, ***Journal of the American Statbtical Association, 76,*** **729–731.**

Chao, A., and Hwang, S.-J. **(1986).** Comparison of confidence intervals for the parameters of the Weibull and extreme value distributions, ***IEEE Transactions on Reliability,*** **35, 111–113.**

Cheng, R. C. H., and Iles, T. C. **(1983).** Confidence bands for cumulative distribution functions of continuous random variables, ***Technometrics,*** **25, 77–86.**

Cheng, R. C. H., and Iles, T. C. **(1988).** One-sided confidence bands for cumulative distribution functions, ***Technometrics,*** **30, 155–159.**

Chiou, P. **(1988).** Shrinkage estimation of scale parameter of the extreme-value distribution, ***IEEE Transactions on Reliability, 37,*** **370–374.**

Chowdhury, J. U., Stedinger, J. R., and Lu, L.-H. **(1991).** Goodness-of-fit tests for regional generalized extreme value flood distributions, ***Water Resources Research,*** **27, 1765–1776.**

Christopeit, N. **(1994).** Estimating parameters of an extreme value distribution by the method of moments, ***Journal of Statistical Planning and Inference,*** **41, 173–186.**

Clough, D. J., and Kotz, S. **(1965).** Extreme value distributions with a special queueing model application, ***CORS Journal,*** **3, 96–109.**

Cockrum, M. B., Larson, R. K., and Taylor, R. W. **(1990).** Distribution modeling and simulation studies in product flammability testing, ***ASA Proceedings of Business and Economic Statistics Section,*** **387–391.**

Coelho, D. P., and Gil, T. P. **(1963).** Studies on extreme double exponential distribution. I. The location parameter, ***Revista da Faculdade de Ciências de Lisboa,*** **10, 37–46.**

Cohen, J. P. **(1986).** Large sample theory for fitting an approximating Gumbel model to maxima, ***Sankhyā, Series A,*** **48, 372–392.**

Cohen, J. P. **(1988).** Fitting extreme value distributions to maxima, ***Sankhyii, Series A,*** **50, 74–97.**

D'Agostino, R. B., and Stephens, M. A. **(eds.) (1986).** ***Goodness-of-fit Techniques,*** New York: Dekker.

Daniels, H. E. **(1942).** A property of the distribution of extremes, ***Biometrika,*** **32, 194–195.**

David, H. A. **(1981).** ***Order Statistics,*** 2d ed., New York: Wiley.

Davidovich, M. I. **(1992).** On convergence of the Weibull-Gnedenko distribution to the extreme value distribution, ***Vestnik Akademii Nauk Belaruss, Ser. Mat.-Fiz.,*** No. 1, Minsk, **103–106.**

Davison, A. C. **(1986).** Approximate predictive likelihood, ***Biometrika,*** 73, **323–332.**

Davison, A. C., and Smith, R. L. **(1990).** Models for exceedances over high thresholds (with comments), ***Journal of the Royal Statistical Society, Series B,*** **52, 393–442.**

de Haan, L. **(1970).** ***On Regular Variation and Its Application to the Weak Convergence of Sample Extremes,*** Amsterdam: Mathematical Centre Tract **32,** Mathematisch Centrum.

de Haan, L. **(1976).** Sample extremes: An elementary introduction, ***Statbtica Neerlandica,*** **30. 161–172.**

Dekkers, A. L. M., and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation, *Annals of Statistics*, **17**, 1795–1832.

Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics*, **17**, 1833–1855.

DiCiccio, T. J. (1987). Approximate inference for the generalized gamma distribution, *Technometrics*, **29**, 32–39.

Divakar, S. (1975). A note on Karlin's admissibility result for extreme value density, *Journal of the Indian Statistical Association*, **13**, 67–69.

Dodd, E. L. (1923). The greatest and least variate under general laws of error, *Transactions of the American Mathematical Society*, **25**, 525–539.

Doganoksoy, N., and Schmee, J. (1991). Comparisons of approximate confidence intervals for the smallest extreme value distribution simple linear regression model under time censoring, *Communications in Statistics—Simulation and Computation*, **20**, 1085–1113.

Downton, F. (1966). Linear estimates of parameters in the extreme value distribution, *Technometrics*, **8**, 3–17.

Dubey, S. D. (1966). Characterization theorems for several distributions and their applications, *Journal of Industrial Mathematics*, **16**, 1–22.

Dubey, S. D. (1969). A new derivation of the logistic distribution, *Naval Research Logistics Quarterly*, **16**, 37–40.

Durrant, N. F. (1978). A nomogram for confidence limits on quantiles of the normal distribution with application to extreme value distributions, *Journal of Quality Technology*, **10**, 155–158.

Eldredge, G. G. (1957). Analysis of corrosion pitting by extreme value statistics and its application to oil well tubing caliper surveys, *Corrosion*, **13**, 51–76.

Engelhardt, M. (1975). On simple estimation of the parameters of the Weibull or extreme-value distribution, *Technomerrics*, **17**, 369–374.

Engelhardt, M., and Bain, L. J. (1973). Some complete and censored results for the Weibull or extreme-value distribution, *Technometrics*, **15**, 541–549.

Engelhardt, M., and Bain, L. J. (1977). Simplified statistical procedures for the Weibull or extreme-value distribution, *Technometrics*, **19**, 323–332.

Engelhardt, M., and Bain, L. J. (1979). Prediction limits and two-sample problems with complete or censored Weibull data, *Technometrics*, **21**, 233–238.

Engelhardt, M., and Bain, L. J. (1982). On prediction limits for samples from a Weibull or extreme-value distribution, *Technometrics*, **24**, 147–150.

Epstein, B. (1948). Application to the theory of extreme values in fracture problems, *Journal of the American Statistical Association*, **43**, 403–412.

Epstein, B. (1958). The exponential distribution and its role in life testing, *Industrial Quality Control*, **15**, 5–9.

Epstein, B. (1960). Elements of the theory of extreme values, *Technometrics*, **2**, 27–41.

Epstein, B., and Brooks, H. (1948). The theory of extreme values and its implications in the study of the dielectric strength of paper capacitors, *Journal of Applied Physics*, **19**, 544–550.

Escobar, L. A., and Meeker, W. Q., Jr. (1986). Elements of the Fisher information matrix for the smallest extreme value distribution and censored data, *Applied Statistics*. **35**. 80–86.

Fahmi, K. J., and Abbasi, J. N. Al. (1991). Application of a mixture distribution of extreme values to earthquake magnitudes in Iraq and conterminous regions, *Geophysical Journal of the Royal Astronomical Society, 107,* 209–217.

Fei, H., Kong, F., and Tang, Y. (1994). Estimations for two-parameter Weibull distributions and extreme-value distributions under multiple Type-I1 censoring, *Preprint.*

Fertig, K. W., and Mann, N. R. (1978). An accurate approximation to the sample distribution of the Studentized extreme value statistic, *Technometrics,* **22,** 83–97.

Fertig, K. W., Meyer, M. E., and Mann, N. R. (1980). On constructing prediction intervals for samples from a Weibull or extreme value distribution, *Technometrics,* **22,** 567–573.

Fisher, R. A. (1934). Two new properties of mathematical likelihood, *Proceedings of the Royal Society, Series A,* **144,** 285–307.

Fisher, R. A., and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society,* **24,** 180–190.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum, *Annales de la Société Polonaise de Mathématique, Cracovie,* *6,* 93–116.

Freimer, M., Kollia, G., Mudholkar, G. S., and Lin, C. T. (1989). Extremes, extreme spacings and outliers in the Tukey and Weibull families, *Communications in Statistics—Theory and Methods,* **18,** 4261–4274.

Frenkel, J. I., and Kontorova, T. A. (1943). A statistical theory of the brittle strength of crystals, *Journal of Physics, USSR, 7,* 108–114.

Fuller, W. E. (1914). Flood flows, *Transactions of the American Society of Civil Engineers,* **77,** 564.

Fung, K. Y., and Paul, S. R. (1985). Comparisons of outlier detection procedures in Weibull or extreme-value distribution, *Communications in Statistics—Simulation and Computation,* **14,** 895–917.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics,* New York: Wiley.

Galambos, J. (1981). Failure time distributions: Estimates and asymptotic results, In *Statistical Distributions in Scientific Work, Volume 5* (eds., C. Taillie, G. P. Patil, and B. A. Baldessari), Dordrecht: Reidel, pp. 309–317.

Galambos, J. (1982). A statistical test for extreme value distributions, In *Nonparametric Statistical Inference* (eds., B. V. Gnedenko, M. L. Puri, and I. Vincze), Amsterdam: North-Holland, pp. 221–230.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics,* 2d ed., Malabar, FL: Krieger.

Garg, M. L., Raja Rao, B., and Redmond, C. K. (1970). Maximum-likelihood estimation of the parameters of the Gompertz survival function, *Applied Statistics, 19,* 152–160.

Geffroy, J. (1958). Contribution a la théorie des valeurs extrêmes, *Publications de l'Institut de Statistique de l'Université de Paris, 7,* No. 3/4, 37–121.

Geffroy, J. (1959). Contribution a la théorie des valeurs extrêmes, II, *Publications de l'Institut de Statistique de l'Université de Paris, 8,* No. 1. 3–65.

Gerisch, W., Struck, W., and Wilke, B. (1991). One-sided Monte Carlo tolerance limit factors for the exact extreme-value distributions from a normal parent distributional, *Computational Statistics Quarterly, 6*, **241–261.**

Gnedenko, B. **(1943).** Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics, 44,* **423–453.**

Goka, T. (1993). Application of extreme-value theory to reliability physics of electronic parts and to on-orbit single event phenomena, Paper presented at the *Conference on Extreme Value Theory and Its Applications,* May **2–7, 1993,** National Institute of Standards, Gaithersburg, MD.

Goldstein, N. **(1963).** Random numbers from the extreme value distributions, *Publications de l'Institut de Statistique de l'Université de Paris,* **12, 137–158.**

Gompertz, B. **(1825).** On the nature of the function expressive of the law of human mortality, *Philosophical Transactions of the Royal Society of London, Series A*, **115, 513–580.**

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. **(1979).** Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form, *Water Resources Research,* **15, 1049–1054.**

Greis, N. P., and Wood, E. F. **(1981).** Regional flood frequency estimation and network design, *Water Resources Research,* **17, 1167–1177.**

Griffith, A. A. (1920). The phenomena of rupture and flow in solids, *Philosophical Transactions of the Royal Society of London, Series A*, **221, 163–198.**

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques, *Annales de l'Institut Henri Poincaré, 4,* **115–158.**

Gumbel, E. J. (1937a). Les intervalles extremes entre les émissions radioactives, *Journal de Physique et de Radium, 8,* **446–452.**

Gumbel, E. J. (1937b). La durée extrême de la vie humaine, *Actualitis Scientifiques et Industrielles*, Paris: Hermann et Cie.

Gumbel, E. J. **(1941).** The return period of flood flows, *Annals of Mathematical Statistics,* **12, 163–190.**

Gumbel, E. J. **(1944).** On the plotting of flood discharges, *Transactions of the American Geophysical Union,* **25, 699–719.**

Gumbel, E. J. **(1945).** Floods estimated by probability methods, *Engineering News-Record, 134,* 97–101.

Gumbel, E. J. **(1947).** The distribution of the range, *Annals of Mathematical Statistics,* **18, 384–412.**

Gumbel, E. J. (1949a). *The Statistical Forecast of Floods,* Bulletin No. **15, 1–21,** Ohio Water Resources Board, Columbus.

Gumbel, E. J. (1949b). Probability tables for the range, *Biometrika,* 36, **142–148.**

Gumbel, E. J. **(1953).** Introduction, In *Probability Tables for the Analysis of Extreme-Value Data,* National Bureau of Standards, Applied Mathematics Series, vol. **22,** Washington, DC: GPO.

Gumbel, E. J. **(1954).** *Statistical Theory of Extreme Values and Some Practical Applications,* National Bureau of Standards, Applied Mathematics Series, vol. 33, Washington, DC: GOP.

Gumbel, E. J. (1958). *Statistics of Extremes,* New York: Columbia University Press.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions, *Publications de l'Institut de Statistique de l'Université de Paris, 9,* 171–173.

Gumbel, E. J. (1961). Sommes et differences de valeurs extrêmes indépendentes, *Comptes Rendus de l'Académie des Sciences, Paris,* **253,** 2838–2839.

Gumbel, E. J. (1962a). Statistical estimation of the endurance limit—An application of extreme-value theory, In *Contributions to Order Statistics* (eds., *A. E.* Sarhan and B. G. Greenberg), New York: Wiley, pp. 406–431.

Gumbel, E. J. (1962b). Statistical theory of extreme values (main results), In *Contributions to Order Statistics* (eds., A. *E.* Sarhan and B. G. Greenberg), New York: Wiley, ch. 6.

Gumbel, E. J. (1962c). Produits et quotients de deux plus grandes valeurs indépendantes, *Comptes Rendus de l'Académie des Sciences, Paris,* **254,** 2132–2134.

Gumbel, E. J. (1962d). Produits et quotients de deux plus petites valeurs indépendantes, *Publications de l'Institut de Statistique de l'Université de Paris,* **11,** 191–193.

Gumbel, E. J. (1965). A quick estimation of the parameters in Fréchet's distribution, *Review of the International Statistical Institute,* **33,** 349–363.

Gumbel, E. J., and Mustafi, C. K. (1966) Comments to: Edward C. Posner, "The application of extreme value theory to error free communication," *Technometrics,* **8,** 363–366.

Gumbel, E. J., and Pickands, J. (1967). Probability tables for the extremal quotient, *Annals of Mathematical Statistics,* **38,** 1541–1551.

Hald, A. (1952). *Statistical Theory with Engineering Applications,* New York: Wiley.

Hall, M. J. (1992). Problems of handling messy field data for engineering decision-making: More on flood frequency analysis, *The Mathematical Scientist,* **17,** 78–88.

Harris, B. (1970). An application of extreme value theory to reliability theory, *Annals of Mathematical Statistics,* **41,** 1456–1465.

Harter, H. L. (1970). *Order Statistics and Their Use in Testing and Estimation,* vol. 2, Washington, DC: GPO.

Harter, H. L. (1978). A bibliography of extreme-value theory, *International Statistical Review,* **46,** 279–306.

Harter, H. L., and Moore, A. H. (1967). A note on estimation from a type 1 extreme-value distribution, *Technometrics,* 9, 325–331.

Harter, H. L., and Moore, A. H. (1968a). Maximum likelihood estimation, from doubly censored samples, of the parameters of The first asymptotic distribution of extreme values, *Journal of the American Statistical Association,* **63,** 889–901.

Harter, H. L., and Moore, A. H. (1968b). Conditional maximum-likelihood estimators, from singly censored samples, of the scale parameters of Type-II extreme-value distributions, *Technometrics,* 10, 349–359.

Hasofer, A. M., and Wang, Z. (1992). A test for extreme value domain of attraction, *Journal of the American Statistical Association, 87,* 171–177.

Hassanein, K. M. (1965). *Estimation of the parameters of the extreme value distribution by order statistics,* National Bureau of Standards, Institute of Applied Technology, Project No. 2776-M.

Hassanein, K. M. (1968). Analysis of extreme-value data by sample quantiles for very large samples, ***Journal of the American Statistical Association, 63,*** 877–888.

Hassanein, K. M. (1969). Estimation of the parameters of the extreme value distribution by use of two or three order statistics, ***Biometrika,*** 56, 429–436.

Hassanein, K. M. (1972). Simultaneous estimation of the parameters of the extreme value distribution by sample quantiles, ***Technometrics,*** **14,** 63–70.

Hassanein, K. M., and Saleh, A. K. Md. E. (1992). Testing equality of locations and quantiles of several extreme-value distributions by use of few order statistics of samples from extreme-value and Weibull distributions, In ***Order Statistics and Nonparametrics: Theory and Applications*** (eds., *P. K.* Sen and I. A. Salama), Amsterdam: North-Holland, pp. 115–132.

Hassanein, K. M., Saleh, A. K. Md. E., and Brown, E. F. (1984). Quantile estimates in complete and censored samples from extreme-value and Weibull distributions, ***IEEE Transactions on Reliability, 33,*** 370–373.

Hassanein, K. M., Saleh, A. K. Md. E., and Brown, E. F. (1986). Estimation and testing of quantiles of the extreme-value distribution, ***Journal of Statistical Planning and Inference,*** 14, 389–400.

Henery, R. J. (1984). An extreme-value model for predicting the results of horse races, ***Applied Statistics,*** **33,** 125–133.

Hooda, B. K., Singh, N. P., and Singh, U. (1990). Estimates of the parameters of Type-11 extreme value distribution from censored samples, ***Communications in Statistics—Theory and Methods, 19,*** 3093–3110.

Hooda, B. K., Singh, N. P., and Singh, U. (1991). Estimation of Gumbel distribution parameters of m-th maxima from doubly censored sample, ***Statktica,*** 51,339–352.

Hopke, P. K., and Paatero, P. (1993). Extreme value estimation applied to aerosol size distributions and related environmental problems, Paper presented at the ***Conference on Extreme Value Theory and Its Applications,*** May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Hosking, J. R. M. (1984). Testing whether the shape parameter is zero in the generalized extreme-value distribution, ***Biometrika,*** 71, 367–374.

Hosking, J. R. M. (1985). Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution, ***Applied Statistics,*** **34,** 301–310.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, ***Technometrics,*** 27, 251–261.

Huesler, J., and Schuepbach, M. (1986). On simple block estimators for the parameters of the extreme-value distribution, ***Communications in Statistics—Simulation and Computation, 15,*** 61–76.

Huesler, J., and Tiago de Oliveira, J. (1988). The usage of the largest observations for parameter and quantile estimation for the Gumbel distribution: An efficiency analysis, ***Publications of Institute of Statistics, University of Paris IV,*** **33(1),** 41–56.

Irwin, J. 0 . (1942). The distribution of the logarithm of survival times when the true law is exponential, ***Journal of Hygiene, Cambridge, 42,*** 328–333.

Jain, D., and Singh, V. P. (1987). Estimating parameters of EV1 distribution for flood frequency analysis, ***Water Resources Research,*** 23, 59–71.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, ***Quarterly Journal of the Royal Meteorological Society,*** **81,** 158–171.

Jenkinson, A. F. (1969). Statistics of Extremes, ***Technical Note No. 98,*** Geneva: World Meteorological Office.

Jeruchim, M. C. (1976). On the estimation of error probability using generalized extreme value theory, ***IEEE Transactions on Information Theory,*** **IT-22,** 108–110.

Joe, H. (1987). Estimation of quantiles of the maximum of *N* observations, ***Biometrika,*** **74,** 347–354.

Johns, M. V., Jr., and Lieberman, G. J. (1966). An exact asymptotically efficient confidence bound for reliability in the case of the Weibull distribution, ***Technometrics,*** **8,** 135–175.

Juncosa, M. L. (1949). On the distribution of the minimum in a sequence of mutually independent random variables, ***Duke Mathematical Journal,*** **16,** 609–618.

Kaminsky, K. S. (1982). A characterization of the Gompertz distribution and a discrete analog, ***Preprint.***

Kanda, J. (1993). Application of an empirical extreme value distribution to load models, Paper presented at the ***Conference on Extreme Value Theory and Its Applications,*** May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Kao, J. H. K. (1958). Computer methods for estimating Weibull parameters in reliability studies, IRE ***Transactions on Reliability and Quality Control,*** **13,** 15–22.

Keating, J. P. (1984). A note on estimation of percentiles and reliability in the extreme-value distribution, ***Statistics & Probability Letters,*** **2,** 143–146.

Kimball, B. F. (1946a). Sufficient statistical estimation functions for the parameters of the distribution of maximum values, ***Annals of Mathematical Statistics,*** **17,** 299–309.

Kimball, B. F. (1946b). Assignment of frequencies to a completely ordered set of sample data, ***Transactions of the American Geophysical Union,*** **27,** 843–846. (Discussion in 28 [1947], 951–953.)

Kimball, B. F. (1949). An approximation to the sampling variance of an estimated maximum value of given frequency based on fit of doubly exponential distribution of maximum values, ***Annals of Mathematical Statistics,*** **20,** 110–113.

Kimball, B. F. (1955). Practical applications of the theory of extreme values, ***Journal of the American Statistical Association,*** **50,** 517–528. (Correction: 50, 1332.)

Kimball, B. F. (1956). The bias in certain estimates of the extreme-value distribution, ***Annals of Mathematical Statistics,*** **27,** 758–767.

Kimball, B. F. (1960). On the choice of plotting positions on probability paper, ***Journal of the American Statistical Association,*** **55,** 546–560.

Kimber, A. C. (1985). Tests for the exponential, Weibull and Gumbel distributions based on the standardized probability plot, ***Biometrika,*** **72,** 661–663.

King, J. R. (1959). Summary of extreme-value theory and its relation to reliability analysis, ***Proceedings of the 12th Annual Conference of the American Society for Quality Control,*** **13,** 163–167.

Kinnison, R. (1989). Correlation coefficient goodness-of-fit test for the extreme-value distribution. ***The American Statistician.*** **43.** 98–100.

Kinnison, R. (1990). Replies to "Comment on 'Correlation coefficient goodness-of-fit test for the extreme-value distribution,'" *The American Statistician, 44,* 61; *44,* 260.

Kotz, S., and Johnson, N. L. (eds.) (1991). *Breakthroughs in Statistics* vol. 1, New York: Springer-Verlag.

Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resources Research,* **15,** 1055–1064.

Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1980). Quantile estimation with more or less floodlike distributions, *Water Resources Research*, 16, 547–555.

Lawless, J. F. (1973). On the estimation of safe life when the underlying life distribution is Weibull, *Technometrics,* **15,** 857–865.

Lawless, J. F. (1975). Construction of tolerance bounds for the extreme-value and Weibull distributions, *Technometrics,* **17,** 255–262.

Lawless, J. F. (1978). Confidence interval estimation for the Weibull and extreme value distributions (with discussion), *Technometrics,* 20, 355–368.

Lawless, J. F. (1980). Inference in the generalized gamma and log-gamma distribution, *Technometrics,* 22, 67–82.

Lawless, J. F. (1982). *Statistical Models & Methods for Lifetime Data,* New York: Wiley.

Lawless, J. F., and Mann, N. R. (1976). Tests for homogeneity of extreme value scale parameters, *Communications in Statistics—Theory and Methods,* 1, 389–405.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes,* New York: Springer:Verlag.

Lehman, E. H. (1963). Shapes, moments and estimators of the Weibull distribution, *Transactions of the IEEE on Reliability,* 12, 32–38.

Lieblein, J. (1953). On the exact evaluation of the variances and covariances of order statistics in samples from the extreme-value distribution, *Annals of Mathematical Statistics,* 24, 282–287.

Lieblein, J. (1954). A *new method of analyzing extreme-value data,* National Advisory Committee on Aeronautics, Technical Note No. 3053, Washington, DC.

Lieblein, J. (1962). Extreme-value distribution, In *Contributions to Order Statistics* (eds., A. *E.* Sarhan and B. G. Greenberg), New York: Wiley, pp. 397–406.

Lieblein, J., and Salzer, H. E. (1957). Table of the first moment of ranked extremes, *Journal of Research of the National Bureau of Standards, 59,* 203–206.

Lieblein, J., and Zelen, M. (1956). Statistical investigation of the fatigue life of deep-grove ball bearings, *Journal of Research of the National Bureau of Standards,* **57,** 273–316.

Lockhart, R. A., O'Reilly, F. J., and Stephens, M. A. (1986a). Tests of fit based on normalized spacings, *Journal of the Royal Statistical Society, Series B,* 48,344–352.

Lockhart, R. A., O'Reilly, F. J., and Stephens, M. A. (1986b). Tests for the extreme value and Weibull distributions based on normalized spacings, *Naval Research Logistics Quarterly, 33,* 413–421.

Lockhart, R. A., and Spinelli, J. J. (1990). Comment on "Correlation coefficient goodness-of-fit test for the extreme-value distribution," *The American Statistician,* 44, 259–260.

Looney, S. W. (1990). Comment on "Correlation coefficient goodness-of-fit test for the extreme-value distribution," *The American Statistician, 44,* 61.

Macleod, A. J. (1989). Comment on "Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution," *Applied Statistics,* 38,198–199.

Mahmoud, M. W., and Ragab, A. (1975). On order statistics in samples drawn from the extreme value distributions, *Mathematische Operationsforschung und Statistik, Series Statistics, 6,* 809–816.

Mann, N. R. (1967a). ***Results on Location and Scale Parameter Estimation with Application to the Extreme-Value Distribution,*** Report ARL67-0023, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Dayton, OH.

Mann, N. R. (1967b). Tables for obtaining the best linear invariant estimates of the parameters of the Weibull distribution, *Technometrics,* 9, 629–645.

Mann, N. R. (1968a). Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions, *Technometrics,* **10,** 231–256.

Mann, N. R. (1968b). ***Results on Statistical Estimation and Hypothesis Testing with Application to the Weibull and Extreme Value Distributions,*** Report ARL 68-0068, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Dayton, OH.

Mann, N. R. (1969). Optimum estimators for linear functions of location and scale parameters, *Annals of Mathematical Statistics,* **40,** 2149–2155.

Mann, N. R. (1976). Warranty periods for production lots based on fatigue-test data, *Engineering Fracture Mechanics,* 8, 123–130.

Mann, N. R. (1982). Optimal outlier tests for a Weibull model—to identify process changes or predict failure times, *Studies in the Management Sciences,* **19,** 261–270.

Mann, N. R., and Fertig, K. W. (1973). Tables for obtaining Weibull confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme-value distribution, *Technometrics,* **15,** 87–102.

Mann, N. R., and Fertig, K. W. (1975). Simplified efficient point and interval estimators for Weibull parameters, *Technometrics,* 17, 361–368.

Mann, N. R., and Fertig, K. W. (1977). Efficient unbiased quantile estimators for moderate-size complete samples from extreme-value and Weibull distributions; Confidence bounds and tolerance and prediction intervals, *Technometrics,* **19,** 87–94.

Mann, N. R., Fertig, K. W., and Scheuer, E. M. (1971). Confidence and tolerance bounds and a new goodness of fit test for the two-parameter Weibull or extreme value distribution with tables for censored samples of size 3(1)25, *Aerospace Research Laboratories Report No. 71-0077,* Wright-Patterson Air Force Base, Dayton, OH.

Mann, N. R., and Saunders, S. C. (1969). On evaluation of warranty assurance when life has a Weibull distribution, *Biometrika,* **56,** 615–625.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data,* New York: Wiley.

Mann, N. R., Scheuer, E. M., and Fertig, K. W. **(1973). A** new goodness-of-fit test for the two-parameter Weibull or extreme-value distribution with unknown parameters, ***Communications in Statistics, 2,*** **383–400.**

Mann, N. R., and Singpurwalla, N. D. **(1982).** Extreme-value distributions, In ***Encyclopedia of Statistical Sciences,*** vol. **2** (eds., S. Kotz, N. L. Johnson, and C. B. Read), New York: Wiley, pp. **606–613.**

Maritz, J. S., and Munro, **A.** H. **(1967).** On the use of the generalized extreme-value distribution in estimating extreme percentiles, ***Biometrics,*** 23, **79–103.**

Massonie, J. P. (1966). Estimation de I'exposant d'une fonction de distribution tronquée, ***Comptes Rendus de l'Académie des Sciences, Paris, 262,*** **350–352.**

McCool, J. I. (1965). The construction of good linear unbiased estimates from the best linear estimates for a smaller sample size, ***Technometrics,*** 7, **543–552.**

McCord, J. R. **(1964).** On asymptotic moments of extreme statistics, ***Annals of Mathematical Statistics,*** **35, 1738–1745.**

McLaren, C. G., and Lockhart, R. A. **(1987).** On the asymptotic efficiency of certain correlation tests of fit, ***Canadian Journal of Statistics,*** **15, 159–167.**

Meeker, W. Q., and Nelson, W. **(1975).** Optimum accelerated life-tests for the Weibull and extreme value distributions, ***IEEE Transactions on Reliability, 24,*** **321–332.**

Metcalfe, A. V., and Mawdsley, J. A. **(1981).** Estimation of extreme low flows for pumped storage reservoir design, ***Water Resources Research,*** **17, 1715–1721.**

Michael, J. R. (1983). The stabilized probability plot, ***Biometrika,*** 70, **11–17.**

Mihram, **G. A. (1975). A** generalized extreme-value density, ***South African Statistical Journal, 9,*** **153–162.**

Miller, I., and Freund, J. **(1965).** ***Probability and Statistics for Engineers,*** Englewood Cliffs, NJ: Prentice-Hall.

Mises, R., von (1923). Über die Variationsbreite einer Beobachtungsreihe, ***Sitzungsberichte der Berliner Mathematischen Gesellschaft, 22,*** **3–8.**

Mises, R., von **(1936).** La distribution de la grande de *n* valeurs, ***Reciew Mathematique Union Interbalcanique,*** **1, 141–160.** Reproduced in Selected Papers of Richard von Mises, II, ***American Mathematical Society*** (1964), pp. **271–294,** Providence, RI.

Moore, A. H., and Harter, H. L. (1966). Point and interval estimation, from one order statistic, of the location parameter of an extreme value distribution with known scale parameter, and of the scale parameter of a Weibull distribution with the known shape parameters, ***Transactions of the IEEE on Reliability,*** **15, 120–126.**

Moore, **A.** H., and Harter, H. L. **(1967).** One-order-statistic conditional estimators of shape parameters of limited and Pareto distributions and scale parameters of Type II asymptotic distributions of smallest and largest values, ***Transactions of the IEEE on Reliability, 16,*** **100–103.**

Mustafi, **C. K.** (1963). Estimation of parameters of the extreme value distribution with limited type of primary probability distribution, ***Bulletin of the Calcutta Statistical Association, 12,*** **47–54.**

Nagaraja, H. N. **(1982).** Record values and extreme value distributions, ***Journal of Applied Probability,*** **19, 233–239.**

Nagaraja, H. N. (1984). Asymptotic linear prediction of extreme order statistics, *Annals of the Institute of Statistical Mathematics,* **36,** 289–299.

Nagaraja, H. N. (1988). Record values and related statistics—A review, *Communications in Statistics—Theory and Methods,* **17,** 2223–2238.

Nelson, W., and Meeker, W. Q. (1978). Theory for optimum accelerated censored life tests for Weibull and extreme value distributions, *Technometrics,* 20, 171–178.

Newman, F. W. (1892). *The Higher Trigonometry, Superrationals of Second Order,* Cambridge: Cambridge University Press.

Nissan, E. (1988). Extreme value distribution in estimation of insurance premiums, *ASA Proceedings of Business and Economic Statistics Section,* 562–566.

Nordquist, J. M. (1945). Theory of largest values, applied to earthquake magnitudes, *Transactions of the American Geophysical Union, 26,* 29–31.

Ogawa, J. (1951). Contributions to the theory of systematic statistics, I, *Osaka Mathematical Journal,* **3,** 175–213.

Ogawa, J. (1952). Contributions to the theory of systematic statistics, II, *Osaka Mathematical Journal, 4,* 41–61.

Okubo, T., and Narita, N. (1980). On the distribution of extreme winds expected in Japan, *National Bureau of Standards Special Publication,* **560-1,** 12 pp.

Oller, J. M. (1987). Information metric for extreme value and logistic probability distributions, *Sankhyā, Series A, 49,* 17–23.

Owen, D. B. (1962). *Handbook of Statistical Tables,* Reading, MA: Addison-Wesley.

Öztürk, A. (1986). On the W test for the extreme value distribution, *Biometrika,* **73,** 738–740.

dztiirk, A. (1987). Weighted least squares estimation of location and scale parameters, *American Journal of Mathematical and Management Sciences,* **7,** 113–129.

Öztürk, A., and Korukoğlu, S. (1988). A new test for the extreme value distribution, *Communications in Statistics—Simulation and Computation,* **17,** 1375–1393.

Pandey, M., and Upadhyay, S. K. (1986). Approximate prediction limit for Weibull failure based on preliminary test estimator, *Communications in Statistics—Theory and Methods,* **15,** 241–250.

Pannullo, J. E., Li, D., and Haimes, Y. Y. (1993). Posterior analysis in assessing risk of extreme events: A conjugate family approach, *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics,* vol. 1, pp. 477–482.

Paul, S. R., and Fung, K. Y. (1986). Critical values for Dixon type test statistics for testing outliers in Weibull or extreme value distributions, *Communications in Statistics—simulation and Computation, 15,* 277–283.

Pericchi, L. P., and Rodriguez-Iturbe, I. (1985). On the statistical analysis of floods, In A *Celebration of Statistics: The ISI Centenary Volume* (eds., A. C. Atkinson and S. E. Fienberg), New York: Springer-Verlag, pp. 511–541.

Phien, H. N. (1991). Maximum likelihood estimation for the Gumbel distribution from censored samples, In *The Frontiers of Statistical Computation, Simulation, and Modeling,* vol. 1 (eds., P. R. Nelson, E. J. Dudewicz, A. Öztürk, and E. C. van der Meulen), Syracuse, NY: American Sciences Press, pp. 271–287.

Posner, E. C. (1965). The application of extreme-value theory to error-free communication. *Technometrics.* 7. 517–529.

Potter, W. D. (1949). *Normalcy tests of precipitation and frequency studies of runoff on small watersheds,* U.S. Department of Agriculture Technical Bulletin, No. 985, Washington, DC: GPO.

Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation, *Biometrika,* 61, 539–544.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika,* 67, 723–724.

Prescott, P., and Walden, A. T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples, *Journal of Statistical Computation and Simulation,* 16, 241–250.

Press, H. (1949). *The application of the statistical theory of extreme value to gust-load problems,* National Advisory Committee on Aeronautics, Technical Note No. 1926, Washington, DC.

Provasi, C. (1987). Exact and approximate means and covariances of order statistics of the standardized extreme value distribution (I type), *Rivista Di Statistica Applicata, 20,* 287–295. (In Italian.)

Pyke, R. (1965). Spacings (with discussion), *Journal of the Royal Statistical Society, Series B,* 27, 395–449.

Rajan, K. (1993). Extreme value theory and its applications in microstructural sciences, Paper presented at the *Conference on Extreme Value Theory and Its Applications,* May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Rantz, S. F., and Riggs, H. C. (1949). *Magnitude and frequency of floods in the Columbia River Basin,* U.S. Geological Survey, Water Supply Paper 1080,317–476.

Rasheed, H., Aldabagh, A. S., and Ramamoorthy, M. V. (1983). Rainfall analysis by power transformation, *Journal of Climate and Applied Meteorology,* 22, 1411–1415.

Reiss, R. D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics,* Berlin: Springer-Verlag.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes,* New York: Springer-Verlag.

Revfeim, K. J. A. (1984a). The cumulants of an extended family of type I extreme value distributions, *Sankhyā, Series B,* 46, 281–284.

Revfeim, K. J. A. (1984b). Generating mechanisms of, and parameter estimators for, the extreme value distribution, *Australian Journal of Statistics,* 26, 151–159.

Revfeim, K. J. A. (1984c). The analysis of maximum wind gusts by direction, *New Zealand Journal of Science,* 27, 365–367.

Revfeim, K. J. A., and Hessell, J. W. D. (1984). More realistic distributions for extreme wind gusts, *Quarterly Journal of the Royal Meteorological Society,* 110, 505–514.

Robinson, J. A. (1983). Bootstrap confidence intervals in location-scale models with progressive censoring, *Technometrics,* 25, 179–187.

Roldan-Canas, J., Garcia-Guzman, A., and Losada-Villasante, A. (1982). A stochastic model for wind occurrence, *Journal of Applied Meteorology,* 21, 740–744.

Rossi, F. (1986). Reply to "Comment on 'Two-component extreme value distribution for flood frequency analysis,'" *Water Resources Research,* 22, 267–269.

Rossi, F., Fiorentino, M., and Versace, P. (1986). Two-component extreme value distribution for flood frequency analysis, ***Water Resources Research,*** **22.**

Scarf, P. A. (1992). Estimation for a four parameter generalized extreme value distribution, ***Communications in Statistics—Theory and Methods,*** **21,** 2185–2201.

Scarf, P. A., and Laycock, P. J. (1993). Applications of extreme value theory in corrosion engineering, Paper presented at the ***Conference on Extreme Value Theory and Its Applications,*** May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Schneider, H., and Weissfeld, L. A. (1989). Interval estimation based on censored samples from the Weibull distribution, ***Journal of Quality Technology,*** **21,** 179–186.

Schuepbach, M., and Huesler, J. (1983). Simple estimators for the parameters of the extreme-value distribution based on censored data, ***Technometrics***, **25,** 189–192.

Sen, P. K. (1961). A note on the large-sample behaviour of extreme sample values from distribution with finite end-points, ***Bulletin of the Calcutta Statistical Association,*** **10,** 106–115.

Sethuraman, J. (1965). On a characterization of the three limiting types of the extreme, ***Sankhyii, Series A*****,** **27,** 357–364.

Shen, H. W., Bryson, M. C., and Ochoa, I. D. (1980). Effect of tail behaviour assumptions on flood predictions, ***Water Resources Research, 16,*** 361–364.

Shibata, T. (1993). Application of extreme value statistics to corrosion, Paper presented at the ***conference on Extreme Value Theory and Its Applications,*** May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Sibuya, M. (1967). On exponential and other random variable generators, ***Annals of the Institute of Statistical Mathematics,*** **13,** 231–237.

Simiu, E., Bietry, J., and Filliben, J. J. (1978). Sampling errors in estimation of extreme winds, ***Journal of the Structural Dicision of the National Bureau of Standards,*** **104,** 491–501.

Simiu, E., and Filliben, J. J. (1975). Statistical analysis of extreme winds, ***National Bureau of Standards Technical Note,*** **868,** 52 pp.

Simiu, E., and Filliben, J. J. (1976). Probability distributions of extreme wind speeds, ***Journal of the Structural Dicision of the National Bureau of Standards,*** **102,** 1861–1877.

Singh, N. P. (1987). Estimation of Gumbel distribution parameters by joint distribution of m extremes, ***Calcutta Statistical Association Bulletin,*** **36,** 101–104.

Singh, R. (1975). Some admissible estimators in extreme value densities, ***Canadian Mathematical Bulletin,*** **18,** 105–110.

Smirnov, N. V. (1952). Limit distributions for the terms of a variational series, ***Transactions of the American Mathematical Society,*** Sec. 1, No. **67** (English translation).

Smith, J. A. (1987). Estimating the upper tail of flood frequency distributions, ***Water Resources Research,*** **23,** 1657–1666.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases, ***Biometrika,*** **72,** 67–90.

Smith, R. L. (1988). Forecasting records by maximum likelihood, ***Journal of the American Statistical Association.*** **83.** 331–338.

Smith, R. L., and Weissman, I. (1985). Maximum likelihood estimation of the lower tail of a probability distribution, ***Journal of the Royal Statistical Society, Series B,*** 47, 285–298.

Smith, R. M. (1977). Some results on interval estimation for the two parameter Weibull or extreme-value distribution, ***Communications in Statistics—Theory and Methods, 2,*** 1311–1322.

Smith, R. M., and Bain, L. J. (1976). Correlation type of goodness-of-fit statistics with censored sampling, ***Communications in Statistics—Theory and Methods, 5,*** 119–132.

Smith, T. E. (1984). A choice probability characterization of generalized extreme value models, ***Applied Mathematics and Computation,*** **14,** 35–62.

Stephens, M. A. (1977). Goodness of fit for the extreme value distribution, ***Biometrika,*** 64, 583–588.

Stephens, M. A. (1986). Tests based on regression and correlation, In ***Goodness-of-fit Techniques*** (eds., R. B. D'Agostino and M. A. Stephens), New York: Dekker, Ch. 5.

Stone, G. C., and Rosen, H. (1984). Some graphical techniques for estimating Weibull confidence intervals, IEEE ***Transactions on Reliability, 33,*** 362–369.

Tawn, J. A. (1992). Estimating probabilities of extreme sea-levels, ***Applied Statistics,*** **41,** 77–93.

Taylor, J. M. (1983). Comparisons of certain distribution functions, ***Mathematische Operationsforschung und Statistik, Series Statistics, 14,*** 397–408.

Taylor, R. W. (1991). The development of burn time models to simulate product flammability testing, ***ASA Proceedings of Business and Economic Statistics Section,*** 339–344.

Teugels, J. L., and Beirlant, J. (1993). Extremes in insurance, Paper presented at the ***Conference on Extreme Value Theory and Its Applications,*** May 2–7, 1993, National Institute of Standards, Gaithersburg, MD.

Thom, H. C. S. (1954). Frequency of maximum wind speeds, ***Proceedings of the American Society of Civil Engineers, 80,*** 104–114.

Thoman, D. R., Bain, L. J., and Antle, C. E. (1970). Reliability and tolerance limits in the Weibull distribution, ***Technometrics,*** **12,** 363–371.

Tiago de Oliveira, J. (1963). Decision results for the parameters of the extreme value (Gumbel) distribution based on the mean and the standard deviation, ***Trabajos de Estadística, 14,*** 61–81.

Tiago de Oliveira, J. (1972). Statistics for Gumbel and Fréchet distributions, In ***Structural Safety and Reliability*** (ed., A. Freudenthal), New York: Pergamon, pp. 94–105.

Tiago de Oliveira, J. (1981). Statistical choice of univariate extreme models, In ***Statistical Distributions in Scientific Work,*** vol. 6 (eds., C. Taillie, G. P. Patil, and B. A. Baldessari), Dordrecht: Reidel, pp. 367–387.

Tiago de Oliveira, J. (1983). Gumbel distribution, In ***Encyclopedia of Statistical Sciences,*** vol. 3 (eds., S. Kotz, N. L. Johnson, and C. B. Read), New York: Wiley, pp. 552–558.

Tikhov, M. S. (1991). On reduction of test duration in the case of censoring, ***Theory of Probability and Its Applications,* 36,** 629–633.

Tiku, M. L., and Singh, M. (1981). Testing the two parameter Weibull distribution, ***Communications in Statistics—Theory and Methods,* 10,** 907–918.

Tiku, M. L., Tan, W. Y., and Balakrishnan, N. (1986). ***Robust Inference,*** New York: Dekker.

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population, ***Biometrika,* 17,** 364–387.

Tsujitani, M., Ohta, H., and Kase, S. (1979). A preliminary test of significance for the extreme-value distribution, ***Bulletin of University of Osaka Prefecture, Section A*, 27,** 187–193.

Tsujitani, M., Ohta, H., and Kase, S. (1980). Goodness-of-fit test for extreme-value distribution, ***IEEE Transactions on Reliability,* 29,** 151–153.

Uzgoren, N. T. (1954). The asymptotic development of the distribution of the extreme values of a sample, ***Studies in Mathematics and Mechanics,*** presented to **R.** von Mises, San Diego, CA: Academic Press, pp. 346–353.

van Montfort, M. A. J. (1970). On testing that the distribution of extreme is of type I when Type-I1 is the alternative, ***Journal of Hydrology,* 11,** 421–427.

Velz, C. J. (1947). Factors influencing self-purification and their relation to pollution abatement, ***Sewage Works Journal,* 19,** 629–644.

Viveros, **R.,** and Balakrishnan, N. (1994). Interval estimation of parameters of life from progressively censored data, ***Technometrics,* 36,** 84–91.

Vogel, **R.** M. (1986). The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses, ***Water Resources Research,* 22,** 587–590.

Wang, Q. J. (1990). Estimation of the GEV distribution from censored samples by method of partial probability weighted moments, ***Journal of Hydrology,* 120,** 103–114.

Wantz, J. W., and Sinclair, R. E. (1981). Distribution of extreme winds in the Bonneville power administration service area, ***Journal of Applied Meteorology,* 20,** 1400–1411.

Watabe, M., and Kitagawa, Y. (1980). Expectancy of maximum earthquake motions in Japan, ***National Bureau of Standards Special Publication,* 560–10, 8** pp.

Watson, G. S. (1954). Extreme values in samples from m-dependent stationary stochastic processes, ***Annals of Mathematical Statistics,* 25,** 798–800.

Weibull, W. (1939). The phenomenon of rupture in solids, ***Ingenior Vetenskaps Akademiens Handlingar,* 153,** 2.

Weibull, W. (1949). A statistical representation of fatigue failures in solids, ***Kunglig Tekniska Högskolans Handlingar,* 27.**

Weinstein, S. B. (1973). Theory and application of some classical and generalized asymptotic distributions of extreme values, ***IEEE Transactions on Information Theory,* IT-19,** 148–154.

Weiss, L. (1961). On the estimation of scale parameters, ***Naval Research Logistics Quarterly,*** 8, 245–256.

Weissman, I. **(1978).** Estimation of parameters and large quantiles based on the k largest observations, ***Journal of the American Statistical Association*, *73*, 812–815.**

Welsh, **A. H. (1986).** On the use of the empirical distribution and characteristic function to estimate parameters of regular variation, ***Australian Journal of Statistics*, 28, 173–181.**

White, **J. S. (1964).** Least-square unbiased censored linear estimation for the log Weibull (extreme value) distribution, ***Journal of Industrial Mathematics*, *14*, 21–60.**

White, J. S. **(1969).** The moments of log-Weibull order statistics, ***Technometrics*, 11, 373–386.**

Wiggins, J. B. **(1991).** Empirical tests of the bias and efficiency of the extreme-value variance estimator for common stocks, ***Journal of Business of the University of Chicago*, *64*, 417–432.**

Winer, P. **(1963).** The estimation of the parameters of the iterated exponential distribution from singly censored samples, ***Biometrics*, 19, 460–464.**

Young, **D. H.,** and Bakir, S. T. (1987). Bias correction for a generalized log-gamma regression model, ***Technometrics*, 29, 183–191.**

Zelenhasic, **E. (1970).** Theoretical probability distributions for flood peaks, ***Hydrology Paper No. 42*,** Colorado State University, Fort Collins.

CHAPTER 23

Logistic Distribution

1 HISTORICAL REMARKS AND GENESIS

An early reference to the use of the logistic function as a growth curve is by Verhulst (1838,1845). The use of the curve for economic demographic purposes has been very popular since the end of the nineteenth century. Many other applications of the logistic curve have also been found over the years. Pearl and Reed (1920, 1924), Pearl, Reed, and Kish (1940), and Schultz (1930) all applied the logistic model as a growth model in human populations as well as in some biological organisms. Schultz (1930) and Oliver (1964) used the logistic function to model agricultural production data. A number of authors including Pearl (1940), Berkson (1944, 1951, 1953), and Finney (1947,1952) discussed applications of the logistic function in bioassay problems. A few more interesting uses of the logistic function are in the analysis of survival data [Plackett (1959)], in the study of income distributions [Fisk (1961)], and in the modeling of the spread of an innovation [Oliver (1969)]. The logistic function and the logistic distribution have found several important applications in many different fields. A book length account of these is due to Balakrishnan (1992). In view of the encyclopedic nature of the treatment given in that book, we refrain from discussing this distribution in great detail and focus mainly on some significant developments concerning the distribution. Interested readers may refer to the volume by Balakrishnan (1992) for more details and relevant references.

The use of logistic function as a growth curve can be based on the differential equation

$$\frac{dF}{dx} = c[F(x) - A][B - F(x)] \tag{23.1}$$

where c, A, and B are constants with $c > 0$, $B > A$. In verbal form (23.1) can be interpreted as rate of growth = [excess over initial (asymptotic) value A] x [deficiency compared with final (asymptotic) value B].

Solution of (23.1) leads to

$$F(x) = \frac{B\,De^{x/c} + A}{De^{x/c} + 1}, \tag{23.2}$$

where D is a constant. As $x \to -\infty$, $F(x) \to A$; as $x \to \infty$, $F(x) \to$ B (if $D \neq 0$). The function $F(x)$ represents "growth" from a lower asymptote A to an upper asymptote B. To make $F(x)$ a proper cumulative distribution function, we put $A = 0$, B = 1; equation (23.2) then becomes

$$F(x) = \frac{De^{x/c}}{1 + De^{x/c}} = [1 + D^{-1}e^{-x/c}]^{-1} \tag{23.3}$$

which is of the logistic distribution form given in the next section with $c = \beta$ and $D = e^{-\alpha/\beta}$.

Equation (23.1) has been used as a model of autocatalysis. This is the name applied to a chemical reaction in which a catalyst M transforms a compound G into two compounds J and K, and J itself acts as a catalyst for the same reaction. If M,, G, = original concentrations of M, G, respectively, and y = common value of concentration of J and K at time t, then the law of mass action in this case is

$$\frac{dy}{dt} = c_1M_0(G_0 - y) + c_2y(G_0 - y) \tag{23.4}$$

(c_1 and c_2 are "catalytic constants" for the actions of E, J, respectively).

The right-hand side of (23.4) can be rearranged to read

$$c_2\left(y + \frac{c_1}{c_2}M_0\right)\left[\left(G_0 + \frac{c_1}{c_2}M_0\right) - \left(y + \frac{c_1}{c_2}M_0\right)\right] \tag{23.5}$$

which is the same form as (23.1) with $F(x)$, x replaced by $(y + c_1M_0/c_2)$, t, respectively, and with $c = c_2$, $A = 0$; B = G, $+ c_1M_0/c_2$.

The logistic distribution arises in a purely statistical manner as the limiting distribution (as $n \to \infty$) of the standardized midrange (average of largest and smallest sample values) of random samples of size n. This result was given by Gumbel (1944). Gumbel and Keeney (1950) [see also Gumbel and Pickands (1967)] showed that a logistic distribution is obtained as the limiting distribution of an appropriate multiple of the "extremal quotient," that is, (largest value)/(smallest value). (See Chapter 22.)

Talacko (1956) has shown that the logistic is the limiting distribution (as $r \to \infty$) of the standardized variable corresponding to $\Sigma_{j=1}^{r} j^{-1}X_j$, where the X_j's are independent random variables each having a type 1 extreme value distribution (see Chapter 22).

Dubey (1969) has shown that the logistic distribution can be obtained as a mixture of extreme value distributions

$$\Pr[X \le x|\eta] = 1 - \exp\left[-\eta\beta \exp\left\{-\frac{x-\alpha}{\beta}\right\}\right], \qquad \eta, \beta > 0$$

[obtained by putting $\theta = \alpha + \beta \log(\eta\beta)$ in Eq. (22.1)], with η having an exponential distribution with density function

$$p_\eta(y) = \beta e^{-\beta y}, \qquad y > 0.$$

Then

$$\begin{aligned}\Pr[X \le x] &= 1 - \beta\int_0^\infty \exp\left[-\beta y\left\{1 + \exp\left(-\frac{x-\alpha}{\beta}\right)\right\}\right] dy \\ &= \left[1 + \exp\left\{-\frac{x-\alpha}{\beta}\right\}\right]^{-1},\end{aligned}$$

which is identical to the logistic distribution given in the next section. More historical details of the distribution may be found in Balakrishnan (1992).

2 DEFINITION

The distribution is most simply defined in terms of its cumulative distribution function

$$\begin{aligned}F_X(x) &= 1 - \left[1 + \exp\left\{\frac{x-\alpha}{\beta}\right\}\right]^{-1} \\ &= \left[1 + \exp\left\{-\frac{x-\alpha}{\beta}\right\}\right]^{-1} \\ &= \frac{1}{2}\left[1 + \tanh\left\{\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)\right\}\right], \qquad \text{with } \beta > 0. \qquad (23.6)\end{aligned}$$

It can be seen that (23.6) defines a proper cumulative distribution function with

$$\lim_{x \to -\infty} F_X(x) = 0,$$

$$\lim_{x \to \infty} F_X(x) = 1.$$

The corresponding probability density function is

$$\begin{aligned}p_X(x) &= \beta^{-1}\left[\exp\left\{\frac{x-\alpha}{\beta}\right\}\right]\left[1+\exp\left\{\frac{x-\alpha}{\beta}\right\}\right]^{-2}\\ &= \beta^{-1}\left[\exp\left\{-\frac{x-\alpha}{\beta}\right\}\right]\left[1+\exp\left\{-\frac{x-\alpha}{\beta}\right\}\right]^{-2}\\ &= (4\beta)^{-1}\operatorname{sech}^2\left\{\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)\right\}.\end{aligned} \tag{23.7}$$

The distribution is sometimes called the *sech-square(d)* distribution.

The function on the right-hand side of (23.6) has been used extensively to represent growth functions (with x representing time). We will be primarily concerned with its use as a distribution function (which can of course include situations in which the random variable represents time). It is worth noting that methods developed for fitting the logistic as a growth curve [e.g., Erkelens (1968) and Balakrishnan (1992, ch. 13)] can also be applied to fit the cumulative logistic distribution.

3 GENERATING FUNCTIONS AND MOMENTS

Making the transformation $Y = (X-\alpha)/\beta$, we obtain, from (23.7), the probability density function of Y

$$p_Y(y) = e^{-y}(1+e^{-y})^{-2} = \tfrac{1}{4}\operatorname{sech}^2\tfrac{1}{2}y. \tag{23.8}$$

The cumulative distribution function of Y is

$$F_Y(y) = (1+e^{-y})^{-1}. \tag{23.9}$$

Equations (23.8) and (23.9) are standard forms for the logistic distribution. [They are not the only standard forms. Equations (23.13) and (23.14), which express the distribution in terms of mean and standard deviation, can also be regarded as standard.]

The moment-generating function of the random variable Y with probability density function (23.8) is

$$\begin{aligned}E[e^{\theta Y}] = M_Y(\theta) &= \int_{-\infty}^{\infty} e^{-(1-\theta)y}(1+e^{-y})^{-2}\,dy\\ &= \int_0^1 \xi^{-\theta}(1-\xi)^{\theta}\,d\xi, \qquad \left[\text{with } \xi = (e^y+1)^{-1}\right]\\ &= B(1-\theta, 1+\theta)\\ &= \pi\theta \operatorname{cosec} \pi\theta.\end{aligned} \tag{23.10}$$

The characteristic function $E(e^{itY})$ is πt cosech πt. The moments of Y may be determined from (23.10), or by direct integration from (23.8). Using the latter method (with r > 0),

$$E[|Y|^r] = 2\int_0^{\infty} y^r e^{-y}(1+e^{-y})^{-2}\,dy$$

$$= 2\int_0^{\infty} y^r \sum_{j=1}^{\infty} (-1)^{j-1} j e^{-jy}\,dy$$

$$= 2 \quad + 1 \quad \overset{m}{\underset{j=1}{}} (-1)^{j-1} j^{-r} \qquad (\text{for r} > 0)$$

$$= 2\Gamma(r+1)(1-2^{-(r-1)})\zeta(r) \qquad (\text{for r} > 1), \qquad (23.11)$$

where $\zeta(r) = \sum_{j=1}^{\infty} j^{-r}$ is the Riemann zeta function (see Chapter 1).

The cumulants are (for r even) $\kappa_r(Y) = 6(2^r - 1)B_r$, where B_r is the rth Bernoulli number (see Chapter 1). If r is odd, $\kappa_r(Y) = 0$. The distribution of Y is symmetrical about y = 0. Putting r = 2,4 in (23.11)

$$\text{var}(Y) = E(Y^2) = 2\cdot 2(1-2^{-1})\left(\frac{\pi^2}{6}\right) = \frac{\pi^2}{3},$$

$$\mu_4(Y) = 2\cdot 24(1-2^{-3})\left(\frac{\pi^4}{90}\right) = \frac{7}{15}\pi^4.$$

The first two moment-ratios of the distribution are

$$\sqrt{\beta_1} = \alpha_3 = 0;\ \beta_2 = \alpha_4 = 4.2.$$

The mean deviation is $2\sum_{j=1}^{\infty}(-1)^{j-1}j^{-1} = 2\log_e 2$. Hence for the logistic distribution

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{2\sqrt{3}\log_e 2}{\pi} = 0.764.$$

Returning to the original form of the distribution (23.6), and recalling that $X = \alpha + \beta Y$, we see that

$$E[X] = \alpha$$

$$\text{var}(X) = \beta^2\pi^2/3. \qquad (23.12)$$

The coefficient of variation is, therefore, $\beta\pi/(\alpha\sqrt{3})$. The moment-ratios (and the ratio of mean deviation to standard deviation), are. of course. the same for X as for Y. The cumulative distribution function of X can be expressed in terms of $E[X] = \xi$ and $\text{var}(X) = \sigma^2$ in the standard form

$$F_X(x) = \left[1 + \exp\left\{-\frac{\pi(x-\xi)}{\sigma\sqrt{3}}\right\}\right]^{-1}. \tag{23.13}$$

The corresponding probability density function is

$$p_X(x) = \frac{\pi}{\sigma\sqrt{3}}\left[\exp\left\{-\frac{\pi(x-\xi)}{\sigma\sqrt{3}}\right\}\right] \cdot \left[1 + \exp\left\{-\frac{\pi(x-\xi)}{\sigma\sqrt{3}}\right\}\right]^{-2}. \tag{23.14}$$

The information-generating function [$(u-1)$-th frequency moment] corresponding to the probability density function (23.8) is

$$\begin{aligned} T_Y(u) &= \int_{-\infty}^{\infty} e^{-uy}(1 - e^{-y})^{-2u}\, dy \\ &= \int_{-\infty}^{\infty} \left(\frac{e^y}{1+e^y}\right)^u (1 + e^y)^{-u}\, dy \\ &= \int_0^1 \xi^{u-1}(1-\xi)^{u-1}\, d\xi \\ &= B(u,u) \\ &= \frac{[\Gamma(u)]^2}{\Gamma(2u)}. \end{aligned} \tag{23.15}$$

The entropy is

$$\begin{aligned} -T'_Y(1) &= \frac{-2\Gamma(1)\Gamma'(1)}{\Gamma(2)} + \frac{2[\Gamma(1)]^2\Gamma'(2)}{\Gamma(2)} \\ &= 2(\psi(2) - \psi(1)) \\ &= 2. \end{aligned}$$

4 PROPERTIES

Gumbel (1961) noted the properties

$$p_Y(y) = F_Y(y)[1 - F_Y(y)], \tag{23.16}$$

$$y = \log_e\left[\frac{F_Y(y)}{1 - F_Y(y)}\right] \tag{23.17}$$

for $p_Y(y)$, with $F_Y(y)$ defined as in (23.8) and (23.9). In general, the inverse distribution function or the quantile function (of probability γ) is

$$\alpha + \beta \log\left(\frac{\gamma}{1-\gamma}\right),$$

and the inverse survival function (of probability γ) is

$$\alpha + \beta \log\left(\frac{1-\gamma}{\gamma}\right).$$

As a result we readily note that the logistic (α, β) distribution arises as the distribution of

$$\alpha + \beta \log\left(\frac{U}{1-U}\right) \quad \text{or} \quad \alpha + \beta \log\left(\frac{e^{-V}}{1-e^{-V}}\right),$$

where U is a standard uniform $(0, 1)$ random variable and V is a standard exponential random variable.

The simple explicit relationships between y, $p_Y(y)$ and $F_Y(y)$ render much of the analysis of the logistic distribution attractively simple. The further fact that the logistic distribution has a shape similar to that of the normal distribution makes it profitable, on suitable occasions, to replace the normal by the logistic to simplify the analysis without too great discrepancies in the theory. Such substitution must be done with care and understanding of the similarities between the two distributions.

If the cumulative distribution functions $G_1(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-u^2/2}\,du$ and $G_2(x) = [1 + \exp(-\pi x/\sqrt{3})]^{-1}$ of the standardized normal and logistic distributions are compared, the differences $G_2(x) - G_1(x)$ vary in the way shown in Figure 23.1. Since both $G_1(x)$ and $G_2(x)$ are symmetric about $x = 0$, only the values for $x \geq 0$ are given. It can be seen that the maximum value of $G_1(x) - G_2(x)$ is about 0.0228, attained when $x = 0.7$. This maximum may be reduced to a value less than 0.01 by changing the scale of x in G_1 and using $G_1(16x/15)$ as an approximation to $G_2(x)$. This also is presented graphically in Figure 23.1. Volodin (1994) has determined that the constant $\pi/\sqrt{3.41}$ [instead of $(15\pi)/16\sqrt{3}$)] gives a better approximation with a maximum absolute difference of 0.0094825 [instead of 0.00953211. He has also indicated that the value of 1.7017456 will provide the best approximation with a maximum absolute difference of 0.0094573.

It should be noted that although there is a close similarity in shape between the normal and logistic distributions, the value of β_2 for logistic is 4.2, considerably different from the value ($\beta_2 = 3$) for the normal distribution. The difference may be attributed largely to the relatively longer tails of the logistic distribution. These can have a considerable effect on the fourth central moment, but a much smaller relative effect on the cumulative distribution function. [We may also note that whereas the standard normal

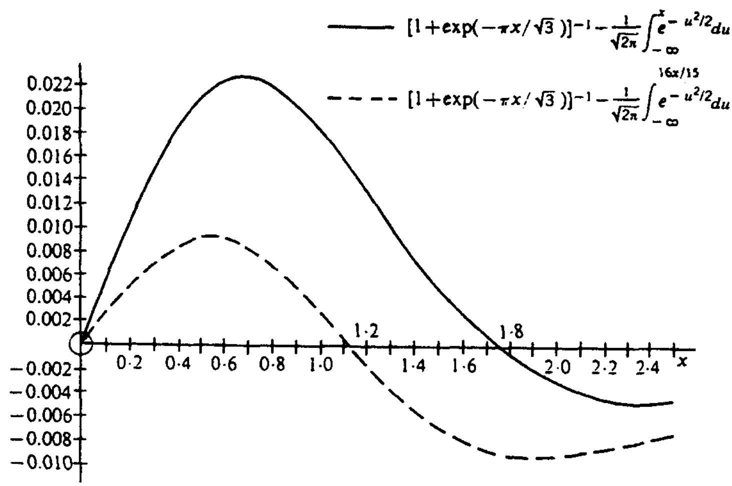


Figure 23.1 Comparison of Logistic and Normal Cumulative Distribution Functions.

curve has points of inflexion at x = ± 1, those of the logistic are x = $\pm(\sqrt{3}/\pi)\log_e(2+\sqrt{3}) = \pm 0.53$.]

The logistic density function is more peaked in the center than the normal density function; see Chew (1968). It is also easily observed that the hazard function is simply proportional to the cumulative distribution function. In fact it is this particular characterizing property of the logistic distribution that makes it useful as a growth curve model.

Noting the fact that $\beta_2 = 4.2$ for the logistic, Mudholkar and George (1978) observed that the logistic distribution very closely approximates Student's t-distribution with nine degrees of freedom. A similar rationale has been applied by George and Ojo (1980) and George, El-Saidi, and Singh (1986) in order to propose some close approximations for Student's t-distribution with υ degrees of freedom using generalized logistic distributions (see Section 10).

With $e(x)$ denoting the mean residual life function or remaining life expectancy function at age x given by

$$e(x) = E[X - x|X > x] = \int_x^\infty \{1 - F_X(t)\}\, dt/\{1 - F_X(x)\}$$

for $x \geq 0$, Ahmed and Abdul-Rahman (1993) have shown that

$$e(x) = \beta\{1 + e^{(x-\alpha)/\beta}\}\log\{1 + e^{-(x-\alpha)/\beta}\}$$

characterizes the logistic distribution in (23.6). They have also presented a number of equivalent conditions in terms of conditional expectations.

An expression for the distribution function of the sum of n i.i.d. logistic variables was obtained by Goel (1975) by using the Laplace transform inverse method for convolutions of Pólya-type functions, a technique developed by Schoenberg (1953) and Hirschman and Widder (1955). Goel (1975) also provided a table of the cdf of the sum of n i.i.d. standard logistic variables for n = 2(1)12, x = 0(0.01)3.99, and for n = 13(1)15, x = 1.20(0.01)3.99; he also presented a table of the quantiles for n = 2(1)15 and a = 0.90, 0.95, 0.975, 0.99, 0.995. George and Mudholkar (1983), on the other hand, derived an expression for the distribution of a convolution of the i.i.d. standard logistic variables by directly inverting the characteristic function. Both these expressions, however, contain a term $(1 - e^x)^{-k}$, k = 1, 2, ..., n, which pose a problem in precision of the computation at the values of x near zero when n is large.

George and Mudholkar (1983) also displayed that a standardized Student's t-distribution provides a very good approximation for the distribution of a convolution of n i.i.d. logistic variables. These authors then compared three approximations: (1) standard normal approximation, (2) Edgeworth series approximation correct to order n^{-1}, and (3) Student's t approximation with $\upsilon = 5n + 4$ degrees of freedom (obtained by equating the coefficient of kurtosis). Of these three the third provides a very good approximation.

Gupta and Han (1992) considered the Edgeworth and Cornish-Fisher series expansions (see Chapter 12 for details) up to order n^{-3} for the distribution of the standardized sample mean

$$T_n = \frac{\sqrt{n}}{\sigma}(\bar{X} - \xi),$$

when X_i's are i.i.d. logistic variables with cdf and pdf as in (23.13) and (23.14). They are given by

$$\begin{aligned} F_{T_n}(t) = \Phi(t) - \phi(t)\Bigg[&\frac{1}{n}\left\{\frac{1}{4!}\cdot\frac{6}{5}H_3(t)\right\} \\ &+ \frac{1}{n^2}\left\{\frac{1}{6!}\cdot\frac{48}{7}H_5(t) + \frac{35}{8!}\left(\frac{6}{5}\right)^2 H_7(t)\right\} \\ &+ \frac{1}{n^3}\left\{\frac{1}{8!}\cdot\frac{432}{5}H_7(t) + \frac{210}{10!}\cdot\frac{48}{7}\cdot\frac{6}{5}H_9(t) + \frac{5775}{12!}\left(\frac{6}{5}\right)^3 H_{11}(t)\right\}\Bigg] \\ &+ O(n^{-7/2}), \end{aligned}$$

and

$$T_n(U_\alpha) = U_\alpha + \frac{1}{n}\left\{\frac{1}{4!}\cdot\frac{6}{5}\left(U_\alpha^3 - 3U_\alpha\right)\right\}$$

$$+\frac{1}{n^2}\left\{\frac{1}{6!}\cdot\frac{48}{7}\left(U_\alpha^5 - 10U_\alpha^3 + 15U_\alpha\right)\right.$$

$$\left. + \frac{35}{8!}\left(\frac{6}{5}\right)^2\left(-9U_\alpha^5 + 72U_\alpha^3 - 87U_\alpha\right)\right\}$$

$$+\frac{1}{n^3}\left\{\frac{1}{8!}\cdot\frac{432}{5}\left(U_\alpha^7 - 21U_\alpha^5 + 105U_\alpha^3 - 105U_\alpha\right)\right.$$

$$+\frac{210}{10!}\cdot\frac{48}{5}\cdot\frac{6}{5}\left(-15U_\alpha^7 + 255U_\alpha^5 - 1035U_\alpha^3 - 855U_\alpha\right)$$

$$\left.+\frac{5775}{12!}\left(\frac{6}{5}\right)\left(243U_\alpha^7 - 3537U_\alpha^5 + 12177U_\alpha^3 - 8667U_\alpha\right)\right\}$$

$$+ O(n^{-7/2}),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf, U_α is the αth quantile of the standard normal distribution, and $H_j(t)$ is the Hermite polynomial defined in Chapter 1.

Table 23.1 A comparison of four approximations for the cdf of the standardized mean T_3 of samples of size 3 from a logistic population

t	$F_{T_3}(t)$	$F_{T_3}(t) - \Phi(t)$	$F_{T_3}(t) - A_1(t)$	$F_{T_3}(t) - A_2(t)$	$F_{T_3}(t) - A_3(t)$
0.05	0.5209	0.0010	0.0000	0.0001	0.0000
0.15	0.5625	0.0029	0.0000	0.0003	0.0000
0.25	0.6033	0.0046	0.0008	0.0005	0.0000
0.45	0.6809	0.0073	−0.0017	0.0007	0.0001
0.65	0.7506	0.0084	−0.0007	0.0007	0.0000
0.85	0.8106	0.0083	−0.0006	0.0007	0.0000
1.00	0.8486	0.0073	−0.0008	0.0004	0.0000
1.20	0.8903	0.0054	−0.0007	0.0002	0.0000
1.45	0.9291	0.0026	−0.0004	0.0000	0.0000
1.75	0.9598	−0.0001	0.0001	−0.0002	0.0000
2.50	0.9918	−0.0020	0.0004	0.0002	0.0000
3.00	0.9975	−0.0012	0.0001	0.0001	0.0000

Note $F_{T_3}(t)$ = exact cdf of the standardized mean taken from Goel (1975), $\Phi(t)$ = standard normal cdf, $A_1(t)$ = Edgeworth series expansion up to order n^{-1}, $A_2(t)$ = cdf of the standardized Student's t with 19 degrees of freedom, $A_3(t)$ = Edgeworth series expansion up to order n^{-3}.

Gupta and Han (1992) compared this approximation with the approximations mentioned earlier, and they showed the approximation to be far better than even the Student's t-approximation suggested by George and Mudholkar (1983). A comparison of these four approximations is presented in Table 23.1 for the sample size $n = 3$ [taken from Gupta and Han (1992)].

From Table 23.1 it is clear that the approximation using the Edgeworth expansion up to order n^{-3}, given by Gupta and Han (1992), is superior to the other three approximations, since its maximum error is about 0.0001 for the range of t considered.

5 ORDER STATISTICS

Let $Y_1' \le Y_2' \le \cdots \le Y_n'$ be the order statistics obtained from a sample of size n from the standard logistic distribution (23.8) and (23.9). Then from the density function of Y_r' $(1 \le r \le n)$ given by

$$p_{Y_r'}(y) = \frac{n!}{(r-1)!(n-r)!}\{F_Y(y)\}^{r-1}\{1 - F_Y(y)\}^{n-r} p_Y(y), \qquad -\infty < y < \infty, \tag{23.18}$$

we obtain the moment-generating function of Y_r' as

$$\begin{aligned} E[e^{\theta Y_r'}] = M_{Y_r'}(\theta) &= \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty}\frac{e^{-(n-r+1)y+\theta y}}{(1+e^{-y})^{n+1}}dy \\ &= \frac{B(r+\theta, n-r+1-\theta)}{B(r, n-r+1)} \\ &= \frac{\Gamma(r+\theta)\Gamma(n-r+1-\theta)}{\Gamma(r)\Gamma(n-r+1)}. \end{aligned} \tag{23.19}$$

An alternative expression for this moment-generating function of Y_r' in terms of Bernoulli numbers and Stirling numbers of the first kind has been given by Gupta and Shah (1965). From (23.19) we obtain

$$E[Y_r'] = \psi(r) - \psi(n - r + 1) \tag{23.20}$$

and

$$\operatorname{var}(Y_r') = \psi'(r) + \psi'(n - r + 1), \tag{23.21}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions, respectively (see Chapter 1).

From (23.19) we obtain the cumulant-generating function of Y_r' as

$$\begin{aligned} K_{Y_r'}(\theta) &= \log M_{Y_r'}(\theta) \\ &= \log \Gamma(r + 8) + \log \Gamma(n - r + 1 - 8) \\ &\quad - \log \Gamma(r) - \log \Gamma(n - r + 1) \end{aligned} \tag{23.22}$$

from which we get the kth cumulant of Y_r' as

$$\kappa_k(Y_r') = \psi^{(k-1)}(r) + (-1)^k \psi^{(k-1)}(n - r + 1), \tag{23.23}$$

$$\kappa_k(Y_r') = (-1)^k \kappa_k(Y_{n-r+1}'), \tag{23.23)'$$

where $\psi^{(k-1)}(\theta) = (d^k/d\theta^k)\log \Gamma(\theta)$ is the polygamma function. The expressions for the first four cumulants were given by Plackett (1958); see also Gumbel (1958).

By starting from the joint density function of X_r' and X_s' $(1 \le r < s \le n)$ and proceeding similarly, an expression for the joint moment-generating function of X_r' and X_s' and the product moment $E[X_r'X_s']$ can be derived; see, for example, Gupta, Qureishi, and Shah (1967) and Gupta and Balakrishnan (1992).

George and Mudholkar (1981a, b, 1982) provided joint characterizations of the logistic and the exponential based on order statistics. They observed that the characteristic function of Y_r' is

$$\phi_{Y_r'}(\theta) = E[e^{i\theta Y_r'}] = \prod_{j=1}^{r-1}\left(1 + \frac{i\theta}{j}\right)\prod_{k=1}^{n-r}\left(1 - \frac{i\theta}{k}\right)\phi_Y(\theta),$$

where $\phi_Y(\theta)$ is the characteristic function of the logistic density in (23.8). From this form they observed that $Y_r' + \Sigma_{k=1}^{n-r} E_{1k} - \Sigma_{j=1}^{r-1} E_{2j}$ is distributed as a standard logistic variable with density in (23.8), where E_{ij}'s are independent exponential random variables with density

$$p_{E_{ij}}(x) = je^{-jx}, \qquad x \ge 0,\ j = 1, 2, \dots, i = 1, 2.$$

Further characterization results of this nature relating the logistic, exponential and Laplace distributions may be found in the works of George and Mudholkar (1981a, b, 1982), George and Rousseau (1987), and Voorn (1987); see George and Devidas (1992) for a review of ail these results.

By making use of the characterizing differential equation (23.16), Shah (1966,1970) derived the following recurrence relations for single and product moments of the order statistics Y, (denoted by $Y_{r:n}$ for obvious notational convenience):

$$E[Y_{1:n+1}^{i+1}] = E[Y_{1:n}^{i+1}] - \frac{(i+1)}{n} E[Y_{1:n}^{i}], \qquad n \geq 1. \tag{23.24}$$

$$E[Y_{r+1:n+1}^{i+1}] = E[Y_{r:n+1}^{i+1}] + \frac{(i+1)(n+1)}{r(n-r+1)} E[Y_{r:n}^{i}], \qquad 1 \leq r \leq n. \tag{23.25}$$

$$E[Y_{r:n+1}Y_{r+1:n+1}] = E[Y_{r:n+1}^2] + \frac{n+1}{n-r+1} \times \left\{ E[Y_{r:n}Y_{r+1:n}] - E[Y_{r:n}^2] - \frac{1}{n-r} E[Y_{r:n}] \right\}, \qquad 1 \leq r \leq n-1. \tag{23.26}$$

$$E[Y_{r+1:n+1}Y_{r+2:n+1}] = E[Y_{r+2:n+1}^2] + \frac{n+1}{r+1} \left\{ E[Y_{r:n}Y_{r+1:n}] - E[Y_{r+1:n}^2] + \frac{1}{r} E[Y_{r+1:n}] \right\}, \qquad 1 \leq r \leq n-1. \tag{23.27}$$

$$E[Y_{r:n+1}Y_{s:n+1}] = E[Y_{r:n+1}Y_{s-1:n+1}] + \frac{n+1}{n-s+2} \left\{ E[Y_{r:n}Y_{s:n}] - E[Y_{r:n}Y_{s-1:n}] - \frac{1}{n-s+1} E[Y_{r:n}] \right\}, \qquad 1 \leq r < s \leq n;\ s - r \geq 2. \tag{23.28}$$

$$E[Y_{r+1:n+1}Y_{s+1:n+1}] = E[Y_{r+2:n+1}Y_{s+1:n+1}] + \frac{n+1}{r+1} \left\{ E[Y_{r:n}Y_{s:n}] - E[Y_{r+1:n}Y_{s:n}] + \frac{1}{r} E[Y_{s:n}] \right\}, \qquad 1 \leq r < s \leq n;\ s - r \geq 2. \tag{23.29}$$

Shah (1966,1970) showed that these recurrence relations are complete in the sense that, by starting with the values of moments of Y, these relations will enable one to determine the single and the product moments of order statistics for all sample sizes in a simple recursive manner.

Birnbaum and Dudman (1963) devoted considerable attention to comparison of distributions of order statistics from normal and logistic distributions. Gupta and Shah (1965) derived the distribution of the sample range from the logistic distribution, and compared it with the distribution of the sample range from the normal distribution for sample sizes 2 and 3. Malik (1980) derived the distribution of the rth quasi-range, $Y_{n-r:n} - Y_{r+1:n}$, for $r = 0, 1, 2, \dots, [(n-1)/2]$. Tarter and Clark (1965) discussed properties of the median. Plackett (1958) used the expression of the cumulants of $Y_{r:n}$ in (23.23) to develop some series approximations for the moments of order statistics from an arbitrary continuous distribution.

Kamps (1991), by considering a general class of distributions satisfying

$$\frac{d}{du}F^{-1}(u) = \frac{1}{d}u^{p}(1-u)^{q-p-1} \text{ on } (0,1),$$

has presented some characterization results through relations for moments of order statistics. The logistic distribution is, of course, a special case of this class (case $p = q = -1$). Reference may also be made to Kamps and Mattner (1993) for some further results in this direction.

The expression for the density function of the sample range, $W = Y_n' - Y_1'$, is

$$p_W(w) = \frac{\sqrt{\pi}\,\Gamma(n)}{2\sqrt{2}\,\Gamma(n+\frac{1}{2})}\left\{1 + \cosh\frac{w}{2}\right\}^{-(n-(1/2))} \times F\left(\frac{1}{2}, \frac{1}{2}; n + \frac{1}{2}; \left(1 - \cosh\frac{w}{2}\right)\right), \quad w > 0, \qquad (23.30)$$

where $F(a, b; c; x) = 1 + (ab/c)(x/1!) + [a(a+1)b(b+1)/\{c(c+1)\}](x^2/2!) + \cdots$ is the hypergeometric function.

Shah (1965) derived the joint density function of W and the midrange, $M = (Y_1' + Y_n')/2$, to be

$$p_{M,W}(m,w) = \frac{n(n-1)\{\sinh(w/2)\}^{n-2}}{4\{\cosh m + \sinh(w/2)\}^{n}}, \qquad w > 0,\ -\infty < m < \infty. \qquad (23.31)$$

By considering the symmetrically truncated logistic distribution with density function

$$p_Y(y) = \begin{cases} \dfrac{1}{1-2Q}\cdot\dfrac{e^{-y}}{(1+e^{-y})^2}, & -\log\left(\dfrac{1-Q}{Q}\right) \le y \le \log\left(\dfrac{1-Q}{Q}\right), \\ 0, & \text{otherwise}, \end{cases} \qquad (23.32)$$

and cumulative distribution function

$$F_Y(y) = \frac{1}{1-2Q}\left\{\frac{1}{1+e^{-y}} - Q\right\}, \qquad -\log\left(\frac{1-Q}{Q}\right) \le y \le \log\left(\frac{1-Q}{Q}\right), \tag{23.33}$$

where Q is the proportion of truncation on the left and the right of the standard logistic density function in (23.8), Balakrishnan and Joshi (1983a) derived several recurrence relations for single and product moments of order statistics. These generalize Shah's results presented in Eqs. (23.24)–(23.29).

Balakrishnan and Kocherlakota (1986) generalized the results of Balakrishnan and Joshi (1983a) by considering the doubly truncated logistic distribution with density function

$$p_Y(y) = \begin{cases} \dfrac{1}{P-Q}\cdot\dfrac{e^{-y}}{(1+e^{-y})^2}, & \log\left(\dfrac{Q}{1-Q}\right) \le y \le \log\left(\dfrac{P}{1-P}\right) \\ 0, & \text{otherwise}, \end{cases} \tag{23.34}$$

where Q and 1 – P are the proportions of truncation on the left and the right of the standard logistic density function in (23.8). For this case Tarter (1966) derived explicit expressions for the single and the product moments of order statistics. Braswell and Mandors (1970a, b) and Braswell and Pewitt (1973) considered the doubly truncated logistic distribution (but referred to it as the FRPDF, finite range probability distribution function) and discussed several inferential issues concerning the location and scale parameters of this distribution.

For a more detailed discussion on order statistics from the logistic distribution, one may refer to Gupta and Balakrishnan (1992).

6 METHODS OF INFERENCE

The maximum likelihood estimators, $\hat{\xi}, \hat{\sigma}$ of the parameters ξ, σ in (23.14) based on a mutually independent set of random variables $X_1, X_2, \ldots, X_n$, each having this distribution, satisfy the equations

$$n^{-1}\sum_{i=1}^{n}\left[1+\exp\left\{\frac{\pi(X_i-\hat{\xi})}{\hat{\sigma}\sqrt{3}}\right\}\right]^{-1} = \frac{1}{2} \tag{23.35}$$

$$n^{-1}\sum_{i=1}^{n}\left(\frac{X_i-\hat{\xi}}{\hat{\sigma}}\right)\frac{1-\exp\{\pi(X_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})\}}{1+\exp\{\pi(X_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})\}} = \frac{\sqrt{3}}{\pi}. \tag{23.36}$$

For large n

$$n \operatorname{var}(\hat{\xi}) \doteq \left(\frac{9}{\pi^2}\right)\sigma^2 \doteq 0.91189\sigma^2 \tag{23.37}$$

$$n \operatorname{var}(\hat{\sigma}) \doteq \left(\frac{9}{3+\pi^2}\right)\sigma^2 \doteq 0.69932\sigma^2. \tag{23.38}$$

Equations (23.35) and (23.36) must be solved by trial and error.

Taking advantage of the similarity in shape between the logistic and normal distributions, initial values of $\hat{\xi}$ and $\hat{\sigma}$ might be taken as the maximum likelihood estimators

$$= n^{-1} \sum_{i=1}^{n} X_i, \quad \text{and} \quad \sqrt{n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2},$$

respectively, appropriate to the latter distribution. Improvements could then be made, using (23.35) and (23.36), by applying, for example, the Newton-Raphson method.

Similarly, if the available sample is Type-II censored, say X'_{r+1}, $X'_{r+2}, \cdots, X'_{n-s}$ with r smallest and s largest observations censored, then the maximum likelihood estimators, $\hat{\xi}$ and $\hat{\sigma}$, of the parameters ξ and σ in (23.14) satisfy the equations

$$(n-r-s) - r\left\{\frac{e^{-\pi(X'_{r+1}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}{1+e^{-\pi(X'_{r+1}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\} + s\left\{\frac{1}{1+e^{-\pi(X'_{n-s}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\}$$
$$-2\sum_{i=r+1}^{n-s}\left\{\frac{e^{-\pi(X'_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}{1+e^{-\pi(X'_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\} = 0, \tag{23.39}$$

$$-(n-r-s)\frac{\sqrt{3}}{\pi} - r\left(\frac{X'_{r+1}-\hat{\xi}}{\hat{\sigma}}\right)\left\{\frac{e^{-\pi(X'_{r+1}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}{1+e^{-\pi(X'_{r+1}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\}$$
$$+ s\left(\frac{X'_{n-s}-\hat{\xi}}{\hat{\sigma}}\right)\left\{\frac{1}{1+e^{-\pi(X'_{n-s}-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\} + \sum_{i=r+1}^{n-s}\left(\frac{X'_i-\hat{\xi}}{\hat{\sigma}}\right)$$
$$-2\sum_{i=r+1}^{n-s}\left(\frac{X'_i-\hat{\xi}}{\hat{\sigma}}\right)\left\{\frac{e^{-\pi(X'_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}{1+e^{-\pi(X'_i-\hat{\xi})/(\hat{\sigma}\sqrt{3})}}\right\} = 0. \tag{23.40}$$

For the case where $r = s = 0$, Eqs. (23.39) and (23.40) reduce to (23.35) and (23.36). Here again (23.39) and (23.40) have to be solved by numerical methods.

By using Monte Carlo simulations and solving the likelihood equations by regula falsi method, Harter and Moore *(1967)* determined the bias, variances, covariance, and conditional variances of $\hat{\xi}$ and $\hat{\sigma}$ for sample sizes $n = 10$ and *20* and various choices of censoring. A table of asymptotic variances and covariance of $\hat{\xi}$ and 6, for various choices of proportions of censoring $p_r = r/n$ and $p_s = s/n$, was presented by Harter and Moore *(1967);* see also Harter *(1970)* and Balakrishnan (1992).

Bain et al. *(1992)* considered the interval estimation of the parameters ξ and a based on Type-II censored samples. For this purpose they presented some simulated percentage points of the pivotal quantities $\sqrt{n}(\hat{\xi} - \xi)/\hat{\sigma}$ and $\sqrt{n}(\hat{\sigma}/\sigma - 1)$. These authors have also presented some tables of lower γ tolerance factors for proportion β, t_γ, where the lower γ tolerance limit for proportion β is given by

$$L(X) = \hat{\xi} - t_\gamma \hat{\sigma}, \tag{23.41}$$

and it is such that

$$\Pr\{1 - F_X(L(X); \xi, \sigma) \geq \beta\} = \gamma. \tag{23.42}$$

Their tables provide values of the factor t_γ for various values of γ and β and for different sample sizes and choices of right censoring (i.e., r = 0).

Due to the symmetry of the logistic distribution, these tolerance factors t_γ may also be used to determine the upper γ tolerance limit for proportion β given by

$$U(X) = \hat{\xi} + t_\gamma \hat{\sigma}. \tag{23.43}$$

Bain et al. *(1992)* also demonstrated how their tables of tolerance factors may be used to determine $100\gamma\%$ lower confidence limits for the reliability function $R_X(t) = 1 - F_X(t; \xi, \sigma)$.

Lawless (1972) discussed the conditional methods of inference for the location and scale parameters, ξ and a, in *(23.14).* Approximate linearization of the maximum likelihood equations (23.35) and (23.36) was effected by Plackett *(1958).* He gave coefficients in estimators $\hat{\xi}', \hat{\sigma}'$, that are quite similar to those of the best linear unbiased estimators for parameters of normal distributions, even when the sample size is no greater than *10.* Another method of approximate linearization was proposed by Tiku *(1968).* Fisk *(1961)* described the maximum likelihood estimation method based on grouped or truncated data; also see Hassanein and Sebaugh (1973).

The parameters ξ, a may also be estimated by the sample mean and standard deviation, m_1', $\sqrt{m_2}$, respectively. The asymptotic efficiency of m_1' is *91.2%;* that of $\sqrt{m_2}$ is *87.4%.* Gupta, Qureishi, and Shah *(1967)* show that the actual efficiency of m_1', as an estimator of ξ, and of $\sqrt{m_2}$, as an estimator of σ, is greater than the asymptotic efficiency when the sample size is small. These estimators are, however, less efficient (about 10% less for m'_2, considerably more so for $\sqrt{m_2}$) than the appropriate best linear unbiased estimators.

A number of methods of fitting parameters of logistic curves were developed in connection with their use as growth curves. Descriptions of such methods can be found in Erkelens (1968), Oliver (1964), Pearl (1940), Rasor (1949), Silverstone (1957), Will (1936), D'Agostino and Massaro (1992), and Tsokos and DiCroce (1992). Many of these are of a heuristic nature and are not based directly on probabilistic considerations, but are nevertheless useful in obtaining quick estimators of the parameters. [The fitting of the distribution is simpler than the fitting as a growth curve, since in the former case there is no need to fit values for A and B, see Eq. (23.1).]

From (23.13) the expected value of

$$f_x = \frac{\text{Number of X's} \le x}{n}$$

is $[1 + \exp\{-\pi(x - \xi)/(\sigma\sqrt{3})\}]^{-1}$. One method of fitting consists of plotting $\log[f_x/(1 - f_x)]$ against x and fitting (often by eye) a straight line

$$\log\left[\frac{f_x}{1 - f_x}\right] = \hat{a} + \hat{b}x \tag{23.44}$$

to the data so obtained. Comparison of coefficients in (23.44) and

$$\log\left[\frac{E(f_x)}{1 - E(f_x)}\right] = \frac{\pi}{\sigma\sqrt{3}}(-\xi + x) \tag{23.45}$$

leads to the estimators

$$\hat{\sigma} = \frac{\pi}{\hat{b}\sqrt{3}}, \quad \hat{\xi} = -\frac{\hat{a}}{\hat{b}}$$

Although various refinements (e.g., in the fitting of the line and in reducing bias) can be introduced, this is quite effective as a quick method. Similar methods are not as effective when a growth curve is being fitted, and values of the upper, and possibly also the lower, asymptotes have to be estimated [Oliver (1964)]. When only quantal response data (i.e., proportions of observations exceeding certain specified values) are available, special methods must be used.

Despite the simple form of the joint probability density functions of the order statistics corresponding to a random sample from a logistic distribution, the variance-covariance matrix does not have a simple analytic form, as in the cases of the exponential and uniform distributions. As a consequence we are very nearly in the same position in regard to construction of best linear unbiased estimators of the parameters ξ and σ, using order statistics, as for the normal distribution. It is again necessary to rely on numerical calculation, using tables rather than analytical formulas.

Gupta, Qureishi, and Shah (1967) showed that the efficiency (relative to the Cramér-Rao lower bounds in (23.37) and (23.38)) of the best linear unbiased estimator of ξ increases from about 95% for n = 5 to about 98% for n = 25; for σ the increase is from about 80% to about 90%.

There are explicit approximate formulas available [Gupta and Gnanadesikan (1966)] for best linear unbiased estimators based on k selected order statistics $X'_{n_1}, X'_{n_2}, \cdots, X'_{n_k}$ (with $1 \leq n_1 < n_2 < \cdots < n_k \leq n$) from n independent random variables $X_1, X_2, \cdots, X_n$ each having probability density function (23.14). These formulas should give useful results when n is large, while n_1/n and n_k/n are not "too near" to 0 or 1, respectively. This method is based on large-sample approximations to the expectations, variances and covariances of order statistics [see, e.g., Ogawa (1951)].

The first of these approximate formulas to be described here can be used to estimate ξ, σ being known. It is

$$\xi^* = \frac{\begin{array}{c}\sum_{i=1}^{k+1}[1-(n_i/n)-(n_{i-1}/n)] \\ \times[(n_i/n)(1-(n_i/n))X'_{n_i} - (n_{i-1}/n)(1-(n_{i-1}/n))X'_{n_{i-1}}] - K_3\sigma\end{array}}{K_1} \tag{23.46}$$

where

$$K_1 = \sum_{i=1}^{k+1}\left(\frac{n_i}{n} - \frac{n_{i-1}}{n}\right)\left(1 - \frac{n_i}{n} - \frac{n_{i-1}}{n}\right)^2$$

$$K_3 = \sum_{i=1}^{k+1}\left(1 - \frac{n_i}{n} - \frac{n_{i-1}}{n}\right)$$
$$\times\left[\frac{n_i}{n}\left(1-\frac{n_i}{n}\right)\log\left\{\frac{n_i/n}{(1-(n_i/n))}\right\} - \frac{n_{i-1}}{n}\left(1-\frac{n_{i-1}}{n}\right)\right.$$
$$\left.\times\log\left\{\frac{n_{i-1}/n}{1-(n_{i-1}/n)}\right\}\right],$$

(X'_{n_0} and $X'_{n_{k+1}}$ are each defined to be zero). Note that the coefficients in ξ^* depend only on the ratios n_i/n.

The variance of ξ^* is approximately $(\sigma^2 n^{-1})K_1^{-1}$. For given k this last quantity is minimized if $n_1, n_2, \ldots, n_k$ are chosen to maximize K_1. This is achieved by taking $n_i = ni/(k+1)$. (In practice of course the nearest suitable integer value would be taken.) With these values of the n_i's we have the estimator

$$\xi^{**} = \frac{6}{k(k+1)(k+2)} \sum_{i=1}^{k} i(k+1-i) X'_{n_i} \tag{23.47}$$

with $\text{var}(\xi^{**}) = (9/\pi^2)(\sigma^2/n)[k^{-1}(k+2)^{-1}(k+1)^2]$.

The **Cramér-Rao** lower bound for an unbiased estimator of ξ is $(9/\pi^2)(\sigma^2/n)$. The relative efficiency of [** is approximately $k^{-1}(k+2)^{-1}(k+1)^2$. This increases with k, from a minimum of 75% when k = 1 up to 100% as k increases. (When k = 1,[** = median $(x_1, x_2, \cdots, x_n)$.) It may be noted that (23.47) is the estimator obtained by Blom's (1956) method. It is also the estimator obtained by multiplying Jung's (1956) estimator by a constant to make it unbiased. The formula for estimating a , ξ being known, is

$$\sigma^* = \frac{Y - \xi K_3}{K_2}, \tag{23.48}$$

where

$$Y = \sum_{i=1}^{k+1} \left[\frac{n_i}{n}\left(1 - \frac{n_i}{n}\right) \log\left\{ \frac{n_i/n}{1-(n_i/n)} \right\} - \frac{n_{i-1}}{n}\left(1 - \frac{n_{i-1}}{n}\right) \times \log\left\{ \frac{n_{i-1}/n}{1-(n_{i-1}/n)} \right\} \right] \times \frac{\{(n_i/n)(1-(n_i/n))X'_{n_i} - (n_{i-1}/n)(1-(n_{i-1}/n))X'_{n_{i-1}}\}}{(n_i/n)-(n_{i-1}/n)}$$

$$K_2 = \sum_{i=1}^{k+1} \frac{[(n_i/n)(1-(n_i/n))\ln\{(n_i/n)(1-(n_i/n))\}}{(n_i/n)-(n_{i-1}/n)} - \frac{(n_{i-1}/n)(1-(n_{i-1}/n))\ln\{(n_{i-1}/n)(1-(n_{i-1}/n))\}]^2}{(n_i/n)-(n_{i-1}/n)},$$

with approximate variance $(\sigma^2/n)K_2^{-1}$.

Gupta and Gnanadesikan (1966) gave a detailed comparison of the estimators of a (ξ not being known) obtained by Blom's (1956,1958) and Jung's (1956) methods. They concluded that these estimators have high efficiencies. Table 23.2, taken from Gupta and Waknis (1965), gives the coefficients a_i in Jung's formula $\sum a_i(X'_{n-i+1} - X'_i)$ modified to make it an unbiased estimator of σ.

The general problem of maximizing K_2, and so minimizing $\text{var}(\sigma^*)$, is rather complex. However, if only two order statistics are to be used then

Table 23.2 Coefficients of the $(n - i + 1)$th-order statistic X'_{n-i+1} in the linear estimator of σ (by Jung's method) modified to make it unbiased

	i													Variance
n	1	2	3	4	5	6	7	8	9	10	11	12	13	σ^2
5	0.3538	0.2038	0											0.1706
6	0.2907	0.2024	0.0715											0.1372
8	0.2125	0.1767	0.1147	0.0396										0.0985
10	0.1663	0.1503	0.1170	0.0737	0.0251									0.0769
15	0.1062	0.1048	0.0955	0.0813	0.0636	0.0436	0.0222	0						0.0496
20	0.0774	0.0787	0.0758	0.0700	0.0622	0.0528	0.0422	0.0307	0.0187	0.0063				0.0366
25	0.0605	0.0625	0.0618	0.0592	0.0553	0.0504	0.0445	0.0381	0.0310	0.0236	0.0159	0.0080	0	0.0293

Note: Computed by using the same approximate covariance matrix as used in Blom's method.

$n_1 \doteq 0.103n$, and $n_2 \doteq 0.897n$ gives approximately the minimum value $1.0227\sigma^2/n$ for $\text{var}(\sigma^*)$. The Cramér-Rao lower bound for the variance of an unbiased estimator is $9(3+\pi^2)^{-1}\sigma^2/n$, and so the estimator

$$\sigma^* = 0.4192\left[X'_{n_1} + X'_{n_2}\right]$$

(with n_1 and n_2 as given above) has approximate efficiency 68.38%. (It may be noted that to get an improved estimator of this form, *four* quantities are needed.)

If neither ξ nor σ are known, the approximate best linear estimators, obtained by similar methods, are

$$\hat{\xi}^* = \Delta^{-1}(K_2X - K_3Y),$$
$$\hat{\sigma}^* = \Delta^{-1}(-K_3X + K_1Y), \qquad (23.49)$$

where

$$\Delta = K_1K_2 - K_3^2,$$
$$X = \sum_{i=1}^{k+1}\left(1 - \frac{n_i}{n} - \frac{n_{i-1}}{n}\right)\left[\frac{n_i}{n}\left(1 - \frac{n_i}{n}\right)X'_{n_i} - \frac{n_{i-1}}{n}\left(1 - \frac{n_{i-1}}{n}\right)X'_{n_{i-1}}\right].$$

[K_1, K_2, and K_3 and Y have been defined in (23.46) and (23.48).]

Simpson (1967), Hassanein (1969, 1974), Chan (1969), Chan, Chan, and Mead (1971,19731, Chan and Cheng (1972, 1974), and Cheng (1975) have all discussed the problem of optimal linear estimation of parameters ξ and σ for the logistic case. All these developments are reviewed by Cheng (1992) who also presents some of the relevant tables. Saleh, Hassanein, and Ali (1992) have discussed the optimal linear estimation of quantiles of the logistic distribution in (23.8) based on selected order statistics and have also presented the required tables; similar work was carried out by Ali and Umbach (1989) for the symmetrically truncated logistic distribution. Linear estimators with polynomial coefficients are examined by Balakrishnan (1992) along the lines of Downton (1966). Raghunandanan and Srinivasan (1970) had proposed some simple linear estimators for$ and σ based on quasi-midranges and quasi-ranges, respectively.

Construction of confidence intervals for the parameters ξ and σ was discussed by Antle, Klimko, and Harkness (1970) who determined the necessary percentage points of the pivotal quantities through Monte Carlo simulations. Schafer and Sheffield (1973) and Bain et al. (1992) have provided further discussions on this issue, with the last authors dealing with Type-I1 censored samples.

Howlader and Weiss (1989) have worked out Bayesian estimators of the reliability function $R_X(t)$ by employing the methods of Lindley and Tierney and Kadane. Both squared-error and log-odds squared-error loss functions were used by these authors. Aguirre and Nikulin (1993) have recently discussed the chi-squared goodness-of-fit test for the logistic distribution. Iqbal (1993) has presented asymptotic expansions for confidence limits for the parameters of the logistic distribution.

7 RECORD VALUES

Let $Y_{U(1)}, Y_{U(2)}, Y_{U(3)}, \cdots$ denote the upper record values arising from a sequence $\{Y_i\}$ of i.i.d. random variables with standard logistic density function (23.8). Then the density function of the nth upper record value $Y_{U(n)}$ is given by

$$p_{Y_{U(n)}}(y) = \frac{1}{(n-1)!}\{-\log(1-F_Y(y))\}^{n-1} p_Y(y), \qquad -\infty < y < \infty, \tag{23.50}$$

and the joint density function of $Y_{U(m)}$ and $Y_{U(n)}$ $(1 \le m < n)$ is

$$\begin{aligned} p_{Y_{U(m)}, Y_{U(n)}}(y_1, y_2) &= \frac{1}{(m-1)!(n-m-1)!}\{-\log(1-F_Y(y_1))\}^{m-1} \\ &\quad \times \frac{p_Y(y_1)}{1-F_Y(y_1)} \\ &\quad \times \{-\log(1-F_Y(y_2)) + \log(1-F_Y(y_1))\}^{n-m-1} p_Y(y_2) \\ &\quad -\infty < y_1 < y_2 < \infty. \end{aligned} \tag{23.51}$$

For this case Balakrishnan, Ahsanullah, and Chan (1994) examined the moment properties of record values from (23.50) and (23.51). For example, from (23.50) we have for $n = 1, 2, \ldots,$

$$\begin{aligned} E[Y_{U(n)}] &= \int_{-\infty}^{\infty} y\, p_{Y_{U(n)}}(y)\, dy \\ &= \int_0^1 F_Y^{-1}(u) \cdot \frac{1}{(n-1)!}\{-\log(1-u)\}^{n-1}\, du \\ &= \int_0^1 (\log u) \frac{1}{(n-1)!}\{-\log(1-u)\}^{n-1}\, du \\ &\quad + \int_0^1 \frac{1}{(n-1)!}\{-\log(1-u)\}^{n}\, du \\ &= n - S_n, \end{aligned} \tag{23.52}$$

where

$$S_n = \sum_{k=1}^{\infty} \frac{1}{k(k+1)^n}, \qquad n = 1, 2, \dots. \tag{23.53}$$

By using the facts that $S_1 = 1$ and $S_{n+1} - S_n = 1 - \zeta(n+1)$, where $\zeta(\cdot)$ is the Riemann zeta function, Balakrishnan, Ahsanullah, and Chan (1994) established that

$$E\left[Y_{U(1)}\right] = 0 \quad \text{and} \quad E\left[Y_{U(n+1)}\right] = E\left[Y_{U(n)}\right] + \zeta(n+1), \qquad n \geq 1. \tag{23.54}$$

Proceeding similarly, they have also derived expressions for the variances and covariances of the upper record values as

$$\operatorname{var}\left(Y_{U(n)}\right) = 2n\zeta(n+1) - n - S_n^2 + 2T_n, \qquad n \geq 1, \tag{23.55}$$

and

$$\operatorname{cov}\left(Y_{U(m)}, Y_{U(n)}\right) = m\{\zeta(m+1) + \zeta(n+1) - 1\} - S_m S_n + \sum_{k=1}^{\infty} \frac{1}{k(k+1)^{n-m}} \sum_{l=1}^{\infty} \frac{1}{l(l+1+k)^m}, \qquad 1 \leq m < n, \tag{23.56}$$

where S_n is as defined in (23.53) and

$$T_n = \sum_{l=2}^{\infty} \frac{1}{l(l+1)^n}\left(1 + \frac{1}{2} + \cdots + \frac{1}{l-1}\right), \qquad n \geq 1. \tag{23.57}$$

By making use of Eqs. (23.55), (23.56), and (23.57), Balakrishnan, Ahsanullah, and Chan (1994) tabulated the means, variances and covariances of the upper record values $Y_{U(n)}$ for n up to 10. Due to the symmetry of the logistic density function in (23.8), these tables also readily give the means (negative of the corresponding entries for upper records), variances, and covariances of the lower record values $Y_{L(n)}$.

These tables were utilized by Balakrishnan, Ahsanullah, and Chan (1994) to derive the best linear unbiased estimators of the location and scale parameters, α and β, in (23.6) based on the observed values of the first n upper record values. Tables of coefficients of these best linear unbiased estimators and the variances of the estimators were presented for n up to 10. The problems of constructing prediction intervals for a future record and of testing for the spuriosity of the current record value were also addressed by these authors, and necessary tables were set up.

8 TABLES

Values of the standard density function $p_Y(y)$ in (23.8) and the corresponding cumulative distribution function $F_Y(y)$ in (23.9) are presented in the collection of tables by Owen (1962). Each function is tabulated to four decimal places for y = 0(0.01)1.00(0.05)3.00. Only positive values of y are needed due to the symmetry of the distribution. Inverse tables give, to four decimal places, values of y and $p_Y(y)$ for which the distribution function $F_Y(y)$ takes the values 0.5(0.1)0.90(0.005)0.99(0.001)0.999(0.0001)0.9999, and some selected higher values up to 0.999999999.

Finney (1947,1952) has presented tables for use in logit analysis. These tables include the logit transformation, $y = \log(F/(1 - F))$. Berkson (1953) has also given short tables of the logit. There are similar tables (Tables XI and XI,) in Fisher and Yates (1957).

Tables of $100\alpha\%$ points, for a = 0.50, 0.75, 0.90, 0.95, 0.975, 0.99, for all order statistics for sample sizes up to 10 and for extreme and central order statistics for sample sizes from 11 to 25 have been presented by Gupta and Shah (1965). These authors have also presented a table of probability integrals of the sample range W, at w = 0.2(0.2)1.0(0.5)4.0, for sample sizes 2 and 3.

Tables of means and standard deviations of order statistics for sample sizes n = 1(1)10, 15, 20, 25 (and some selected order statistics for n = 100) were given by Birnbaum and Dudman (1963). Tables of covariances of order statistics for sample sizes up to 10 have been presented by Shah (1966). These tables were extended by Gupta, Qureishi, and Shah (1967) for sample sizes up to 25, and by Balakrishnan and Malik (1994) for sample sizes up to 50 (to ten decimal accuracy). For the symmetrically truncated logistic distribution in (23.32), Balakrishnan and Joshi (1983b) presented tables of means, variances, and covariances of order statistics for sample sizes up to 10 and for the proportion of truncation Q = 0.01, 0.05(0.05)0.20.

Tables of asymptotic variances and covariance of the maximum likelihood estimators of ξ and σ in (23.14) were presented by Harter and Moore (1967) and Harter (1970) for various choices of proportions of censoring on the left and the right of the sample. Tables of coefficients of the best linear unbiased estimators of ξ and σ and the variances and covariance of these estimators were tabulated by Gupta, Qureishi, and Shah (1967) for sample sizes n = 2(1)5(5)25 and some selected choices of censoring. Balakrishnan (1991) presented more exhaustive tables covering sample sizes n = 2(1)25(5)40 and all possible choices of censoring. Cheng (1992) has given tables of optimal spacings, the corresponding coefficients, and the variances for the asymptotic best linear unbiased estimators of ξ and σ based on k selected order statistics. Similar tables for the asymptotic best linear unbiased estimators of logistic quantiles (based on k optimally selected order statistics) are presented by Saleh, Hassanein, and Ali (1992). Bain et al. (1992) give tables of percentage points of pivotal quantities involving the maximum likelihood estimators of ξ and σ from which one can construct confidence intervals for the parameters ξ and σ based on complete or Type-11 censored samples. They also present tables of tolerance factors that are necessary in determining lower and upper tolerance limits, as well as lower confidence limits for the reliability, based on given complete or Type-II censored samples. Similar tables of tolerance factors involving best linear unbiased estimators of ξ and σ were constructed by Hall (1975). Balakrishnan and Fung (1992) extended the tables of Hall, giving one-sided tolerance factors for sample sizes up to 40 and also presenting two-sided tolerance factors. D'Agostino and Massaro (1992) give tables of critical points of various goodness-of-fit tests useful for testing the validity of the assumption of the logistic distribution for the data at hand; also see D'Agostino and Stephens (1986).

Balakrishnan et al. (1991) prepared tables of means, variances, and covariances of logistic order statistics in the presence of a single location or scale outlier. These tables are used by Balakrishnan (1992) in examining the robustness features of various linear estimators of the location and scale parameters, ξ and σ, of the logistic population.

9 APPLICATIONS

Some important uses of the logistic curve or distribution have already been mentioned. These include use in describing growth and as a substitute for the normal distribution. Possibly [see Berkson (1944, 1951, 1955, 1957)] included in the latter is its use in the analysis of quantal response data. This type of analysis has already been described (Chapter 13) in connection with the use of the normal distribution in probit analysis. If a logistic distribution is used, in place of a normal distribution, to represent the population tolerance distribution then the analysis is carried out in terms of logits instead of probits.

The logit Y and the corresponding observed proportion P are connected by the equation

$$P = (1 + e^{-Y})^{-1},$$

that is,

$$Y = \log\left(\frac{P}{1 - P}\right).$$

Given observed proportions $P_i = D_i/n_i$ of "deaths" at "dosages" x_i ($i = 1, \dots, k$), the logits $Y_i = \log[P_i/(1 - P_i)]$ are calculated. The estimation of values of the constants α, β in the equation

$$P = (1 + e^{-(\alpha+\beta x)})^{-1}$$

is to be based on the k independent binomial proportions P_i, or equivalently on the k independent logits Y_i. The maximum likelihood estimators $\hat{\alpha}, \hat{\beta}$ of α, β, respectively, satisfy the equations

$$\sum_{i=1}^{k} n_i(P_i - \hat{P}_i) = 0 = \sum_{i=1}^{k} n_i x_i (P_i - \hat{P}_i), \tag{23.58}$$

where

$$\hat{P}_i = \left[1 + \exp\left\{-\left(\hat{\alpha} + \hat{\beta} x_i\right)\right\}\right]^{-1}.$$

An iterative method for solving these equations, linked with the idea of fitting a weighted regression of Y_i on x_i, can be constructed in an exactly similar way to that described for probit analysis. In one respect the calculations are simpler, as the weight per unit observation corresponding to P_i is $P_i(1 - P_i)$, which is simpler than the corresponding formula for probit analysis. As is to be expected from the similarity in shape of the normal and logistic distributions, the results of probit and logit analysis of the same data are usually very similar. Agreement is particularly good in respect of estimates of the median of the tolerance distribution [Finney (1947)]. Systems of analysis, using an assumed logistic form for residual variation, were worked out for 2^n factorial experiments by Dyke and Patterson (1952), and for the general linear hypothesis by Grizzle (1961). Multiple comparisons, using the logistic distribution, were discussed by Reiersøl (1961).

The logistic function was also used in studies concerning physiochemical phenomenon by Pearl and Reed (1929), psychological issues by Birnbaum and Dudman (1963), Lord (1965), Sanathanan *(1974)* and Formann (1982), and geological issues by Aitchison and Shen (1980). Vieira and Hoffmann *(1977)* applied the logistic function to weight-gain data of Holstein cows, while Glasbey *(1979)* applied to the weight-gain analysis of Ayrshire steer calves. Leach (1981) and Oliver (1982) used the logistic model for the growth of human population.

The problem of medical diagnosis through the logistic discriminant function was introduced first by Cox *(1966)* and Day and Kerridge (1967), and later extended by Anderson *(1972,1973,1974)*. Wijesinha et al. *(1983)* and Begg and Gray *(1984)* applied the polychotomous logistic regression model to a large data set of patients where there were many distinct diagnostic categories. Breslow and Powers *(1978)* compared the prospective and the retrospective models of the logistic regression analysis based on data from the Oxford Childhood Cancer Survey reported by Kneale (1971). Johnson (1985) applied the logistic regression to estimate the survival time of diagnosed leukemia patients.

McCullagh (1977) extended some simple odds ratio statistics and presented an application concerning the degrees of pneumoconiosis in coal miners. Greenland (1985) discussed some extensions of logistic models to the modeling of probabilities of ordinal responses and illustrated an application in an analysis of the dependence of chronic obstructive respiratory disease prevalence on smoking and age. Bonney (1986) introduced the regressive logistic model to merge the goals and methodologies of both the epidemiologist and geneticist in the study of familial disease and other binary traits. Kay and Little (1986) used the logistic regression model to analyze the hemolytic uremic syndrome data secured from a number of children. The logistic distribution, the logistic growth model, and the logistic regression model have found numerous other applications. Interested readers may refer to the volume by Balakrishnan (1992).

10 GENERALIZATIONS

Several different forms of generalizations of the logistic distribution have been proposed in the literature. The type I generalized logistic distribution has cumulative distribution function

$$F_Y(y) = \frac{1}{(1+e^{-y})^{\alpha}}, \qquad -\infty < y < \infty, \alpha > 0. \tag{23.59}$$

The standard logistic distribution function in (23.9) corresponds to the case $\alpha = 1$ in (23.59). Dubey (1969) observed that if Y, given η, has an extreme value distribution with density function $\eta e^{-y} e^{-\eta e^{-y}}$ ($\eta > 0$) and η has a gamma distribution with density function $e^{-\eta}\eta^{\alpha-1}/\Gamma(\alpha)$ ($\alpha > 0$), then the unconditional distribution of Y is the type I generalized logistic distribution in (23.59). Zelterman (1987a, b) considered the three-parameter form of the type I distribution in (23.59) (by introducing location and scale parameters) and discussed methods of estimation of the parameters. Balakrishnan and Leung (1988a) and Zelterman (1989) studied order statistics from this distribution, while Balakrishnan and Leung (1988b) presented tables of means, variances and covariances of order statistics for various values of the shape parameter α. By making use of these tables, Balakrishnan and Leung (1988b) also derived the best linearunbiased estimators of the location and scale parameters (with α being assumed known) and presented the necessary tables. Gerstenkorn (1992) discussed the estimation of the parameter α.

The type I distribution in (23.59) is negatively skewed for $0 < \alpha < 1$ and positively skewed for $\alpha > 1$. Ahuja and Nash (1967) showed that if Y has the type I generalized logistic distribution in (23.59), then $-\alpha Y$ behaves like a standard exponential random variable when α is close to 0, and $Y - \ln \alpha$ behaves like an extreme value random variable with distribution function $e^{-e^{-y}}$ (see Chapter 22) when α is large.

The type II generalized logistic distribution has cumulative distribution function

$$F_Y(y) = 1 - \frac{e^{-\alpha y}}{(1 + e^{-y})^{\alpha}}, \qquad -\infty < y < \infty, \alpha > 0. \tag{23.60}$$

It is easy to observe that if Y has a type I generalized distribution in (23.59), then $-Y$ has a type II generalized distribution in (23.60). Consequently the type II distribution in (23.60) is positively skewed for $0 < \alpha < 1$ and negatively skewed for $\alpha > 1$.

The type III generalized logistic distribution has density function

$$p_Y(y) = \frac{1}{B(\alpha, \alpha)} \frac{e^{-\alpha y}}{(1 + e^{-y})^{2\alpha}}, \qquad -\infty < y < \infty, \alpha > 0. \tag{23.61}$$

The standard logistic density function in (23.8) corresponds to the case $\alpha = 1$ in (23.61). It is clear that the density in (23.61) is symmetric about 0 for every α. Davidson (1980) showed that the moment-generatine function for this distribution is

$$\begin{aligned} E[e^{\theta Y}] &= \frac{1}{B(\alpha, \alpha)} \int_{-\infty}^{\infty} \frac{e^{-(\alpha - \theta)y}}{(1 + e^{-y})^{2\alpha}} dy \\ &= \frac{\Gamma(\alpha - \theta)\Gamma(\alpha + \theta)}{\{\Gamma(\alpha)\}^2}, \qquad -\alpha < \theta < \alpha. \end{aligned} \tag{23.62}$$

From (23.62) we obtain the mean, variance, and the coefficients of skewness and kurtosis to be

$$\begin{aligned} &E[Y] = 0, \quad \text{var}(Y) = 2\psi'(\alpha), \\ &\sqrt{\beta_1(Y)} = 0, \quad \beta_2(Y) = 3 + \frac{\psi'''(\alpha)}{2\{\psi'(\alpha)\}^2}. \end{aligned} \tag{23.63}$$

Thus the type III distribution has thicker tails than the normal distribution. Further, for large values of α, $\sqrt{2/\alpha}\, Y$ behaves like a standard normal variable.

Gumbel (1944) characterized the type III distribution as the limiting distribution of the αth midrange, $(X_{\alpha:n} + X_{n-\alpha+1:n})/2$, from certain symmetric distributions. Davidson (1980) established that the difference of two independent and identically distributed extreme value random variables is distributed as the type III. Cutler (1992) has shown that the type III distribution arises (asymptotically) in a natural way from statistics based on the kth nearest neighbor distance. George and Ojo (1980) and George, El-Saidi, and Singh (1986) developed an approximation to Student's t-distribution with ν degrees of freedom based on the type III distribution in (23.61). Specifically, by matching the coefficient of kurtosis in (23.63) with that of the Student's t (see Chapter 28), they recommend the use of $\alpha = (\nu - 3.25)/5.5$ as the appropriate type III shape parameter for the required approximation.

The type IV generalized logistic distribution with density function

$$p_Y(y) = \frac{1}{B(p,q)} \frac{e^{-qy}}{(1+e^{-y})^{p+q}}, \qquad -\infty < y < \infty,\ p,q > 0 \quad (23.64)$$

was studied by Prentice (1976) and Kalbfleisch and Prentice (1980). It is readily observed that types I, II, and III are all special cases of this distribution; further, the type IV density function in (23.64) is the density of $-\ln Z$ when qZ/p has a central F-distribution with $(2p, 2q)$ degrees of freedom (see Chapter 27). If Y has a type IV density function as in (23.64), it is of interest to note that $-Y$ also has a type IV distribution with shape parameters p and q interchanged. The moment-generating function of Y is

$$\begin{aligned} E[e^{\theta Y}] &= \frac{1}{B(p,q)} \int_{-\infty}^{\infty} \frac{e^{-(q-\theta)y}}{(1+e^{-y})^{p+q}} dy \\ &= \frac{\Gamma(p+\theta)\Gamma(q-\theta)}{\Gamma(p)\Gamma(q)}, \qquad -p < \theta < q, \end{aligned} \quad (23.65)$$

from which we get

$$E[Y] = \psi(p) - \psi(q), \quad \text{var}(Y) = \psi'(p) + \psi'(q),$$

$$\sqrt{\beta_1(Y)} = \frac{\psi''(p) - \psi''(q)}{\{\psi'(p) + \psi'(q)\}^{3/2}}, \quad \beta_2(Y) = 3 + \frac{\psi'''(p) + \psi'''(q)}{\{\psi'(p) + \psi'(q)\}^2}. \quad (23.66)$$

From (23.66) we observe that the type IV distribution is positively skewed if p > q, negatively skewed if p < q, and symmetric if p = q (becomes type III in this case). George and Ojo (1980) and George and Singh (1987) presented infinite series expressions for the first four cumulants of Y.

Prentice (1976) proposed the type IV distribution as an alternative for modeling binary response data to the usual logistic model. Kalbfleisch and Prentice (1980) considered the four-parameter form of (23.65) (by introducing location and scale parameters) in survival analysis. They have also shown that the type IV distribution goes in limit to lognormal (when $p \to \infty$) and Weibull (when $p = 1$ and $q \to \infty$); also see Farewell and Prentice (1977). The type IV distribution is referred to as the exponential generalized beta of the second type (denoted by EGB2) by McDonald (1991).

All of these generalized logistic distributions can be regarded as special cases of a large class of distributions introduced by Perks (1932). Perks, a British actuary, was primarily interested in obtaining a general function for graduating life-table data, but his formulas are of general applicability.

Perks proposed, as a general form for the probability density function of a random variable, Y, the ratio

$$p_Y(y) = \frac{\sum_{j=0}^{m} a_j e^{-j\theta y}}{\sum_{j=0}^{m'} b_j e^{-j\theta y}}. \tag{23.67}$$

The quantities a_j, b_j, θ are real parameters. There must be relationships among the values of these parameters to ensure that the conditions

$$p_Y(y) \ge 0 \quad \text{and} \quad \int_{-\infty}^{\infty} p_Y(y)\,dy = 1$$

are satisfied. It is always possible to take $\theta = 1$, by a suitable choice of scale, and evidently all a_j's and b_j's can be multiplied by the same (nonzero) constant, without affecting $p_Y(y)$.

A particularly interesting subclass of symmetrical distributions is obtained by taking $m = 1, m' = 2$; $a_0 = 0, b_0 = b_2$. Then (23.67) becomes

$$\begin{aligned} p_Y(y) &= \frac{a_1 e^{-y}}{b_0 + b_1 e^{-y} + b_0 e^{-2y}} \\ &= \frac{c_1}{e^y + c_2 + e^{-y}}, \qquad \text{with } c_1 + \frac{a_1}{b_0}, c_2 = \frac{b_1}{b}. \end{aligned} \tag{23.68}$$

Taking $c_1 > 0$, the condition $p_Y(y) \ge 0$, for all y, requires that $c_2 \ge -2$. The condition $\int_{-\infty}^{\infty} p_Y(y)\,dy = 1$ excludes $c_2 = -2$, but, for all $c_2 > -2$, (23.68) can represent a probability density function. The logistic distribution is obtained by putting $c_2 = 2$, giving (23.8).

Another form of generalized logistic distribution has been considered by Hosking (1989, 1991). The cumulative distribution function for this family is given by

$$F_Y(y) = \frac{1}{1 + (1 - ky)^{1/k}}, \qquad y \le \frac{1}{k} \text{ when } k > 0,\ y \geq \frac{1}{k} \text{ when } k < 0, \tag{23.69}$$

and the probability density function is given by

$$p_Y(y) = \frac{(1 - ky)^{(1/k)-1}}{\{1 + (1 - ky)^{1/k}\}^2}, \qquad y \le \frac{1}{k} \text{ when } k > 0,\ y \geq \frac{1}{k} \text{ when } k < 0. \tag{23.70}$$

The density function above reduces to the standard logistic density function in (23.8) for the case when the shape parameter $k \to 0$. It may be noted from (23.70) that if Y has the above generalized logistic distribution with $k > 0$, then $-Y$ also has the generalized logistic distribution with shape parameter $-k$ (< 0). From (23.69), we obtain the rth raw moment of Y to be

$$\begin{aligned} E[Y^r] &= \int_0^1 \{F_Y^{-1}(u)\}^r \, du \\ &= \int_0^1 \frac{1}{k^r} \cdot \left\{1 - \left(\frac{1-u}{u}\right)^k\right\}^r du \\ &= \frac{1}{k^r} \sum_{i=0}^{r} (-1)^i \binom{r}{i} B(1 - ki, 1 + ki) \\ &= \frac{1}{k^r} \sum_{i=0}^{r} (-1)^i \binom{r}{i} \Gamma(1 + ki)\Gamma(1 - ki) \end{aligned} \tag{23.71}$$

for $r \in (-1/k, 1/k)$. By using the characterizing differential equation

$$(1 - ky)p_Y(y) = F_Y(y)\{1 - F_Y(y)\}, \tag{23.72}$$

Balakrishnan and Sandhu (1994a) established several recurrence relations for the single and the product moments of order statistics. These authors proved

the following specific relationships:

$$E\left[Y_{1:n+1}^{i+1}\right] = \left\{1 + \frac{k(i+1)}{n}\right\} E\left[Y_{1:n}^{i+1}\right] - \frac{i+1}{n} E\left[Y_{1:n}^{i}\right],$$

$$n \geq 1;\ i = 0, 1, \ldots. \quad (23.72a)$$

$$E\left[Y_{r+1:n+1}^{i+1}\right] = E\left[Y_{r:n+1}^{i+1}\right] + \frac{(i+1)(n+1)}{r(n-r+1)} \left\{E\left[Y_{r:n}^{i}\right] - kE\left[Y_{r:n}^{i+1}\right]\right\},$$

$$1 \leq r \leq n;\ i = 0, 1, \ldots. \quad (23.72b)$$

$$E[Y_{r:n+1}Y_{r+1:n+1}] = \frac{n+1}{n-r+1} \left\{\left(1 + \frac{k}{n-r}\right) E[Y_{r:n}Y_{r+1:n}] - \frac{r}{n+1} E\left[Y_{r+1:n+1}^{2}\right] - \frac{1}{n-r} E[Y_{r:n}]\right\},$$

$$(23.72c)$$

$$E[Y_{r:n+1}Y_{s:n+1}] = E[Y_{r:n+1}Y_{s-1:n+1}] + \frac{n+1}{n-s+2} \left\{\left(1 + \frac{k}{n-s+1}\right) E[Y_{r:n}Y_{s:n}] - E[Y_{r:n}Y_{s-1:n}] - \frac{1}{n-s+1} E[Y_{r:n}]\right\},$$

$$1 \leq r < s \leq n;\ s - r \geq 2. \quad (23.72d)$$

$$E[Y_{r+1:n+1}Y_{r+2:n+1}] = \frac{n+1}{r+1} \left\{\frac{1}{r} E[Y_{r+1:n}] + \left(1 - \frac{k}{r}\right) E[Y_{r:n}Y_{r+1:n}] - \frac{n-r}{n+1} E\left[Y_{r+1:n+1}^{2}\right]\right\},$$

$$1 \leq r \leq n - 1. \quad (23.72e)$$

$$E[Y_{r+1:n+1}Y_{s+1:n+1}] = E[Y_{r+2:n+1}Y_{s+1:n+1}] + \frac{n+1}{r+1} \left\{\frac{1}{r} E[Y_{s:n}] + \left(1 - \frac{k}{r}\right) E[Y_{r:n}Y_{s:n}] - E[Y_{r+1:n}Y_{s:n}]\right\},$$

$$1 \leq r < s \leq n;\ s - r \geq 2. \quad (23.72f)$$

$$E[Y_{r:n+1}Y_{r+2:n+1}] = E[Y_{r:n+1}Y_{r+1:n+1}] - r\{E[Y_{r+1:n+1}Y_{r+2:n+1}] - E[Y_{r+1:n+1}^2]\} + \frac{n+1}{n-r}\{E[Y_{r:n}] - kE[Y_{r:n}Y_{r+1:n}]\},$$

$$1 \le r \le n-1. \quad (23.72g)$$

$$E[Y_{r:n+1}Y_{s+1:n+1}] = E[Y_{r:n+1}Y_{s:n+1}] - \frac{r}{s-r}\{E[Y_{r+1:n+1}Y_{s+1:n+1}] - E[Y_{r+1:n+1}Y_{s:n+1}]\} + \frac{n+1}{(n-s+1)(s-r)}\{E[Y_{r:n}] - kE[Y_{r:n}Y_{s:n}]\},$$

$$1 \le r < s \le n;\ s-r \ge 2. \quad (23.72h)$$

When $k \to 0$, these recurrence relations reduce to the results in Eqs. (23.24)–(23.29) due to Shah (1966,1970). Hosking (1989,1991) introduced location and scale parameters into the distribution in (23.69) and studied the probability-weighted moment estimation of the three parameters with relation to maximum likelihood estimates; see Chen and Balakrishnan (1994) for some further remarks in this direction. Balakrishnan and Sandhu (1994a) have discussed the best linear unbiased estimation of parameters.

In considering the truncated form of the generalized logistic distribution in (23.69) and its properties, Balakrishnan and Sandhu (1994b) established similar recurrence relations for the single and the product moments of order statistics. The cumulative distribution function of this truncated distribution is

$$F(y) = \frac{1}{(P-Q)\{1+(1-ky)^{1/k}\}}, \qquad Q_1 \le y \le P_1 \le \frac{1}{k},\ k>0$$

and the probability density function

$$p(y) = \frac{(1-ky)^{(1/k)-1}}{(P-Q)\{1+(1-ky)^{1/k}\}^2}, \qquad Q_1 \le y \le P_1 \le \frac{1}{k},$$

where $Q_1 = 1 - [(1-Q)/Q]^k/k$ and $P_1 = 1 - [(1-P)/P]^k/k$. Here Q (> 0) and 1 − P (> 0) are the proportions of truncation on the left and right of the generalized logistic distribution in (23.69), respectively. The characterizing differential equation for this distribution is given by

$$(1-ky)p(y) = F(y)\{1-(P-Q)F(y)\}$$

which has been exploited by Balakrishnan and Sandhu (1994b) in order to derive several recurrence relations for single and product moments of order statistics.

11 RELATED DISTRIBUTIONS

Talacko (1956) studied the hyperbolic secant distribution, obtained by putting $c_2 = 0$ in (23.68), giving

$$p_Y(y) = \frac{c_1}{e^y + e^{-y}} = \frac{1}{2}c_1 \operatorname{sech} y. \tag{23.73}$$

The condition $\int_{-\infty}^{\infty} p_Y(y)\,dy = 1$ requires that $c_1 = 2/\pi$ so that

$$p_Y(y) = \pi^{-1} \operatorname{sech} y \tag{23.74}$$

and the cumulative distribution function is

$$F_Y(y) = \tfrac{1}{2} + \pi^{-1} \tanh^{-1}(\sinh y). \tag{23.75}$$

Note that if Y has this distribution then e^Y has a half-Cauchy distribution (Chapter 16). The distribution of the sum of n independent random variables, each having the same hyperbolic secant distribution, has been derived by Baten (1934).

Returning for a moment to the more general form of distribution (23.68) the characteristic function $E[e^{itY}]$ is [Talacko (1956)]

$$\frac{\pi}{\cos^{-1}(c_2/2)} \frac{\sinh[t \cos^{-1}(c_2/2)]}{\sinh t\pi} \quad \text{for } -2 < c_2 < 0, 0 < c_2 \leq 2, \tag{23.76}$$

$$\frac{\pi}{\cosh^{-1}(c_2/2)} \frac{\sin[t \cosh^{-1}(c_2/2)]}{\sinh t\pi} \quad \text{for } c_2 > 2. \tag{23.77}$$

The logistic distribution appears as a limiting case, letting $c_2 \to 2$. The values of $\cos^{-1}(c_2/2)$ are taken in the range 0 to π. For the hyperbolic secant distribution, the characteristic function is $\operatorname{sech}(\pi t/2)$, and the rth absolute moment about zero is

$$\begin{aligned}\nu_r' = E[|Y|^r] &= \frac{4}{\pi}\int_0^\infty y^r e^{-y}(1+e^{-2y})^{-1}\,dy \\ &= \frac{4}{\pi}\int_0^\infty \sum_{j=0}^\infty (-1)^j y^r e^{-(2j+1)y}\,dy \\ &= \frac{4}{\pi}\sum_{j=0}^\infty (-1)^j \int_0^\infty y^r e^{-(2j+1)y}\,dy \\ &= \frac{4}{\pi}\Gamma(r+1)\sum_{j=0}^\infty (-1)^j(2j+1)^{-(r+1)}. \qquad (23.78)\end{aligned}$$

The expected value of Y equals zero; $\text{var}(Y) = (4/\pi) \cdot 2 \cdot \pi^3/32 = (1/4)\pi^2$, $\mu_4(Y) = 5\pi^4/15$; $\beta_2(Y) = \alpha_4(Y) = 5$. The mean deviation is $4c/\pi$ (where $c = 0.916$ is Catalan's constant). For this distribution

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{8c}{\pi^2} = 0.742.$$

Harkness and Harkness (1968) have investigated the properties of a class of distributions having characteristic functions

$$(\text{sech}\,\theta t)^\rho, \qquad \rho > 0, \theta > 0,$$

which they term generalized hyperbolic secant distributions. For integer values of ρ the distributions are those of sums of independent identically distributed hyperbolic secant random variables. They have shown that for ρ even ($= 2n$) the density function is

$$p_X(x) = \left[\frac{4^{n-1}x}{(2n-1)!2\theta^2}\cdot \text{cosech}\frac{\pi x}{2\theta}\right]\prod_{j=1}^{n-1}\left(\frac{x^2}{4\theta^2}+j^2\right), \qquad (23.79)$$

while for ρ odd ($= 2n + 1$)

$$p_X(x) = \left[\frac{2^{2n-1}}{(2n)!\theta}\cdot \text{sech}\frac{\pi x}{2\theta}\right]\prod_{j=1}^{n}\left\{\frac{x^2}{4\theta^2}+\left(j-\frac{1}{2}\right)^2\right\}. \qquad (23.80)$$

Fisk (1961) has shown that the Pareto distribution (Chapter 20) can be regarded as a form related to the logistic for certain extreme values of the variable. Making the substitution $e^Y = (T/t_0)^n$ in (23.9) (with $n > 0$, $t_0 > 0$), we have

$$\Pr[T < t] = \Pr\left[(T/t_0)^n < \left(\frac{t}{t_0}\right)^n\right] = (t/t_0)^n\left[1 + \left(\frac{t}{t_0}\right)^n\right]^{-1}.$$

If t be small compared with t_0, then approximately

$$\Pr[T < t] \propto t^n.$$

If t be large compared with t_0, then

$$\Pr[T > t] = \left(\frac{t}{t_0}\right)^{-n}\left[1 + \left(\frac{t}{t_0}\right)^{-n}\right]^{-1}$$

and approximately

$$\Pr[T > t] \propto t^{-n}.$$

Let $\{D_j\}_{j=1}^{\infty}$ be a sequence of independent double exponential random variables (see Chapter *24)* with density function

$$p_{D_j}(d) = \frac{j}{2}e^{-j|d|}, \qquad -\infty < d < \infty, j = 1,2,\dots .$$

Then $\sum_{j=1}^{\infty} D_j$ is distributed as a standard logistic random variable with density in *(23.8)*. Since difference of two i.i.d. exponential random variables is distributed as double exponential, we have $\sum_{j=1}^{\infty}(E_{1j} - E_{2j})$ to be distributed as standard logistic with density in (23.8), where E_{ij}'s are independent exponential random variables with density function

$$p_{E_{ij}}(x) = je^{-jx}, \qquad x \geq 0, j = 1,2,\dots, i = 1,2.$$

Incidentally these two representations prove that the logistic distribution is infinitely divisible.

Galambos and Kotz (1979) provided an interesting joint characterization of the logistic and the exponential distribution as follows: Suppose that X is a continuous random variable with distribution function $G_X(x)$, which is symmetric about 0. Then

$$\Pr[X > -x | X < x] = 1 - e^{-\lambda x}, \qquad x > 0,$$

if and only if $G_X(x) = F_Y(\lambda x)$ with $F_Y(\cdot)$ being the standard logistic distribution function in *(23.9)*.

Baringhaus (1980) presented a characterization result that connects the geometric and the logistic distributions: Let $Z_1, Z_2, \cdots$ be i.i.d. random variables with nondegenerate distribution function $G(z)$, and let N be a positive integer-valued random variable independent of Z's, with ρ the generating function of N. Let $G(z)$ be symmetric about 0, and let γ be a real-valued function of $\theta \in (0, 1)$. Then

$$\frac{\rho(\theta G(z))}{\rho(\theta)} = G(z + \gamma(\theta)), \qquad -\infty < z < \infty,$$

if and only if $G(z) = F_Y(az)$, for some $a > 0$, where $F_Y(\cdot)$ is the standard logistic distribution function in (23.9) and ρ is the generating function of a geometric distribution. Voorn (1987) has generalized this result.

Balakrishnan (1985) considered the folded form of the standard logistic distribution (23.8) and termed it the half ***logistic distribution.*** The density function of this distribution is given by

$$p_X(x) = \frac{2e^{-x}}{(1 + e^{-x})^2}, \qquad x \ge 0, \tag{23.81}$$

and the cumulative distribution function is given by

$$F_X(x) = \frac{1 - e^{-x}}{1 + e^{-x}}, \qquad x \ge 0. \tag{23.82}$$

Use of this distribution as a possible life-time model has been suggested by Balakrishnan (1985) who has established several recurrence relations for the single and the product moments of order statistics. Balakrishnan and Puthenpura (1986) derived the best linear unbiased estimators of the location and scale parameters for the two-parameter half logistic distribution and presented the necessary tables. Balakrishnan and Wong (1991) derived approximate maximum likelihood estimators for the two parameters based on Type-I1 censored samples. Balakrishnan and Chan (1992) considered the scaled half logistic distribution and discussed different methods of inference for the scale parameter of the distribution. Similar to the generalized logistic distribution in (23.69), Balakrishnan and Sandhu (1994c) defined a ***generalized*** half ***logistic distribution*** with distribution function

$$F_X(x) = \frac{1 - (1 - kx)^{1/k}}{1 + (1 - kx)^{1/k}}, \qquad 0 \le x \le \frac{1}{k}, \; k > 0, \tag{23.83}$$

and probably density function

$$p_X(x) = \frac{2(1-kx)^{(1/k)-1}}{\{1+(1-kx)^{1/k}\}^2}, \qquad 0 \le x \le \frac{1}{k}, k > 0. \qquad (23.84)$$

This family includes the half logistic distribution in (23.82) as the shape parameter $k \to 0$. Balakrishnan and Sandhu (1994c) have discussed various properties of this distribution and derived several recurrence relations for the single and the product moments of order statistics generalizing the results of Balakrishnan (1985). These developments have been extended by Balakrishnan and Sandhu (1994d) to the truncated form of the generalized half logistic density function in (23.84) given by

$$p(x) = \frac{2(1-kx)^{(1/k)-1}}{P\{1+(1-kx)^{1/k}\}^2}, \qquad 0 \le x \le P_1, k > 0,$$

where $1 - P$ $(0 < P \le 1)$ is the proportion of truncation on the right of the generalized half logistic density in (23.84), and $P_1 = (1 - [(1-P)/(1+P)]^k)/k$.

Similar to the case of the normal distribution (see Chapter 12, Section 4.3), three translated families of logistic distributions were proposed by Tadikamalla and Johnson (1982a). These three systems of distributions are derived by ascribing the standard logistic distribution in (23.8) to

$$Y = \gamma + \delta \log X, \qquad X > 0 \text{ (for } L_L \text{ system)} \qquad (23.85)$$

$$Y = \gamma + \delta \log\left\{\frac{X}{1-X}\right\}, \qquad 0 < X < 1 \text{ (for } L_B \text{ system)}, \qquad (23.86)$$

and

$$Y = \gamma + \delta \sinh^{-1} X, \qquad -\infty < X < \infty \text{ (for } L_U \text{ system)}. \qquad (23.87)$$

The family of distributions arising from (23.85), called as the log-logistic distributions, were first studied by Shah and Dave (1963). The probability density function is given by

$$p_X(x) = \frac{\delta e^{\gamma} x^{\delta-1}}{(1+e^{\gamma}x^{\delta})^2}, \qquad x \ge 0, \delta > 0. \qquad (23.88)$$

The distributions belong to Burr's type XII family of distributions (see Chapter 12, Section 4.5). Dubey (1966) called them Weibull-exponential distributions and fitted them to business failure data. The density in (23.88) is unimodal; when 6 ≥ 1, the mode is at $x = 0$ (giving a reverse J-shaped curve), and when 6 < 1, the mode is at $x = e^{-\gamma}[(\delta - 1)/(\delta + 1)]$. The cumulative distribution function is

$$F_X(x) = \frac{1}{1 + e^{-\gamma} x^{-\delta}}, \qquad x \geq 0, \delta > 0, \tag{23.89}$$

and the rth raw moment of X is given by

$$E[X^r] = e^{-r\gamma/\delta} \frac{r\pi}{\delta} \operatorname{cosec} \frac{r\pi}{\delta}, \qquad r = 1, 2, \dots . \tag{23.90}$$

Johnson and Tadikamalla (1992) have discussed methods of fitting the four-parameter form of the log-logistic distributions. Shoukri, Mian, and Tracy (1988) examined probability-weighted moment estimators for the three-parameter form of the distributions and compared them with the maximum likelihood estimators. Best linear unbiased estimation of the location and scale parameters (with shape parameter 6 being assumed known) has been discussed by Balakrishnan, Malik, and Puthenpura (1987). Nearly best linear unbiased estimation of the location and scale parameters based on complete as well as singly and doubly Type-II censored samples has been discussed by Ragab and Green (1987). Ali and Khan (1987) and Balakrishnan and Malik (1987) have established some recurrence relations satisfied by the single and the product moments of order statistics from log-logistic and truncated log-logistic distributions.

The density function of X corresponding to the L_B-system transformation in (23.86) is given by

$$p_X(x) = e^{\gamma} \frac{x^{\delta-1}(1-x)^{\delta-1}}{\left\{(1-x)^{\delta} + e^{\gamma} x^{\delta}\right\}^2}, \qquad 0 < x < 1. \tag{23.91}$$

There is a single mode if $\delta > 1$ or an antimode if 6 < 1 at the unique value of x ∈ (0, 1) satisfying the equation

$$e^{\gamma} = \frac{\delta - 1 + 2x}{\delta + 1 - 2x} \left(\frac{1-x}{x}\right)^{\delta}.$$

If 6 > 1, the mode is at $x \lessgtr 1/2$ according as $\gamma \lessgtr 0$. If 6 < 1, the density is U-shaped with the antimode being in $((1 - \delta)/2, 1/2)$ for $\gamma < 0$ and in $(1/2, (1 + \delta)/2)$ for $\gamma > 0$. The cumulative distribution is given by

$$F_X(x) = \frac{1}{1 + e^{-\gamma}[x/(1-x)]^{-\delta}}, \qquad 0 < x < 1. \tag{23.92}$$

The rth raw moment given by

$$E[X^r] = \int_0^1 \left\{1 + e^{\gamma/\delta}(1-u)^{1/\delta}u^{-1/\delta}\right\}^{-r} du \qquad (23.93)$$

needs to be numerically computed. Johnson and Tadikamalla (1992) have discussed methods of fitting this family of distributions.

The density function of X corresponding to the L_U-system transformation in (23.87) is given by

$$p_X(x) = \frac{\delta e^{\gamma}}{\sqrt{x^2+1}} \frac{\left\{x + \sqrt{x^2+1}\right\}^{\delta}}{\left[1 + e^{\gamma}\left\{x + \sqrt{x^2+1}\right\}^{\delta}\right]^2}, \qquad -\infty < x < \infty, \quad (23.94)$$

and the corresponding distribution function is

$$F_X(x) = \frac{1}{1 + e^{-\gamma - \delta \sinh^{-1} x}}, \qquad -\infty < x < \infty. \qquad (23.95)$$

The density in (23.94) is unimodal with mode at the unique value of x satisfying the equation

$$\delta\left[1 - e^{\gamma}\left\{x + \sqrt{x^2+1}\right\}\right] = \frac{x}{\sqrt{x^2+1}}.$$

The rth raw moment of X can be written

$$E[X^r] = \frac{1}{2^r}\sum_{i=0}^{r}(-1)^i\binom{r}{i}e^{-(r-2i)\gamma/\delta}(r-2i)\frac{\pi}{\delta}\operatorname{cosec}(r-2i)\frac{\pi}{\delta} \qquad (23.96)$$

provided that $r < \delta$. If $r \geq \delta$, $E[X^r]$ is infinite. Tadikamalla and Johnson (1982b) have presented tables of δ and γ/δ corresponding to specified values of $\sqrt{\beta_1(X)}$ and $\beta_2(X)$, along with mean and standard deviation of X. Bowman and Shenton (1981) presented formulas from which γ and δ can be determined from the given values of $\sqrt{\beta_1(X)}$ and $\beta_2(X)$. Reference may be made to Johnson and Tadikamalla (1992) for further details. The (β_1, β_2) region for the three transformed logistic distributions are shown in Figure 23.2.

Shah (1963) has discussed mixtures of two logistic distributions. For more details on this and other related distributions, the reader may refer to Balakrishnan (1992).

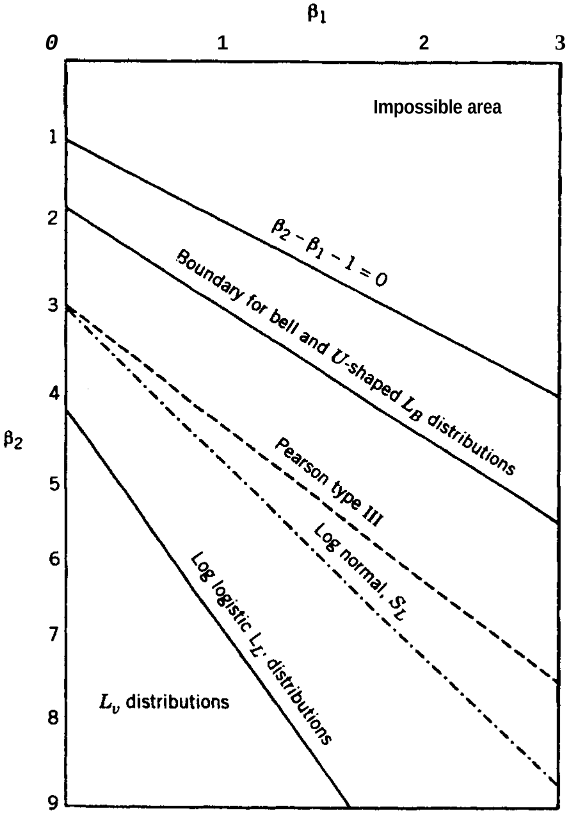


Figure 23.2 β_1, β_2 region for L_U-, L_L-, L_B-distributions.

REFERENCES

Aguirre, N., and Nikulin, M. **(1993).** Chi-squared goodness-of-fit test for the family of logistic distributions, ***Report,*** Institute of Stochastics, University of Bordeaux, France.

Ahmed, A. N., and Abdul-Rahman, A. A. (1993). On characterization of the normal, one sided normal, lognormal and logistic distributions via conditional expectations, ***Pakistan Journal of Statistics, Series B, 9,*** **19–30.**

Ahuja, J. C., and Nash, S. W. **(1967).** The generalized Gompertz-Verhulst family of distributions, ***Sankhyā, Series A, 29,*** **141–156.**

Aitchison, J., and Shen, S. M. **(1980).** Logistic-normal distributions: Some properties and uses, ***Biometrika,*** 67, **261–272.**

Ali, M. Masoom, and Khan, A. H. **(1987).** On order statistics from the log-logistic distribution, ***Journal of Statistical Planning and Inference, 17,*** **103–108.**

Ali, M. Masoom, and Umbach, D. **(1989).** Estimation of quantiles of symmetrically truncated logistic distribution using a few optimally selected order statistics, ***Journal of Information and Optimization Sciences, 10,*** **303–307.**

Anderson, J. A. (1972). Separate sample logistic discrimination, ***Biometrika,*** 59, 19–35.

Anderson, J. A. (1973). Logistic discrimination with medical applications, In ***Discriminant Analysis and Applications*** (ed., *T.* Cacoullos), San Diego, CA. Academic Press, pp. 1–15.

Anderson, J. A. (1974). Diagnosis by logistic discriminant function: Further practical problems and results, ***Applied Statistics,*** 23, 397–404.

Antle, C. E., Klimko, L., and Harkness, W. (1970). Confidence intervals for, the parameters of the logistic distribution, ***Biometrika,*** **57**, 397–402.

Bain, L. J., Balakrishnan, N., Eastman, J. A., Engelhardt, M., and Antle, C. E. (1992). Reliability estimation based on MLEs for complete and censored samples, ***Chapter*** 5 of Balakrishnan, N. (1992).

Balakrishnan, N. (1985). Order statistics from the half logistic distribution, ***Journal of Statistical Computation and Simulation, 20,*** 287–309.

Balakrishnan, N. (1991). Best linear unbiased estimates of the location and scale parameters of logistic distribution for complete and censored samples of sizes 2(1)25(5)40, ***Report,*** McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N. (ed.) (1992). ***Handbook of the Logistic Distribution,*** New York: Dekker.

Balakrishnan, N., Ahsanullah, M., and Chan, P. S. (1994). On logistic record values and associated inference, ***Joumal of Applied Statistical Science*** (to appear).

Balakrishnan, N., and Chan, P. S. (1992). Estimation for the scaled half logistic distribution under Type II censoring, ***Computational Statistics & Data Analysis,*** **13**, 123–141.

Balakrishnan, N., Chan, P. S., Ho, K. L., and Lo, K. K. (1991). Means, variances and covariances of logistic order statistics in the presence of an outlier, ***Report,*** McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N., and Fung, K. Y. (1992). Tolerance limits and sampling plans based on censored samples, ***Chapter*** 14 of Balakrishnan, N. (1992).

Balakrishnan, N., and Joshi, P. C. (1983a). Single and product moments of order statistics from symmetrically truncated logistic distribution, ***Demonstratio Mathematica,*** **16**, 833–841.

Balakrishnan, N., and Joshi, P. C. (1983b). Means, variances, and covariances of order statistics from symmetrically truncated logistic distribution, ***Journal of Statistical Research,*** **17**, 56–61.

Balakrishnan, N., and Kocherlakota, S. (1986). On the moments of order statistics from the doubly truncated logistic distribution, ***Journal of Statistical Planning and Inference,*** **13**, 117–129.

Balakrishnan, N., and Leung, M. Y. (1988a). Order statistics from the Type I generalized logistic distribution, ***Communications in Statistics—Simulation and Computation,*** **17**, 25–50.

Balakrishnan, N., and Leung, M. Y. (1988b). Means, variances and covariances of order statistics, BLUE's for the Type I generalized logistic distribution, and some applications, ***Communications in Statistics—Simulation and Computation,*** **17**, 51–84.

Balakrishnan, N., and Malik, H. J. (1987). Moments of order statistics from truncated log-logistic distribution, ***Journal of Statistical Planning and Inference,*** **17**, 251–267.

Balakrishnan, N., and Malik, H. J. (1994). Means, variances and covariances of logistic order statistics for sample sizes up to fifty, Selected Tables in Mathematical Statistics (to appear).

Balakrishnan, N., Malik, H. J., and Puthenpura, S. (1987). Best linear unbiased estimation of location and scale parameters of the log-logistic distribution, Communications in Statistics—Theory and Methods, **16,** 3477–3495.

Balakrishnan, N., and Puthenpura, S. (1986). Best linear unbiased estimators of location and scale parameters of the half logistic distribution, Journal of Statistical Computation and Simulation, **25,** 193–204.

Balakrishnan, N., and Sandhu, R. A. (1994a). Recurrence relations for single and product moments of order statistics from a generalized logistic distribution, Report, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Sandhu, R. A. (1994b). Relationships for moments of order statistics from a truncated generalized logistic distribution, Report, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Sandhu, R. A. (1994c). Order statistics from a generalized half logistic distribution, Report, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Sandhu, R. A. (1994d). Recurrence relations for single and product moments of order statistics from a truncated generalized half logistic distribution, Report, McMaster University, Hamilton, Canada.

Balakrishnan, N., and Wong, K. H. T. (1991). Approximate MLEs for the location and scale parameters of the half-logistic distribution under Type-I1 right-censoring, IEEE Transactions on *Reliability*, **40,** 140–145.

Baringhaus, L. (1980). Eine simultane Charakterisierung der geometrischen Verteilung und der logistchen Verteilung, Metrika, **27,** 237–242.

Baten, W. D. (1934). The probability law for the sum of n independent variables, each subject to the law $(1/(2h))\text{sech}(\pi x/(2h))$, Bulletin of the American Mathematical *Society*, **40,** 284–290.

Begg, C. B., and Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions, Biometrika, **71,** 11–18.

Berkson, J. (1944). Application of the logistic function to bio-assay, Journal of the American Statistical Association, **39,** 357–365.

Berkson, J. (1951). Why I prefer logits to probits, Biornetrics, 7, 327–339.

Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay and quantal response, based on the logistic function, Journal of the American Statistical Association, **48,** 565–599.

Berkson, J. (1955). Maximum likelihood and minimum χ^2 estimates of the logistic function, Journal of the American Statistical Association, **50,** 130–162.

Berkson, J. (1957). Tables for the maximum likelihood estimates of the logistic function, Biornetrics, **13,** 28–34.

Birnbaum, A., and Dudman, J. (1963). Logistic order statistics, Annals of *Mathematical* Statistics, **34,** 658–663.

Blom, G. (1956). On linear estimates with nearly minimum variances, *Arkiv* for Matematik. **3.** 365–369.

Blom, G. (1958). ***Statistical Estimates and Transformed Beta Variables,*** New York: Wiley.

Bonney, G. E. (1986). Regressive logistic models for familial disease and other binary traits, ***Biornetrics,*** **42,** 611–625.

Bowman, K. O., and Shenton, L. R. (1981). Explicit accurate approximations for fitting the parameters of L_U, In ***Statistical Distributions in Scientific Work,*** vol. 5 (eds., C. Taillie, G. P. Patil, and B. A. Baldessari), Dordrecht: Reidel, pp. 231–240.

Braswell, R. N., and Manders, C. F.-M. (1970a). A new finite range probability distribution function (FRPDF) with parameters–Nomogram and tables, ***Reports in Statistical Applied Research, JUSE,*** **17,** 67–76.

Braswell, R. N., and Manders, C. F.-M. (1970b). On testing and application of a new finite range probability distribution function, ***Reports in Statistical Applied Research, JUSE, 17,*** 77–83.

Braswell, R. N., and Pewitt, T. C. (1973). Generalizations of the finite range probability distribution function of Braswell and Manders, ***Reports in Statistical Applied Research, JUSE,*** **20,** 18–28.

Breslow, N., and Powers, W. (1978). Are there two logistic regressions for retrospective studies? ***Biornetrics,*** 34, 100–105.

Chan, L. K. (1969). Linear quantile estimates of the location and scale parameters of the logistic distribution, ***Statistische Hefte,*** **10,** 277–282.

Chan, L. K., Chan, N. N., and Mead, E. R. (1971). Best linear unbiased estimates of the parameters of the logistic distribution based on selected order statistics, ***Journal of the American Statistical Association, 66,*** 889–892.

Chan, L. K., Chan, N. N., and Mead, E. R. (1973). Tables for the best linear unbiased estimate based on selected order statistics from the normal, logistic, Cauchy and double exponential distributions. ***Mathematics of Computation,*** **27,** 445–446.

Chan, L. K., and Cheng, S. W. (1972). Optimum spacing for the asymptotically best linear estimate of the location parameter of the logistic distribution when samples are complete or censored, ***Statistische Hefte,*** **13,** 41–57.

Chan, L. K., and Cheng, S. W. (1974). An algorithm for determining the asymptotically best linear estimate of the mean from multiply censored logistic data, ***Journal of the American Statistical Association, 69,*** 1027–1030.

Chen, G., and Balakrishnan, N. (1994). The infeasibility of probability weighted moments estimation of some generalized distributions, In ***Recent Advances in Life-testing and Reliability*** (ed., *N*. Balakrishnan), Boca Raton, FL: CRC Press, pp. 565–573.

Cheng, S. W. (1975). A unified approach to choosing optimum quantiles for the ABLE's, ***Journal of the American Statistical Association,*** **70,** 155–159.

Cheng, S. W. (1992). Linear estimation based on selected order statistics, ***Section*** 4.3 of Balakrishnan, N. (1992).

Chew, V. (1968). Some useful alternatives to the normal distribution, ***The American Statistician, 22,*** 22–24.

Cox, D. R. (1966). Some procedures connected with the logistic qualitative response curve, In ***Research Papers in Statistics*** (ed., F. *N*. David), New York: Wiley, pp. 55–71.

Cutler, C. D. (1992). kth nearest neighbors and the generalized logistic distribution, *Section 18.2* of Balakrishnan, N. (1992).

D'Agostino, R. B., and Massaro, J. M. (1992). Goodness-of-fit tests, *Chapter* 13 of Balakrishnan, N. (1992).

D'Agostino, R. B., and Stephens, M. A. (eds.) (1986). *Goodness-of-fit Techniques,* New York: Dekker.

Davidson, R. R. (1980). Some properties of a family of generalized logistic distributions, In *Statistical Climatology, Developments in Atmospheric Science,* vol. 13 (eds., S. Ikeda et al.), Amsterdam: Elsevier.

Day, N. E., and Kerridge, D. F. (1967). A general maximum likelihood discriminant, *Biometrics,* **23,** 313–323.

Downton, F. (1966). Linear estimates with polynomial coefficients, *Biometrika,* **53,** 129–141.

Dubey, S. D. (1966). Transformation for estimation of parameters, *Journal of the Indian Statistical Association, 4,* 109–124.

Dubey, S. D. (1969). A new derivation of the logistic distribution, *Naval Research Logistics Quarterly,* **16,** 37–40.

Dyke, G. V., and Patterson, H. D. (1952). Analysis of factorial arrangements when the data are proportions, *Biornetrics,* 8, 1–12.

Erkelens, J. (1968). A method of calculation for the logistic curve, *Statistica Neerlandica,* **22,** 213–217. *(In* Dutch.)

Farewell, V. T., and Prentice, R. L. (1977). A study of distributional shape in life testing, *Technometrics,* **19,** 69–76.

Finney, D. J. (1947). The principles of biological assay, *Journal of the Royal Statistical Society, Series B,* **9,** 46–91.

Finney, D. J. (1952). *Statistical Method in Biological Assay,* New York: Hafner.

Fisher, R. A., and Yates, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research,* 5th ed., London: Oliver and Boyd; New York: Hafner.

Fisk, P. R. (1961a). Estimation of location and scale parameters in a truncated grouped sech square distribution, *Journal of the American Statistical Association,* **56,** 692–702.

Fisk, P. R. (1961b). The graduation of income distributions, *Econometrica,* **29,** 171–185.

Formann, A. K. (1982). Linear logistic latent class analysis, *Biometrical Journal,* **24,** 171–190.

Galambos, J., and Kotz, S. (1978). *Characterizations of Probability Distributions,* Lecture Notes in Mathematics, No. 675, Berlin: Springer-Verlag.

George, E. O., and Devidas, M. (1992). Some related distributions, *Chapter 10* of Balakrishnan, N. (1992).

George, E. O., El-Saidi, M., and Singh, K. (1986). A generalized logistic approximation of the Student *t* distribution, *Communications in Statistics—Simulation and Computation,* **15,** 1199–1208.

George, E. O., and Mudholkar, G. S. (1981a). A characterization of the logistic distribution by sample median, *Annals of the Institute of Statistical Mathematics,* **33.** 125–129.

George, E. O., and Mudholkar, G. S. (1981b). Some relationships between the logistic and the exponential distributions, In ***Statistical Distributions in Scientific Work,*** vol. 4, (eds., C. Taillie, G. P. Patil, and B. A. Baldessari), Dordrecht: Reidel, pp. 401–409.

George, E. O., and Mudholkar, G. S. (1982). On the logistic and exponential laws, ***Sankhyii, Series A, 44,*** 291–293.

George, E. O., and Mudholkar, G. S. (1983). On the convolution of logistic random variables, ***Metrika,*** 30, 1–13.

George, E. O., and Ojo, M. O. (1980). On a generalization of the logistic distribution, ***Annals of the Institute of Statistical Mathematics,*** **32,** 161–169.

George, E. O., and Rousseau, C. C. (1987). On the logistic midrange, ***Annals of the Institute of Statistical Mathematics,*** **39,** 627–635.

George, E. O., and Singh, K. (1987). An approximation of F distribution by binomial probabilities, ***Statistics & Probability Letters,*** **5,** 169–173.

Gerstenkorn, T. (1992). Estimation of a parameter of the logistic distribution, ***Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Proesses***, Prague: Academia Publishing House of the Czechoslovak Academy of Sciences, pp. 441–448.

Glasbey, C. A. (1979). Correlated residuals in non-linear regression applied to growth data, ***Applied Statistics,*** 28, 251–259.

Goel, P. K. (1975). On the distribution of standardized mean of samples from the logistic population, ***Sankhyii, Series*** B, ***36,*** 165–172.

Greenland, S. (1985). An application of logistic models to the analysis of ordinal responses, ***Biometrical Journal,*** **27,** 189–197.

Grizzle, J. E. (1961). A new method for testing hypotheses and estimating parameters for the logistic model, ***Biometrics***, **17,** 372–385.

Gumbel, E. J. (1944). Ranges and midranges, ***Annals of Mathematical Statistics,*** **15,** 414–422.

Gumbel, E. J. (1958). ***Statistics of Extremes,*** New York: Columbia University Press.

Gumbel, E. J. (1961). Bivariate logistic distributions, ***Journal of the American Statistical Association, 56,*** 335–349.

Gumbel, E. J., and Keeney, R. D. (1950). The extremal quotient, ***Annals of Mathematical Statistics,*** **21,** 523–538.

Gumbel, E. J., and Pickands, J. (1967). Probability tables for the extreme quotient, ***Annals of Mathematical Statistics,*** 38, 1441–1451.

Gupta, S. S., and Balakrishnan, N. (1992). Logistic order statistics and their properties, ***Chapter*** 2 of Balakrishnan, N. (1992).

Gupta, S. S., and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution, ***Biometrika,*** 53, 565–570.

Gupta, S. S., and Han, S. (1992). Selection and ranking procedures for logistic populations, In ***Order Statistics and Nonparametrics: Theory and Applications*** (eds., P. K. Sen and I. A. Salama), Amsterdam: Elsevier, pp. 377–404.

Gupta, S. S., Qureishi, A. S., and Shah, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics, ***Technometrics,*** 9. 43–56.

Perks, W. F. (1932). On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries*, **58**, 12–57.

Plackett, R. L. (1958). Linear estimation from censored data, *Annals of Mathematical Statistics*, **29**, 131–142.

Plackett, R. L. (1959). The analysis of life-test data, *Technometrics*, **1**, 9–19.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves, *Biornetrics*, **32**, 761–768.

Ragab, A., and Green, J. (1987). Estimation of the parameters of the log-logistic distribution based on order statistics, *American Journal of Mathematical and Management Sciences*, *7*, 307–323.

Raghunandanan, K., and Srinivasan, R. (1970). Simplified estimation of parameters in a logistic distribution, *Biometrika*, 57, 677–678.

Rasor, E. A. (1949). The fitting of logistic curves by means of a nomograph, *Journal of the American Statistical Association*, *44*, 548–553.

Reed, L. J., and Berkson, J. (1929). The application of the logistic function to experimental data, *Journal of Physical Chemistry*, 33, 760–779.

Reiersøl, O. (1961). Linear and non-linear multiple comparisons in logit analysis, *Biometrika*, **48**, 359–365.

Saleh, A. K. Md. E., Hassanein, K. M., and Ali, M. Masoom (1992). Estimation of quantiles using selected order statistics, *Section* 4.4 of Balakrishnan, N. (1992).

Sanathanan, L. (1974). Some properties of the logistic model for dichotomous response, *Journal of the American Statistical Association*, **69**, 744–749.

Schafer, R. E., and Sheffield, T. S. (1973). Inferences on the parameters of the logistic distribution, *Biornetrics*, **29**, 445–455.

Schoenberg, I. J. (1953). On Pólya frequency functions and their Laplace transformations, *Journal d'Analyse Mathematique*, *1*, 331–374.

Schultz, H. (1930). The standard error of a forecast from a curve, *Journal of the American Statistical Association*, **25**, 139–185.

Shah, B. K. (1963). A note on method of moments applied to a mixture of two logistic populations, *Journal of the M.S. University of Baroda (Science Number)*, *12*, 21–22.

Shah, B. K. (1965). Distribution of midrange and semirange from logistic population, *Journal of the Indian Statistical Association*, 3, 185–188.

Shah, B. K. (1966). On the bivariate moments of order statistics from a logistic distribution, *Annals of Mathematical Statistics*, 37, 1002–1010.

Shah, B. K. (1970). Note on moments of a logistic order statistics, *Annals of Mathematical Statistics*, *41*, 2151–2152.

Shah, B. K., and Dave, P. H. (1963). A note on log-logistic distribution, *Journal of the M.S. University of Baroda (Science Number)*, **12**, 15–20.

Shoukri, M. M., Mian, I. U. H., and Tracy, D. S. (1988). Sampling properties of estimates of the log-logistic distribution, with application to Canadian precipitation data, *Canadian Journal of Statistics*, **16**, 223–226.

Silverstone, H. (1957). Estimating the logistic curve, *Journal of the American Statistical Association*, **52**, 567–577.

Simpson, J. S. (1967). *Simultaneous linear estimation of the mean and standard deviation of the normal and logistic distributions by the use of selected order* *statistics from doubly censored samples,* M.Sc. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH.

Tadikamalla, P. R., and Johnson, N. L. **(1982a).** Systems of frequency curves generated by transformation of logistic variables, ***Biometrika,*** 69, **461–465.**

Tadikamalla, P. **R.,** and Johnson, N. L. **(1982b).** Tables to facilitate fitting L_U distribution, ***Communications in Statistics—Simulation and Computation,*** *11,* **249–271.**

Talacko, J. **(1956).** Perks' distributions and their role in the theory of Wiener's stochastic variables, ***Trabajos de Estadistica, 7,*** **159–174.**

Tarter, M. E. **(1965).** ***Order statistics moment and product moment relationships for the truncated logistic and other distributions,*** Manuscript, University of Michigan, Ann Arbor.

Tarter, M. E. **(1966).** Exact moments and product moments of the order statistics from the truncated logistic distribution, ***Journal of the American Statistical Association, 61.*** **514–525.**

Tarter, M. E., and Clark, V. A. **(1965).** Properties of the median and other order statistics of logistic variates, ***Annals of Mathematical Statistics, 36,*** **1779–1786.**

Tiku, M. **L. (1968).** Estimating the parameters of normal and logistic distributions from censored samples, ***Australian Journal of Statistics, 10,*** **64–74.**

Tsokos, **C.** P., and **DiCroce,** P. **S. (1992).** Applications in health and social sciences, ***Chapter*** 17 of Balakrishnan, N. **(1992).**

Verhulst, P. **J. (1838).** Notice sur la loi que la population suit dans sons accroissement, ***Corr. Math. et Physique, 10,*** **113–121.**

Verhulst, **P.** J. **(1845).** Recherches **mathématiques** sur la loi d'accroissement de la population, ***Académie de Bruxelles, 18,*** **1–38.**

Vieira, S., and Hoffmann, R. **(1977).** Comparison of the logistic and the Gompertz growth function considering additive and multiplicative error terms, ***Applied Statistics***, 26, **143–148.**

Volodin, N. **(1994).** ***Personal communication.***

Voorn, W. J. **(1987).** Characterization of the logistic and log-logistic distributions by extreme value related stability with random sample sizes, ***Journal of Applied Probability, 24,*** **838–851.**

Wijesinha, A., Begg, C. B., Funkenstein, H. H., and **McNeil,** B. J. **(1983).** Methodology for the differential diagnosis of a complex data set: A case study using data from routine CT-scan examinations, ***Medical Decision Making, 3,*** **133–154.**

Zelterman, D. **(1987a).** Parameter estimation in the generalized logistic distribution, ***Computational Statistics & Data Analysis, 5,*** **177–184.**

Zelterman, D. **(1987b).** Estimation of percentage points by simulation, ***Computational Statistics & Data Analysis, 5,*** **107–125.**

Zelterman, **D. (1989).** Order statistics of the generalized logistic distribution, ***Computational Statistics & Data Analysis, 7,*** **69–77.**

Zelterman, **D.,** and Balakrishnan, N. **(1992).** Univariate generalized distributions, ***Chapter*** 9 of Balakrishnan, N. **(1992).**

CHAPTER 24

Laplace (Double Exponential) Distributions

1 DEFINITION, GENESIS, AND HISTORICAL REMARKS

The double exponential distribution was discovered by Pierre Laplace (1774) as the distribution form for which the likelihood function is maximized by setting the location parameter equal to the median of the observed values of an odd number of independent and identically distributed random variables. This result appeared in Laplace's fundamental paper on symmetric distributions for describing errors of measurement, and it is appropriately known as the ***first law of Laplace***.

The probability density function of the Laplace distribution is

$$p_X(x) = \frac{1}{2\phi}e^{-|x-\theta|/\phi}, \qquad -\infty < x < \infty, \phi > 0. \tag{24.1}$$

Another mode of genesis for this distribution is as the distribution of the difference of two independent and identically distributed exponential random variables.

Incidentally Laplace (1774) proceeded further. He replaced the median by the arithmetic mean as the value maximizing the likelihood function and derived the corresponding distribution to be the normal distribution (see Chapter 13). This result is called as the ***second law of Laplace***. Stigler (1975) has chronicled the many significant contributions made by Laplace and their impact on the subject.

The Laplace distribution in (24.1) is known under different names. One of the most common ones is the ***double exponential.*** It should be mentioned, however, that this name has also been applied to the extreme value distribution (see Chapter 22). A distinction can be made in terminology by calling the

Laplace distribution as double exponential and the extreme value distribution as doubly exponential. In Greenwood, Olkin, and Savage (1962), the Laplace distribution is termed the two-tailed exponential; Feller (1966) referred to it as the bilateral exponential, and Weida (1935) called it as Poisson's first law of error.

Yellott (1977) has very nicely elucidated the relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. Ord (1983) has presented a brief review of various significant developments relating to this distribution.

2 MOMENTS, GENERATING FUNCTIONS, AND PROPERTIES

A standard form of the probability density function (24.1) is obtained by putting $\theta = 0, \phi = 1$, giving

$$p_X(x) = \tfrac{1}{2}e^{-|x|}. \tag{24.2}$$

(This form is sometimes called *Poisson's first* law of error.) The characteristic function corresponding to this probability density function is

$$E[e^{itX}] = \tfrac{1}{2}(1 + it)^{-1} + \tfrac{1}{2}(1 - it)^{-1} = (1 + t^2)^{-1}. \tag{24.3}$$

It is interesting to note that the Fourier transform pair, (24.2) and (24.3), occur in reverse order for the Cauchy distribution (see Chapter 16). The moment-generating function is $(1 - t^2)^{-1}$. The cumulant-generating function is

$$-\log(1 - t^2), \tag{24.4}$$

and the rth cumulant is

$$\kappa_r(X) = \begin{cases} 0 & \text{if } r \text{ is odd,} \\ 2[(r-1)!] & \text{if } r \text{ is even.} \end{cases} \tag{24.5}$$

The rth central moment is

$$\mu_r(X) = \begin{cases} 0 & \text{if } r \text{ is odd,} \\ r! & \text{if } r \text{ is even.} \end{cases} \tag{24.6}$$

Takano (1988) has discussed the Lévy representation of the characteristic function in the d-dimensional Euclidean case (including the case $d = 1$).

The distribution is symmetrical about $x = 0$; the values of the first two moment ratios are

$$\sqrt{\beta_1} = 0, \quad \beta_2 = 6. \tag{24.7}$$

β_2 of 6 reflects the slower rate of decay in the tails of the distribution compared to the normal.

The mean deviation is

$$\nu_1 = E[|X|] = 1. \tag{24.8}$$

So for the **Laplace** distribution

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{1}{\sqrt{2}} \doteqdot 0.707. \tag{24.9}$$

For the more general distribution **(24.1)**, the ratios (24.7) and (24.9) have the same values as for the standard form (24.2). The expected value and standard deviation of **(24.1)** are 8 and $\sqrt{2}\,\phi$, respectively.

The information-generating function is

$$\int_{-\infty}^{\infty} (2\phi)^{-u} \exp\{-u|x-\theta|/\phi\}\, dx = (2\phi)^{1-u} u^{-1}. \tag{24.10}$$

The entropy is $1 + \log(2\phi)$.

The probability density function has a maximum at $x = 6$, where there is a cusp. The form of the function is sketched in Figure 24.1. The cumulative distribution function is

$$F_X(x) = \begin{cases} \dfrac{1}{2}\exp\left[-\dfrac{\theta - x}{\phi}\right], & x \le \theta, \\[2ex] 1 - \dfrac{1}{2}\exp\left[-\dfrac{x-\theta}{\phi}\right], & x \ge \theta. \end{cases} \tag{24.11}$$

The lower and upper quartiles are $8 + \phi \log 2 \doteq \theta \pm 0.6934$.

The probability density function expressed in terms of the expected value, ξ, and the standard deviation, σ, is

$$p_X(x) = (\sigma\sqrt{2})^{-1} \exp\left[-\frac{\sqrt{2}\,|x-\xi|}{\sigma}\right]. \tag{24.12}$$

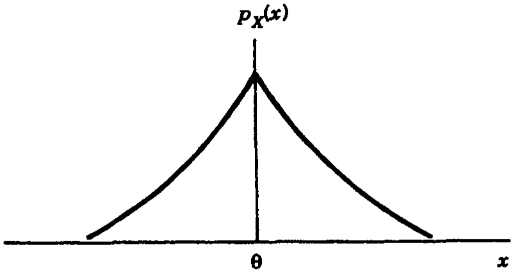


Figure 24.1 Laplace density function.

The upper and lower quartiles are $\xi \pm \sigma \cdot 2^{-1/2} \log 2 = \xi \pm 0.490$~For the normal distribution the corresponding values are $\xi \pm 0.674\sigma$. This difference reflects the sharp peak in the Laplace distribution. For quantiles further in the tails the comparison is reversed because the Laplace probability density function decreases as $\exp[-\sqrt{2}\,|x - \xi|/\sigma]$ while the normal decreases as

$$\exp\left\{-\frac{1}{2}\left[\frac{(x-\xi)^2}{\sigma^2}\right]\right\}.$$

For example the upper and lower 1% points of the Laplace distribution are $\xi \pm 2.722\sigma$, compared with $\xi \pm 2.3260$ for the normal distribution.

It is also easily noted that the Laplace density in (24.1) with $\theta = 0$ is exactly the same as the distribution of $V_1 - V_2$ when V_1 and V_2 are i.i.d. Exponential (ϕ) random variables. Special tables are not needed for numerical calculations connected with the Laplace distribution, as standard tables of the exponential function can be used.

A kurtosis comparison of the Laplace and the Cauchy distributions was made by Balanda (1987). Although β_2 is 6 for the Laplace and infinite for the Cauchy, Balanda pointed out that this moment-based comparison is inadequate, since it does not recognize the dominant features of the two distributions: the Laplace's dramatic peak and the Cauchy's long tail. Horn (1983), for example, identified the Laplace as being more peaked than the Cauchy, while Rosenberger and Gasko (1983) identified the Cauchy as having heavier tails than the Laplace. Since the moment-based orderings are not useful, as none of the Cauchy's moments are finite (see Chapter 161, Balanda (1987) carried out a kurtosis comparison using kurtosis orderings along the lines of van Zwet (1964).

We note that if $\theta = 0$, the probability density function of the arithmetic mean $(\bar{X})$ is

$$p_{\bar{X}}(x) = \frac{n}{\phi^n(n-1)!}\frac{d^{n-1}}{dv^{n-1}}\left\{\frac{e^{-nv|x|}}{(1+\phi v)^n}\right\}\Bigg|_{v=\phi^{-1}}$$

$$= \frac{(n/\phi)e^{-n|x|/\phi}}{2^{2n-1}(n-1)!}\sum_{j=0}^{n-1}\frac{2^j(2n-j-2)!}{j!(n-j-1)!}\left|\frac{nx}{\phi}\right|^j. \qquad (24.13)$$

Cases where the underlying variables are double exponential, in the distribution of the arithmetic mean and the distributions of some other statistics, have been discussed by many authors including Hausdorff (19011, Craig (1932), Weida (1935), and Sassa (1968). The density function of $\bar{X}$ in (24.13) has been used by Balakrishnan and Kocherlakota (1986) in studying the effects of nonnormality on $\bar{X}$-charts as summarized by the true probabilities of false alarm (α) and of true alarm $(1 - \beta)$. These authors have shown that no modification to the control charts is necessary in the case of the double exponential distribution, as both a and $1 - \beta$ attain values close to their normal counterparts.

Sansing (1976) has discussed the t-statistic arising from a double exponential distribution. Gallo (1979) has derived the sampling distributions of the sum of Laplace variables and the sum of absolute values of the variables, in addition to the distribution of the related t-statistic. Dobrogowski (1976) and Findeisen (1982) have discussed some additional properties of the Laplace distribution.

3 ORDER STATISTICS

The simple explicit form of $F_X(x)$, as given in (24.11) leads to simple explicit forms for the distributions of order statistics connected with the Laplace distribution. If $X_1' \le X_2' \le \cdots \le X_n'$ denote the order statistics corresponding to n independent random variables $X_1, X_2, \ldots, X_n$, each having probability density function (24.1)(so that X_r' is the rth smallest of $X_1, X_2, \ldots, X_n$), then the probability density function of X_r' is

$$p_{X_r'}(x) = \begin{cases} \dfrac{n!}{(r-1)!(n-r)!} \cdot \dfrac{1}{2\phi}\left\{1 - \dfrac{1}{2}e^{-(\theta-x)/\phi}\right\}^{n-r} \\ \times \exp\left[-\dfrac{(r+1)(\theta-x)}{\phi}\right] & \text{for } x \le \theta, \\ \dfrac{n!}{(r-1)!(n-r)!} \cdot \dfrac{1}{2\phi}\left\{1 - \dfrac{1}{2}e^{-(x-\theta)/\phi}\right\}^{r-1} \\ \times \exp\left[-\dfrac{(n-r+1)(x-\theta)}{\phi}\right] & \text{for } x \ge 0. \end{cases} \tag{24.14}$$

The sth moment of X_r' about θ is

$$E\left[(X_r' - \theta)^s\right] = \phi^s \frac{n!\Gamma(s+1)}{(r-1)!(n-r)!} \times \left[(-1)^s \sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} 2^{-(r+j+1)}(r+j)^{-(s+1)} + \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} 2^{-(n-r+2+j)} \times (n-r+1+j)^{-(s+1)}\right]. \tag{24.15}$$

In particular, if n is odd, the distribution of the median is obtained by putting $r = (n+1)/2$ in (24.14). This distribution is symmetrical about 0;

the expected value of the median is 8, and the variance is

$$\frac{4\phi^2 n!}{[(n - 11/21!} \sum_{j=0}^{(n-1)/2} (-1)^j$$

$$\times \left[j! \left(\frac{n-1}{2} - j \right) ! 2^{j+(n+1)/2} \left\{ \frac{1}{2}(n+1) + j \right\}^3 \right]^{-1}. \quad (24.16)$$

For any value of n, the expected value of the largest of X,, $X_2, \cdots, X_n$ is

$$E[X_n'] = \theta + \phi n \left[\sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} 2^{-(j+2)}(j+1)^{-2} - 2^{-(1+n)} n^{-2} \right]. \quad (24.17)$$

The expected value of the smallest of X,, $X_2, \cdots, X_n$ is by symmetry 28 – $E(X_n')$, and the expected value of the range W $(= X_n' - X_1')$ is

$$E[W] = 2\phi n \left[\sum_{j=0}^{n-1} (-1)^j \binom{n-1}{j} 2^{-(j+2)}(j+1)^{-2} - 2^{-(1+n)} n^{-2} \right]$$

$$= a_n \phi. \quad (24.18)$$

Edwards (1948) gave the values a, = 2.7708, a, = 3.1771. Edwards also presented the following formulas for the cumulative distribution function of range for $n = 4, 5$: For $n = 4$,

$$F_W(w) = 1 + \tfrac{15}{8} e^{-w} - 3e^{-2w} + \tfrac{1}{8} e^{-3w} - \tfrac{3}{4} w e^{-w}(4 + e^{-w}), \qquad w \ge 0. \quad (24.19)$$

For $n = 5$,

$$F_W(w) = 1 + \tfrac{77}{12} e^{-w} - \tfrac{57}{8} e^{-2w} - \tfrac{1}{4} e^{-3w} - \tfrac{1}{24} e^{-4w} - \tfrac{5}{4} w e^{-w}(4 + 3e^{-w}),$$

$$w \ge 0. \quad (24.20)$$

An alternative interesting way of deriving moments of order statistics from the double exponential distribution was presented by Govindarajulu (1963). His method in fact applies to a general symmetric distribution and is as follows: Let X_1' I X_2' I $\cdot\cdot\cdot$ I X_n' denote the order statistics obtained from a random sample of size n from a symmetric distribution (about zero, without loss of any generality) with cdf $F_X(x)$. Let $Y_{1:n}' \le Y_{2:n}' \le \cdots \le Y_{n:n}'$ denote the order statistics obtained from a random sample of size n from the corresponding folded distribution (folded at zero) with cdf $G_Y(y) = 2F_X(y) - 1$, $y \ge 0$. Then, as Govindarajulu (1963) showed, we have the relations

$$E\left[X_r'^k\right] = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}E\left[Y_{r-i:n-i}'^k\right] + (-1)^k\sum_{i=r}^{n}\binom{n}{i}E\left[Y_{i-r+1:i}'^k\right]\right\}, \qquad 1 \le r \le n, \quad (24.21)$$

and, for $1 \le r < s \le n$,

$$E[X_r'X_s'] = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}E[Y_{r-i:n-i}'Y_{s-i:n-i}'] - \sum_{i=r}^{s-1}\binom{n}{i}E[Y_{i-r+1:i}']E[Y_{s-i:n-i}'] + \sum_{i=s}^{n}\binom{n}{i}E[Y_{i-s+1:i}'Y_{i-r+1:i}']\right\}. \quad (24.22)$$

From (24.21) and (24.22) we take $F_X(x)$ to be the standard double exponential distribution with pdf as in (24.2) and corresponding $G_Y(y)$ to be the standard exponential distribution, and make use of the explicit expressions of the means, variances, and covariances of standard exponential order statistics (see Chapter 18) to obtain

$$E[X_r'] = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_1(r-i, n-i) - \sum_{i=r}^{n}\binom{n}{i}S_1(i-r+1, i)\right\}, \qquad 1 \le r \le n, \quad (24.23)$$

$$E\left[X_r'^2\right] = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_2(r-i, n-i) + \sum_{i=r}^{n}\binom{n}{i}S_2(i-r+1, i)\right\}, \qquad 1 \le r \le n, \quad (24.24)$$

$$E[X_r'X_s'] = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}S_3(r-i, s-i, n-i) - \sum_{i=r}^{s-1}\binom{n}{i}S_1(i-r+1, i)S_1(s-i, n-i) + \sum_{i=s}^{n}\binom{n}{i}S_3(i-s+1, i-r+1, i)\right\}, \qquad 1 \le r < s \le n. \quad (24.25)$$

In these formulas, for $1 \leq r \leq n$,

$$S_1(r,n) = \sum_{i=n-r+1}^{n} \frac{1}{i},$$

$$S_2 = \sum_{i=n-r+1}^{n} \frac{1}{i^2} + (S_1(r,n))^2,$$

and for $1 \leq r < s \leq n$,

$$S_3(r,s,n) = \sum_{i=n-r+1}^{n} \frac{1}{i^2} + S_1(r,n) \, . \, S_1(s,n).$$

Using Eqs. (24.23)–(24.25) Govindarajulu (1966) tabulated the means, variances, and covariances of standard double exponential order statistics for sample sizes up to 20.

Recently Balakrishnan, Govindarajulu, and Balasubramanian (1993) have given an elegant probabilistic interpretation for the results in (24.21) and (24.22), and used it to establish some generalizations.

Govindarajulu's results in (24.21) and (24.22) were extended by Balakrishnan (1988) to the case when the order statistics arise from a single-scale–outlier model consisting of $n - 1$ symmetric variables and one single symmetric scale outlier. These results were used, along with explicit expressions for the single and the product moments of order statistics from a single-scale–outlier exponential model [Barnett and Lewis (1994)], by Balakrishnan and Ambagaspitiya (1988) to study the robustness features of various linear estimators of the parameters 8 and ϕ of the Laplace distribution (24.1). Balakrishnan's (1988) results for the single-scale–outlier double exponential model have been extended in Balakrishnan (1989) to the case when the order statistics $X_1', X_2', \cdots, X_n'$ arise from n independent and nonidentically distributed Laplace random variables. Akahira and Takeuchi (1990) have discussed the loss of information associated with the order statistics from the double exponential distribution and related estimators of the parameters θ and ϕ (see Section 4).

Lien, Balakrishnan, and Balasubramanian (1992) discussed the moments of order statistics from a doubly truncated Laplace distribution with density function

$$p_X(x) = \frac{1}{2(1-P-Q)} e^{-|x|}, \qquad \log(2Q) \leq x \leq -\log(2P), \quad (24.26)$$

where Q and P are the proportions of truncation on the left and right of the standard Laplace density (24.2). They used these moments of order statistics to derive best linear unbiased estimators of the location and scale parameters of a doubly truncated Laplace distribution. They also developed some results for order statistics from a general nonoverlapping mixture model from which the results for the doubly truncated Laplace distribution in (24.26) are deduced as special cases. Khan and Khan (1987) have also derived some recurrence relations for the moments of order statistics from the doubly truncated Laplace distribution (24.26).

4 METHODS OF INFERENCE

4.1 Maximum Likelihood Estimation

Given observed values of n mutually independent random variables $X_1, \cdots, X_n$, each with probability density function (24.1), the likelihood function is

$$-n\log(2\phi) - \phi^{-1}\sum_{j=1}^{n}|X_j - \theta|. \tag{24.27}$$

Whether the value of 4 is known or not, any value $\hat{\theta}$ minimizing $\sum_{j=1}^{n}|X_j - \theta|$ with respect to 8 is a maximum likelihood estimator of θ. If n is odd, then $\hat{8}$ is uniquely defined as the median of $X_1, X_2, \cdots, X_n$. This result was obtained by Keynes (1911), who also conjectured that this is a characterization of the Laplace distribution (as, indeed, it effectively is). If n is even, then $\hat{\theta}$ can be any value between the $\frac{1}{2}n$th and $(\frac{1}{2}n + 1)$th-order statistics from $X_1, X_2, \cdots, X_n$. The arithmetic mean of these two values is convenient to use, and is an unbiased estimator of θ (as is the median when n is odd).

If 4 (as well as θ) is unknown, a maximum likelihood estimator of ϕ is

$$n^{-1}\sum_{j=1}^{n}|X_j - \hat{\theta}|, \tag{24.28}$$

where $\hat{\theta}$ is a maximum likelihood estimator of θ. If 8 is known, but ϕ is unknown (most commonly, 8 = 0), then the maximum likelihood estimator of ϕ is

$$n^{-1}\sum_{j=1}^{n}|X_j - \theta|. \tag{24.29}$$

The distribution of the median has been discussed in Section 24.3. Although the median is a maximum likelihood estimator of 8, and unbiased, it is not a minimum variance unbiased estimator of 8. Indeed for small values of n (the sample size) it is possible to construct unbiased estimators with smaller variance than the median (e.g., see Table 24.1). Norton (1984) and Hombas (1986) have described the use of calculus to find the maximum likelihood estimators.

Table 24.1 Coefficients and variances of best linear estimators of location (θ) and scale (ϕ) parameters

n	r		Coefficients of X'_n	X'_{n-1}	X'_{n-2}	X'_{n-3}	X'_{n-4}	Variances
2	0	θ	0.5000					1.0000
		ϕ	0.6667					0.7778
3	0	θ	0.1481	0.7037				0.5895
		ϕ	0.4444	0.0000				0.4321
4	0	θ	0.0473	0.4527				0.4155
		ϕ	0.3077	0.2145				0.2986
4	1	θ		0.5000				0.4201
		ϕ		1.4545				0.8512
5	0	θ	0.0166	0.2213	0.5241			0.3169
		ϕ	0.2331	0.2264	0.0000			0.2290
5	1	θ		0.2378	0.5244			0.3174
		ϕ		0.8727	0.0000			0.4387
6	0	θ	0.0063	0.1006	0.3931			0.2548
		ϕ	0.1876	0.1943	0.1132			0.1858
6	1	θ		0.1069	0.3931			0.2548
		ϕ		0.6135	0.1824			0.2996
6	2	θ			0.5000			0.2609
		ϕ			2.2857			0.8866
7	0	θ	0.0025	0.0455	0.2386	0.4267		0.2122
		ϕ	0.1572	0.1631	0.1439	0.0000		0.1565
7	1	θ		0.0480	0.2386	0.4267		0.2122
		ϕ		0.4677	0.2104	0.0000		0.2288
7	2	θ			0.2862	0.4276		0.2134
		ϕ			1.3061	0.0000		0.4468
8	0	θ	0.0010	0.0208	0.1316	0.3465		0.1814
		ϕ	0.1355	0.1391	0.1391	0.0718		0.1351
8	1	θ		0.0219	0.1316	0.3465		0.1814
		ϕ		0.3767	0.1910	0.0987		0.1856
8	2	θ			0.1533	0.3467		0.1816
		ϕ			0.1977	0.1605		0.3020
8	3	θ				0.5000		0.1873
		ϕ				3.1411		0.9078
9	0	θ	0.0004	0.0097	0.0698	0.2374	0.3654	0.1581
		ϕ	0.1191	0.1211	0.1251	0.1013	0.0000	0.1190
9	1	θ		0.0101	0.0698	0.2374	0.3654	0.1581
		ϕ		0.3153	0.1643	0.1331	0.0000	0.1562
9	2	θ			0.0799	0.2374	0.3655	0.1581
		ϕ			0.7023	0.1955	0.0000	0.2295
9	3	θ				0.3166	0.3668	0.1596
		ϕ				1.7451	0.0000	0.4534
10	0	θ	0.0002	0.0046	0.0364	0.1478	0.3110	0.1399
		ϕ	0.1063	0.1074	0.1110	0.1061	0.0504	0.1062
10	1	θ		0.0047	0.0364	0.1478	0.3310	0.1399
		ϕ		0.2714	0.1410	0.1347	0.0640	0.1350
10	2	θ			0.0412	0.1478	0.3110	0.1399
		ϕ			0.5665	0.1854	0.0881	0.1857
10	3	θ				0.1887	0.3113	0.1403
		ϕ				1.2218	0.1448	0.3044
10	4	θ					0.5000	0.1452
		ϕ					4.0125	0.9220

Balakrishnan and Cutler (1994) have recently derived explicitly the **maximum** likelihood estimators of the parameters 8 and ϕ based on **symmetrically** Type-I1 censored samples. To this end, let $X'_{r+1} \le X'_{r+2} \le \dots \le X'_{n-r}$ be the symmetrically Type-I1 censored sample available from a sample of size n, where the smallest r and the largest r observations have been censored. Then the likelihood function based on the given censored sample is

$$L(\theta, \phi) = \frac{n!}{(r!)^2} [F_X(X'_{r+1})\{1 - F_X(X'_{n-r})\}]^r \prod_{i=r+1}^{n-r} p_X(X'_i), \quad (24.30)$$

where $p_X(x)$ and $F_X(x)$ are as in Eqs. (24.1) and (24.11). If θ is in the interval $[X'_{r+1}, X'_{n-r}]$, the likelihood function in (24.30) becomes

$$L(\theta, \phi) = \frac{n!}{2^n (r!)^2 \phi^{n-2r}} \exp\left\{ \frac{-r}{\phi}(X'_{n-r} - X'_{r+1}) - \sum_{i=r+1}^{n-r} \left| \frac{X'_i - \theta}{\phi} \right| \right\}, \quad (24.31)$$

from which it is clear that the "restricted MLE" of θ is

$$\hat{\theta} = \begin{cases} X'_{m+1} & \text{when } n = 2m + 1, \\ \text{any value in } [X'_m, X'_{m+1}] & \text{when } n = 2m. \end{cases} \quad (24.32)$$

If $\theta < X'_{r+1}$, the likelihood function in (24.30) becomes

$$L(\theta, \phi) = \frac{n!}{(r!)^2 (2\phi)^{n-2r}} \left[\left\{ 1 - \frac{1}{2} e^{-(X'_{r+1} - \theta)/\phi} \right\} \frac{1}{2} e^{-(X'_{n-r} - \theta)/\phi} \right]^r \times e^{-\sum_{i=r+1}^{n-r} (X'_i - \theta)/\phi}, \quad (24.33)$$

which can be shown to be a monotonic increasing function in 8. If $\theta > X'_{n-r}$, the likelihood function in (29.30) becomes

$$L(\theta, \phi) = \frac{n!}{(r!)^2 (2\phi)^{n-2r}} \left[\frac{1}{2} e^{(X'_{r+1} - \theta)/\phi} \cdot \left\{ 1 - \frac{1}{2} e^{(X'_{n-r} - \theta)/\phi} \right\} \right]^r \times e^{-\sum_{i=r+1}^{n-r} (\theta - X'_i)/\phi}, \quad (24.34)$$

which can be shown to be a monotonic decreasing function in 8.

Consequently, the "restricted MLE" of θ in (24.32) becomes the global MLE as well. Now, substituting $\hat{\theta}$ in (24.30) and maximizing $L(\hat{\theta}, \phi)$ with respect to ϕ, we derive the MLE of ϕ as, when n = 2m + 1,

$$\hat{\phi} = \frac{1}{2m+1-2r}\left\{\sum_{i=m+2}^{2m+1-r} X_i' - \sum_{i=r+1}^{m} X_i' + r\{X_{2m+1-r}' - X_{r+1}'\}\right\}, \quad (24.35)$$

and, when n = $2m$,

$$\hat{\phi} = \frac{1}{2m-2r}\left[\sum_{i=m+1}^{2m-r} X_i' - \sum_{i=r+1}^{m} X_i' + r\{X_{2m-r}' - X_{r+1}'\}\right]. \quad (24.36)$$

Balakrishnan and Cutler (1994) have discussed the bias and the efficiencies of these estimators as compared to the best linear unbiased estimators presented by Govindarajulu (1966) and described below. Balakrishnan and Cutler (1994) have also derived such explicit MLEs for 8 and 4 based on Type-II right-censored samples.

4.2 Best Linear Unbiased Estimation

Let $X_{r+1}' \le X_{r+2}' \le \cdots \le X_{n-s}'$ be the doubly Type-II censored sample available from a sample of size n, where the smallest r and the largest s observations have been censored. From Eqs. (24.23)–(24.25), we can compute the means, variances, and covariances of order statistics from the standard Laplace distribution (24.2), and we denote them by μ_i, σ_{ii}, and σ_{ij}. Further let us denote

$$\mathbf{X} = (X_{r+1}', X_{r+2}', \cdots, X_{n-s}')^T,$$

$$\boldsymbol{\mu} = (\mu_{r+1}, \mu_{r+2}, \cdots, \mu_{n-s})^T,$$

$$\mathbf{1} = (1, 1, \cdots, 1)^T_{(n-r-s)\times 1},$$

$$\boldsymbol{\Sigma} = ((\sigma_{ij}))_{i,j=r+1,\dots,n-s}.$$

Then the best linear unbiased estimators of 8 and ϕ, based on the given doubly Type-II censored sample, are given by [see David (1981), Balakrishnan and Cohen (1991, pp. 80–82)]

$$\begin{aligned}\theta^* &= \left\{\frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\mathbf{1}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}\right\}\mathbf{X}\\ &= \sum_{i=r+1}^{n-s} a_i X_i' \end{aligned} \quad (24.37)$$

and

$$\phi^* = \left\{ \frac{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} - \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2} \right\} \mathbf{X}$$

$$= \sum_{i=r+1}^{n-s} b_i X_i'. \tag{24.38}$$

The variances and covariance of these estimators are given by

$$\operatorname{var}(\theta^*) = \phi^2 \left\{ \frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2} \right\} = \phi^2 V_1, \tag{24.39}$$

$$\operatorname{var}(\phi^*) = \phi^2 \left\{ \frac{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2} \right\} = \phi^2 V_2, \tag{24.40}$$

$$\operatorname{cov}(\theta^*, \phi^*) = -\phi^2 \left\{ \frac{\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}{(\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}) - (\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2} \right\} = -\phi^2 V_3. \tag{24.41}$$

For the case where the available sample is symmetrically Type-II censored (i.e., $r = s$), $\operatorname{cov}(\theta^*, \phi^*)$ in (24.41) will equal 0, since $\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} = 0$. Further in this case the coefficient of X_i' in θ^* in (24.37) is the same as that of X_{n-i+1}'; the coefficient of X_i' in ϕ^* in (24.38) is the same as that of X_{n-i+1}' in magnitude but is of opposite sign. Govindarajulu (1966) presented tables of the coefficients a_i and b_i and the values of V_1, V_2, and V_3 for sample sizes up to 20 and all possible choices of $r = s$. Balakrishnan, Chandramouleeswaran, and Ambagaspitiya (1994) presented similar tables for Type-II right-censored samples of sizes up to 20 with $r = 0$ and $s = 0(1)n - 2$.

Table 24.1 [from Govindarajulu (1966)] gives the coefficients of θ^* and ϕ^* in samples of sizes $n = 2(1)10$ and $r = s = 0(1)[(n - 2)/2]$. The final column gives the values of $\operatorname{var}(\theta^*)/\phi^2$ and $\operatorname{var}(\phi^*)/\phi^2$.

Sarhan (1954) has compared the variances of the best linear estimator of θ, the median (defined as the arithmetic mean of $X_{n/2}'$ and $X_{(n+1)/2}'$ when n is even), the arithmetic mean $(n^{-1}\sum_{j=1}^{n} X_j')$, and the midrange $[\frac{1}{2}(X_1' + X_n')]$. These are all unbiased estimators of θ. Table 24.2 presents the efficiencies (inverse ratio of variances, expressed as a percentage) of the last three estimators relative to the first. Figure 24.2*a–c* represent these values diagrammatically. The irregular appearance of Figure 24.2*c* is associated with the different definition of the median in samples of odd and even sizes.

We note that the estimator $n^{-1}\sum|X_i - \theta|$ of ϕ (with θ known) is distributed as $(2n)^{-1}\phi \times (\chi^2$ with $2n$ degrees of freedom). The distribution of

Table 24.2 Efficiency of various estimators of 8, relative to best linear unbiased estimator

Estimator (%)	Sample Size n			
	2	3	4	5
Arithmetic mean	100.00	88.43	82.80	79.21
Midrange	100.00	67.90	49.65	38.29
Median	100.00	92.27	98.90	90.23

Note: Chu and Hotelling (1955) showed that var(median) is less than var(arithmetic mean) for $n \geq 7$.

$n^{-1}\Sigma|X_i - \hat{\theta}|$, where $\hat{\theta}$ is a median value, has been studied by Karst and Polowy (1963).

In Table 24.3 the coefficients for the best linear unbiased estimators of the mean 9 and the standard deviation $\sigma = \phi\sqrt{2}$ and the values of $\text{var}(\theta^*)/\sigma^2$ and $\text{var}(\sigma^*)/\sigma^2$ [taken from Sarhan (1954)] are presented for $n = 3$, 4, and 5 for the case of right censoring with $r = 0$ and $s = 1(1)n - 2$. As we mentioned earlier, Balakrishnan, Chandramouleeswaran, and Ambagaspitiya (1994) present an extended form of this table for samples sizes up to 20.

If ϕ is known, confidence limits for 9 may be based on the distribution of the median, $\hat{\theta}$. If θ is known, then confidence limits for ϕ may be obtained by using the fact that the distribution of $n^{-1}\Sigma_{j=1}^{n}|X_j - \theta|$ is that of $(2n)^{-1}\phi$ X (χ^2 with 2n degrees of freedom). The limits of a $100(1 - \alpha)\%$ confidence

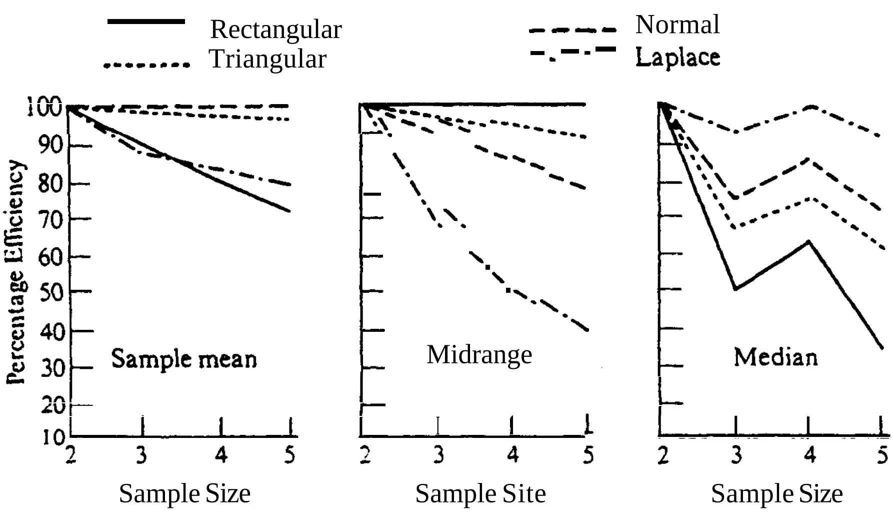


Figure 24.2 Percentage efficiencies of the sample mean, midrange. and median in different populations.

Table 24.3 Coefficients of best linear unbiased estimators of expected value (θ) and standard deviation ($\sigma = \phi\sqrt{2}$) for samples censored by omission of s largest observed values

n	s	Estimator of	Coefficients of X'_1	X'_2	X'_3	X	Variance $\times\sigma^{-2}$	Relative Efficiency[a]
3	1	θ	−0.3300	1.3000				
		σ	−1.3578	1.3578				
4	1	θ	0.0662	0.3333	0.6004		0.1860	(98.23)
		σ	−0.7332	−0.2129	0.9461		0.3339	(89.42)
	2	θ	0.0000	1.0000			0.3335	(62.29)
		σ	−1.2563	1.2563			0.9457	(31.58)
5	1	θ	0.0114	0.2163	0.5243	0.2479	0.1586	(99.88)
		σ	−0.4331	−0.4191	0.0037	0.8484	0.3097	(73.88)
	2	θ	−0.6649	0.1666	0.8998		0.1724	(91.85)
		σ	−0.6655	−0.6233	1.2889		0.4634	(49.37)
	3	θ	−0.5641	1.5641			1.2743	(2.24)
		σ	−1.3925	1.3925			2.8481	(8.03)

"Relative efficiency (inverse ratio of variance to that of best linear unbiased estimator using the complete sample) is shown, as a percentage, in parentheses.

interval for 4 are then

$$2\sum_{j=1}^{n} \frac{|X_j-\theta|}{\chi^2_{2n,1-\alpha/2}} \quad \text{and} \quad 2 \sum_{j=1} \frac{|X_j-\theta|}{\chi^2_{2n,\alpha/2}}. \tag{24.42}$$

If neither 8 nor ϕ is known it would be possible to construct confidence intervals for 8 and ϕ, respectively, using the distributions of

$$\frac{\hat{\theta} - \mathrm{e}}{\sum_{j=1}^{n}|X_j-\hat{\theta}|} \quad \text{and} \quad \phi^{-1}\sum_{j=1}^{n}|X_j-\hat{\theta}|, \tag{24.43}$$

which are pivotal quantities for the parameters θ and 4, respectively. Bain and Engelhardt (1973) have determined exact distributions for $n = 3$ and $n = 5$ and have provided approximate distributions for larger n. These authors have also given the asymptotic distributions of the pivotal quantities and the powers of the associated tests of hypotheses.

For the case of complete as well as Type-II censored samples, Balakrishnan, Chandramouleeswaran, and Arnbagaspitiya (1994) considered three pivotal quantities by using the best linear unbiased estimators θ^* and ϕ^* in Eqs. (24.37) and (24.38) and their variances in Eqs. (24.39) and (24.40),

$$\frac{\theta^*-\theta}{\phi\sqrt{V_1}}, \quad \frac{\mathrm{e}^*-\mathrm{e}}{\phi^*\sqrt{V_1}}, \quad \text{and} \quad \frac{(\phi^*/\phi)-1}{\sqrt{V_2}} \tag{24.44}$$

and making inferences on 8 when ϕ is known, on θ when ϕ is unknown, and on ϕ when θ is unknown, respectively. These authors presented some percentage points of all three pivotal quantities for sample sizes up to 20 for various choices of censoring. Edgeworth approximations for the distributions of the pivotal quantities in (24.44) have been discussed by Balakrishnan, Chandramouleeswaran, and Govindarajulu (1994) who have also examined their accuracy.

Referring to ***the tables*** of ***BLUEs given*** by ***Govindarajulu (1966), Srini-*** vasan and Wharton (1982) discussed the derivation of one-sided and two-sided confidence bands on the entire cumulative distribution function $F_X(x)$. These bands are constructed using the Kolmogorov–Smirnov–type statistics. For example, the two-sided band on $F_X(x;\theta,\phi)$ is based on the statistic

$$L_n = \sup_{-\infty<x<\infty} |F_X(x;\theta,\phi) - F_X(x;\theta^*,\phi^*)|, \qquad (24.45)$$

while the one-sided upper confidence contour for $F_X(x;\theta,\phi)$ is based on the statistic

$$L_n^+ = \sup_{x\geq 0}\{F_X(x;\theta,\phi) - F_X(x;\theta^*,\phi^*)\}. \qquad (24.46)$$

For a $\in (0, 1)$, if 1, is the ath quantile of L, (i.e., $\Pr[L_n \leq l_\alpha]$ = a), then a two-sided confidence band for $F_X(x;\theta,\phi)$ with confidence level a is given by the planar region bounded above by the function y = $\min\{F_X(x;\theta^*,\phi^*) +$ I„ 1) and bounded below by y = $\max\{F_X(x;0^*,\phi^*) - l_\alpha, 0\}$. Srinivasan and Wharton (1982) have presented tables of simulated percentage points of L_n and L_n^+ for n up to 20; these are presented in Tables 24.4 and 24.5. Srinivasan and Wharton (1982) have also discussed some large-sample approximations for the percentage points of L_n and L_n^+. For example, by using the asymptotic result that $Z_1 = \sqrt{n}\,\theta^*$ and $Z_2 = \sqrt{n}\,(\phi^* - 1)$ are independent standard normal variables (in the case of the standard Laplace distribution), they have shown that the limiting distribution of $\sqrt{n}\,L_n$ is the same as that of the random variable $\sup|X_0(y)|$, where

$$X_0(y) = \tfrac{1}{2}e^{-|y|}(U + Vy), \qquad -\infty < y < \infty.$$

This expression readily gives approximate quantiles for $\sqrt{n}\,L_n$ when n is large. Srinivasan and Wharton (1982) have mentioned that this asymptotic approximation works quite effectively for $n > 30$.

By considering just the location-Laplace model (with $\phi = 1$), Sugiura and Naing (1989) derived improved estimators of 8 in the form of a weighted linear combination of the sample median and pairs of order statistics (with symmetric distance to both sides from the sample median) and by minimizing with respect to weights and distances. The resulting estimator has been

Table 24.4 Simulated percentage points l_α of the statistic L_n

n	α 0.80	0.85	0.90	0.95	0.99
5	0.31	0.35	0.39	0.45	0.56
6	0.29	0.32	0.35	0.41	0.52
7	0.26	0.29	0.33	0.38	0.48
8	0.25	0.27	0.31	0.36	0.46
9	0.23	0.26	0.29	0.34	0.44
10	0.22	0.24	0.27	0.32	0.41
11	0.21	0.23	0.26	0.31	0.39
12	0.20	0.22	0.25	0.30	0.38
13	0.19	0.22	0.24	0.28	0.36
14	0.18	0.21	0.23	0.27	0.34
15	0.18	0.20	0.22	0.26	0.33
16	0.17	0.19	0.22	0.25	0.32
17	0.16	0.18	0.21	0.24	0.31
18	0.16	0.18	0.20	0.24	0.31
19	0.16	0.18	0.20	0.23	0.31
20	0.15	0.17	0.19	0.23	0.29

Table 24.5 Simulated percentage points l_α^+ of the statistic L_n^+

n	α 0.80	0.85	0.90	0.95	0.99
5	0.23	0.27	0.31	0.38	0.51
6	0.21	0.24	0.29	0.35	0.47
7	0.19	0.22	0.26	0.32	0.44
8	0.18	0.21	0.25	0.31	0.42
9	0.16	0.19	0.23	0.28	0.39
10	0.16	0.18	0.22	0.27	0.38
11	0.15	0.17	0.21	0.26	0.36
12	0.14	0.17	0.20	0.25	0.34
13	0.13	0.16	0.19	0.24	0.34
14	0.13	0.15	0.18	0.23	0.32
15	0.12	0.14	0.18	0.22	0.30
16	0.12	0.14	0.17	0.21	0.29
17	0.12	0.14	0.17	0.21	0.28
18	0.12	0.14	0.17	0.21	0.28
19	0.11	0.13	0.16	0.20	0.27
20	0.11	0.13	0.15	0.19	0.26

shown to have smaller asymptotic variance in the second order; see also Akahira (1987,1990) and Akahira and Takeuchi (1993).

4.3 Simplified Linear Estimation

By considering the ith quasi-range, $W_i = X'_{n-i+1} - X'_i$, and the ith quasi-midrange, $V_i = (X'_{n-i+1} + X'_i)/2$, Raghunandanan and Srinivasan *(1971)* proposed simplified linear estimators of the parameters ϕ and 8. Their estimator of $\theta, \tilde{\theta}$, is defined to be that V_i with the smallest variance. The estimator $\tilde{\theta}$ is presented in Table *24.6* for n up to 20, and the efficiency of this estimator relative to the best linear unbiased estimator of 8 based on the complete sample of size n (Table *24.1)* is also presented in this table. The estimator 8 presented in the table is also applicable when the available sample is symmetrically Type-I1 censored with $r \le i - 1$. We may note that for $n = 3$ and 5, the estimator in Table *24.6* is simply the sample median.

Raghunandanan and Srinivasan's estimator of ϕ, which is based on a symmetrically Type-I1 censored sample with r smallest and r largest observations censored, is defined to be the one with minimum variance among linear estimators of the form

$$\tilde{\phi} = D \sum_{i=r+1}^{n/2} c_i W_i, \tag{24.47}$$

Table 24.6 Estimator $\tilde{\theta}$ and its efficiency

n	i	$\text{var}(\tilde{\theta})/\phi^2$	$\text{Eff}(\tilde{\theta})$
3	2	0.638890	0.923
4	2	0.420135	0.989
5	3	0.351180	0.902
6	3	0.260905	0.977
7	3	0.225805	0.940
8	4	0.187310	0.968
9	4	0.164795	0.959
10	5	0.145225	0.963
11	5	0.129605	0.967
12	6	0.118125	0.960
13	6	0.106670	0.970
14	7	0.099285	0.959
15	7	0.090540	0.972
16	7	0.085190	0.960
17	8	0.078575	0.972
18	8	0.074175	0.967
19	9	0.069350	0.973
20	9	0.065670	0.970

Table 24.7 Estimator $\tilde{\phi}$ and its efficiency

n	r	$\tilde{\phi}$	$\text{var}(\tilde{\phi})/\phi^2$	$\text{Eff}(\tilde{\phi})$
4	0	$0.289157\,(W_1 + W_2)$	0.300624	0.993
5	0	$0.231325\,(W_1 + W_2)$	0.229000	1.000
6	0	$0.183486\,(W_1 + W_2 + W_3)$	0.186515	0.996
6	1	$0.666667\,W_2$	0.304009	0.985
7	0	$0.157274\,(W_1 + W_2 + W_3)$	0.156500	1.000
7	1	$0.390721\,(W_2 + W_3)$	0.234731	0.975
8	0	$0.134254\,(W_1 + W_2 + W_3 + W_4)$	0.135438	0.997
8	1	$0.324571\,(W_2 + W_3)$	0.188570	0.984
8	2	$0.967133\,W_3$	0.303726	0.994
9	0	$0.119337\,(W_1 + W_2 + W_3 + W_4)$	0.119000	1.000
9	1	$0.282882\,(W_2 + W_3)$	0.158812	0.983
9	2	$0.790855\,W_3$	0.233068	0.985
10	0	$0.108696\,(W_1 + W_2 + W_3 + W_4)$	0.106392	0.998
10	1	$0.238741\,(W_2 + W_3 + W_5)$	0.137784	0.980
10	2	$0.681084\,W_3$	0.190810	0.973
10	3	$1.267536\,W_4$	0.305295	0.997

where each c_i takes the values 0 or 1 and D is the constant that would make the estimator unbiased; as a result D is given by

$$D = \left\{2 \sum_{i=r+1}^{n/2} c_i \mu_{n-i+1}\right\}^{-1}. \tag{24.48}$$

In Table 24.7 the estimator $\tilde{\phi}$ is presented for $n = 4(1)10$ for various choices of r. The efficiency of this estimator relative to the BLUE based on the symmetrically Type-11 censored sample (see Table 24.1) is also presented in this table. More elaborate tables have been provided by Raghunandanan and Srinivasan (1971). Similar simplified linear estimators for the normal case have been discussed in Chapter 13.

Iliescu and Vodă (1973) have presented minimum mean-square-error estimator of the form

$$b_n \sum_{i=1}^{[n/2]} W_i$$

for the parameter ϕ. These authors have also presented the appropriate values of b_n for $n = 2(1)100$. It may be noted that this estimator is essentially of the same form as Raghunandanan and Srinivasan's simplified linear estimator (24.47) (with all c_i's taken to be 1).

4.4 Asymptotic Best Linear Unbiased Estimation

Chan and Chan (1969) discussed the estimation of the parameters 8 and ϕ based on selected order statistics. Ahsanullah and Rahim (1973) discussed the simplified estimation of 8 and ϕ based on optimally selected order statistics from a middle-censored sample.

By using the theory of the asymptotic best linear unbiased estimation *(ABLUEs)* of parameters developed by Ogawa *(1951)*, Cheng (1978) established the following:

1. The optimal spacing $\{\lambda_i\}$ for the ABLUE θ^* is just a single-point spacing $\{\frac{1}{2}\}$, which is independent of the value of the number (k) of selected order statistics, and hence it is unique with $\mathrm{ARE}(\theta^*) = 1$;
2. The optimal spacing $\{\lambda_i\}$ for the ABLUE ϕ^* is not symmetric about the point $\frac{1}{2}$ when k is odd. However, when k is even, the optimal spacing may be symmetric about the point $\frac{1}{2}$.

Cheng (1978) presented the optimal spacing $\{\lambda_i\}$ and the corresponding coefficients $\{b_i\}$ for the ABLUE +*, and the asymptotic efficiency of this estimator relative to the Rao-Cramer lower bound. In Table 24.8, these values are presented for k = 1(1)10 [taken from Cheng (1978)].

Ali, Umbach, and Hassanein (1981) discussed the ABLUE of quantiles x_ξ of the Laplace distribution in (24.11) [i.e., $F_X(x_\xi) = \xi$] based on two optimally chosen order statistics. They have provided the explicit form of this estimator of x_ξ as

$$x_\xi^{**} = 0.255X'_{[0.30506\xi n]+1} + 0.745X'_{[1.50134\xi n]+1}$$

for $0.0352 \le \xi \le 0.3330$

$$= -\frac{z_\xi}{1.59362}X'_{[0.10159n]+1} + \left(1 + \frac{z_\xi}{1.59362}\right)X'_{[n/2]+1}$$

for $\xi < 0.0352$ and $0.3330 < \xi < 0.5$

$$= X'_{[n/2]+1} \qquad \text{for } \xi = 0.5$$

$$= \left(1 - \frac{z_\xi}{1.59362}\right)X'_{[n/2]+1} + \frac{z_\xi}{1.59362}X'_{[0.89841n]+1}$$

for $0.5 < \xi < 0.6670$ and $\xi > 0.9648$

$$= 0.745X'_{[(1.50134\xi - 0.50134)n]+1} + 0.255X'_{[(0.30536\xi + 0.69494)n]+1}$$

for $0.6670 \le 5 \le 0.9648$, (24.49)

Table 24.8 The optimal spacing $\{\lambda_i\}$, the coefficients $\{b_i\}$, and the ARE(ϕ^{**}) of the ABLUE $+^{**}$

k	1	2	3	4	5
λ_1	0.079297	0.101594	0.033422	0.036723	0.016452
λ_2		0.898406	0.164490	0.180735	0.080968
λ_3			0.890440	0.819265	0.223999
λ_4				0.963277	0.810958
λ_5					0.961589
b_1	−0.543063	−0.313750	−0.091528	−0.089540	−0.038492
b_2		0.313750	−0.267216	−0.261596	−0.112454
b_3			0.299908	0.261596	−0.221847
b_4				0.089540	0.262269
b_5					0.089868
ARE(ϕ^{**})	0.292036	0.647609	0.730316	0.820263	0.854828

k	6	7	8	9	10
λ_1	0.017277	0.008980	0.009478	0.005620	0.005752
λ_2	0.085029	0.044197	0.046645	0.027661	0.028311
λ_3	0.235233	0.122269	0.129043	0.076523	0.078322
λ_4	0.764767	0.259890	0.274288	0.162653	0.166477
λ_5	0.914971	0.763397	0.725712	0.296500	0.303472
λ_6	0.982723	0.914475	0.870957	0.720691	0.696528
λ_7		0.982622	0.953355	0.868595	0.833523
λ_8			0.990522	0.952501	0.921678
λ_9				0.990349	0.971689
λ_{10}					0.994248
b_1	−0.038779	−0.019766	−0.020451	−0.011994	−0.012141
b_2	−0.113294	−0.057748	−0.059746	−0.035044	−0.035472
b_3	−0.223845	−0.114094	−0.118045	−0.069236	−0.070084
b_4	0.223845	−0.192711	−0.195351	−0.114579	−0.115982
b_5	0.113294	0.224331	0.195351	−0.171350	−0.173167
b_6	0.038779	0.111746	0.118045	0.197019	0.173167
b_7		0.038250	0.059746	0.118894	0.115982
b_8			0.020451	0.060176	0.070084
b_9				0.020597	0.035472
b_{10}					0.012141
ARE(ϕ^{**})	0.891047	0.908654	0.926909	0.937134	0.947572

where z_ξ is the quantile of the standard Laplace distribution (24.2). They have reported some selected asymptotic efficiency values of this estimator relative to $X'_{[n\xi]+1}$. For example, for $\xi = 0.1$, 0.2, 0.4, and 0.5, the asymptotic relative efficiencies are 122%, 128%, 147%, and 100%, respectively. Further discussion on this problem has been provided by Saleh, Ali, and Umbach (1983). Umbach, Ali, and Saleh (1984) have discussed hypothesis testing using ABLUEs based on optimal spacings.

4.5 Conditional Inference

Kappenman (1975) discussed conditional confidence intervals for the parameters θ and ϕ; see also the note by Edwards (1974). With $\hat{\theta}$ = sample median and $\hat{\phi}$ as in (24.28) being the MLEs of θ and ϕ, and with

$$a_i = \frac{X_i' - \hat{\theta}}{\hat{\phi}}, \qquad i = 1, 2, \dots, n, \tag{24.50}$$

being the ancillary statistics (only n − 2 of which are independent), the joint conditional density function, given the ancillary statistics, is

$$p(\hat{\theta}, \hat{\phi} | \boldsymbol{a}) = K \cdot \frac{1}{\phi^2}\left(\frac{\hat{\phi}}{\phi}\right)^{n-2} \exp\left\{-\frac{\hat{\phi}}{\phi}\sum_{i=1}^{n}\left|\frac{\hat{\theta}-\theta}{\hat{\phi}} + a_i\right|\right\}. \tag{24.51}$$

Defining $U = (\hat{\theta} - \theta)/\hat{\phi}$ and $V = \hat{\phi}/\phi$ to be the pivotal quantities for θ and ϕ, respectively, the conditional joint density function of U and V, given the ancillary statistics, a, is obtained from (24.51) to be

$$p_{U,V}(u, v | \boldsymbol{a}) = K' v^{n-1} e^{-v\sum_{i=1}^{n}|u+a_i|}, \tag{24.52}$$

where the normalizing constant K' is

$$K' = \frac{1}{2\Gamma(n-1)}\left[B_n(a_1, a_2, \dots, a_n)c(\hat{\theta})\right]^{n-1}. \tag{24.53}$$

Here

$$c(t) = \sum_{i=1}^{n} |a_i - t| \tag{24.54}$$

and

$$B_n(a_1, a_2, \dots, a_n)$$

$$= \left\{\sum_{i=1}^{n}\left(\frac{c(\hat{\theta})}{c(a_i)}\right)^{n-1}\frac{1}{(2i-n)(n+2-2i)}\right\}^{-1/(n-1)} \quad \text{if } n \text{ is odd}$$

$$= \left\{(n-1)(a_{(n/2)+1} - a_{n/2})\frac{1}{2c(\hat{\theta})} + \frac{1}{2}\right.$$

$$\left. + \sum_{\substack{i=1 \\ i \neq n/2, n/2+1}}^{n}\left(\frac{c(\hat{\theta})}{c(a_i)}\right)^{n-1}\frac{1}{(2i-n)(n+2-2i)}\right\}^{-1/(n-1)}$$

if n is even.

(24.55)

Uthoff (1973) essentially calculated the constant K' in (24.53) while developing the most powerful location and scale invariant test of the normal distribution against the double exponential distribution.

From (24.52) the marginal conditional density function of U is obtained as

$$p_U(u|\boldsymbol{a}) = K'\Gamma(n)\left\{\sum_{i=1}^{n}|u + a_i|\right\}^{-n} \tag{24.56}$$

from which the conditional 100(1 − a)% confidence interval for the parameter θ can be produced as $(\hat{\theta} - u_2\hat{\phi}, \hat{\theta} - u_1\hat{\phi})$ by finding two constants u_1 and u_2 such that

$$\Pr[U \le u_1|\boldsymbol{a}] = \Pr[U \ge u_2|\boldsymbol{a}] = \alpha/2.$$

Further the marginal conditional density function of V can be similarly obtained from (24.52), from which we get

$$\begin{aligned}\Pr[v_1 < V < v_2|\boldsymbol{a}] = K'\Bigg\{ & \frac{\Gamma(n-1; v_2c(a_1)) - \Gamma(n-1; v_1c(a_1))}{n(c(a_1))^{n-1}} \\ & + \sum_{i=1}^{n-1}\frac{\Gamma(n-1; v_2c(a_i)) - \Gamma(n-1; v_1c(a_i))}{(2i-n)(c(a_i))^{n-1}} \\ & - \sum_{i=1}^{n-1}\frac{\Gamma(n-1; v_2c(a_{i+1})) - \Gamma(n-1; v_1c(a_{i+1}))}{(2i-n)(c(a_{i+1}))^{n-1}} \\ & + \frac{\Gamma(n-1; v_2c(a_n)) - \Gamma(n-1; v_1c(a_n))}{n(c(a_n))^{n-1}}\Bigg\},\end{aligned} \tag{24.57}$$

where $\Gamma(n-1; z) = \int_0^z e^{-t}t^{n-2}\,dt$, $0 < z < \infty$, is the incomplete gamma function. The conditional 100(1 − a)% confidence interval for the parameter

Table 24.9 Comparison of expected lengths of 100(1 − a)% conditional and unconditional confidence intervals for θ (with $\phi = 1$)

$1-\alpha$	0.90		0.95		0.98	
n	Conditional	Unconditional	Conditional	Unconditional	Conditional	Unconditio
3	3.352	3.641	4.740	4.975	7.495	7.649
5	2.113	2.273	2.575	2.912	3.542	3.787
9	1.375	1.498	1.698	1.949	2.119	2.316
15	0.997	1.061	1.214	1.326	1.484	1.525
33	0.631	0.682	0.761	0.830	0.917	0.942

ϕ can be produced as $(\hat{\phi}/v_2, \hat{\phi}/v_1)$ by finding the constants v_1 and v_2 such that $\Pr[v_1 < V < v_2|a]$ in (24.57) equals $1 - \alpha$.

Grice, Bain, and Engelhardt (1978) compared the conditional confidence intervals for θ resulting from (24.56) with the unconditional confidence intervals for θ, based on MLEs, derived from (24.53). Through Monte Carlo simulations they observed that the conditional method gives slightly better results (narrower confidence intervals), although a close agreement develops as n gets larger. For example, the expected lengths of the conditional and unconditional confidence intervals are presented in Table 24.9 for some selected sample sizes and choices of $1 - \alpha$ [taken from Grice, Bain, and Engelhardt (1978)].

4.6 Other Developments

Asrabadi (1985) discussed the minimum variance unbiased estimator for ϕ and exact confidence intervals based on it. Harter, Moore, and Curry (1979) proposed some adaptive robust estimates of location and scale parameters of symmetric populations and examined their performance in the case of the Laplace distribution. Joshi (1984) discussed an expansion of the Bayes risk in the case of the double exponential family. Ramsey (1971) derived the small-sample power functions for some nonparametric tests of location when the sample is assumed to have come from the Laplace distribution; also see Schlittgen (1979).

Awad and Fayoumi (1985) discussed the estimation of $\Pr[Y < X]$ when X and Y are distributed as Laplace. Patel (1986) considered the double exponential case while discussing in general the estimation of finite mixtures of distributions. Yen and Moore (1988) proposed a modified goodness-of-fit test for testing the validity of the Laplace distribution for a given sample. Damsleth and El-Shaarawi (1989) examined ARMA models with double exponentially distributed noise, while Shamma, Amin, and Shamma (1991) discussed a double exponentially weighted moving average control procedure with variable sampling intervals. Ulrich and Chen (1987) considered a bivariate form of the Laplace distribution and its generalizations. Efron (1986) discussed in detail the double exponential families and their use in generalized linear regression. Hwang and Chen (1986) derived improved confidence sets for the coefficients of a linear model with spherically symmetric errors. Some other interesting issues relating to the Laplace distribution can be seen in the works of Brown and Resnick (1977), Hall and Joiner (1983), Loh (1984), Parker (1988), and Davis and Resnick (1988).

5 TOLERANCE LIMITS AND PREDICTION INTERVALS

Based on a complete sample of size n and making use of the MLEs of θ and ϕ described in Section 24.4.1, Bain and Engelhardt (1973) discussed the determination of tolerance limits (approximately). A function $L(X_1',\cdots,X_n')$ is said to be a lower one-sided (β,γ) tolerance limit if

$$\Pr\left[\int_L^{\infty} p_X(x)\,dx > \beta\right] = \gamma. \tag{24.58}$$

Taking $L(X_1',\cdots,X_n') = \hat{\theta} - b\hat{\phi}$, and with $k_\beta = \log\{2(1-\beta)\}$ and $\xi_\beta = \theta + k_\beta\phi$, we have $F_X(\xi_\beta) = 1 - \beta$ and

$$\begin{aligned}\Pr\left[\int_L^{\infty} p_X(x)\,dx \ge \beta\right] &= \Pr\left[\hat{\theta} - b\hat{\phi} < \theta + k_\beta\phi\right]\\ &= \Pr\left[\frac{\hat{\theta}-\theta}{\phi} - b\frac{\hat{\phi}}{\phi} < k_\beta\right]\\ &= \Pr\left[U_n\left(\frac{b}{n}\right) < k_\beta\right]\\ &= \gamma_\beta,\end{aligned} \tag{24.59}$$

where P, $= n(\hat{\theta} - \theta)/\phi$ and $P_2 = n\hat{\phi}/\phi$ are pivotal quantities for 8 and 4, and $U_n(c) =$ P, $- cP_2$. Thus $L(X_1',\cdots,X_n') = \hat{\theta} - b\hat{\phi}$ is the desired (β,γ_β) lower tolerance limit, and any desired probability can be obtained by the proper choice of b. For specified β and γ, Bain and Engelhardt (1973) have used the approximation

$$\Pr\left[U_n\left(\frac{b}{n}\right) < k_\beta\right] \simeq \Phi\left(\frac{\sqrt{n}\,(k_\beta - b)}{\sqrt{1+b^2}}\right) = \gamma \tag{24.60}$$

to get an approximate expression for the tolerance factor b as

$$b \simeq \frac{1}{n - z_\gamma^2}\left\{-nk_\beta + z_\gamma\left[n\left(1 + k_\beta^2\right) - z_\gamma^2\right]^{1/2}\right\}, \tag{24.61}$$

where z_γ denotes the standard normal yth quantile.

Due to the symmetry of the Laplace distribution, $U(X_1',\cdot\cdot\cdot,X_n') = \hat{\theta} + b\hat{\phi}$ is an upper (β,γ) tolerance limit. Hence an approximate upper (β,γ) tolerance limit can be determined by using the approximate expression of the tolerance factor b in Eq. (24.61).

Kappenman (1977) followed the conditional method elaborated in Section 24.4.5 to derive conditional tolerance intervals. In this approach the interval $(\hat{\theta} - b\hat{\phi}, \infty)$ becomes a lower y probability (conditional) tolerance interval for proportion β (≥ 0.5), where

$$b = -a_h - \frac{c(a_h)}{n-2h} + \frac{1}{n-2h}$$
$$\times \left\{ e^{k_\beta(n-2h)} \left[\frac{p(n-2h)}{K'\Gamma(n-1)} + \left(c(a_h)\right)^{-n+1} \right] \right\}^{-1/(n-1)}, \quad (24.62)$$

where $k_\beta = \log(2(1-\beta))$, a_i's are the ancillary statistics in (24.50), $c(t)$ is defined in (24.54), K' is the normalizing constant defined in (24.531, h is the largest positive integer (≥ 2) such that

$$K'\Gamma(n-1) \left\{ \frac{1}{n(c(a_1))^{n-1}} + \sum_{i=1}^{h-1} \frac{1}{n-2i} \right.$$
$$\left. \times \left[\frac{1}{(c(a_{i+1}))^{n-1}} - \frac{1}{(c(a_i))^{n-1}} \right] \right\} \leq 1 - \gamma, \quad (24.63)$$

and p is the difference between $1 - \gamma$ and the value of the left-hand side of (24.63) so determined. One needs to find h from (24.63) by successively setting h = 2, 3, ... or by trial and error.

Due to the symmetry of the Laplace distribution, an upper γ probability (conditional) tolerance interval for proportion β is obtained by simply replacing b by $-b$ in (24.621, replacing $1 - \gamma$ by γ in (24.63), and taking p as the difference between y and the value of the left-hand side of (24.63) so determined. Then the required upper tolerance interval will be $(-\infty, \hat{\theta} - b\hat{\phi})$.

Shyu and Owen (1986a, b, 1987) provide further discussions on one-sided as well as two-sided tolerance intervals, and they also present some valuable tables of necessary tolerance factors. By noting that all the above mentioned works are based on complete samples, Balakrishnan and Chandramouleeswaran (1994a) used the BLUEs of θ and ϕ (elaborated in Section 4.2) to develop lower and upper tolerance limits based on Type-II censored samples. They considered the lower (β, γ) tolerance limit to be of the form $L(X'_1, \cdots, X'_n) = \theta^* - b\phi^*$, and presented extensive tables of the tolerance factor b for n = 5(1)10, 12, 15, 20, level of right-censoring $s = 0(1)[n/2]$, $\beta = 0.500(0.025)0.975$, and $\gamma = 0.75$, 0.85, 0.90, 0.95, 0.98, 0.99 and 0.995. Once again, due to the symmetry of the Laplace distribution, these tables also enable the determination of the upper (β, γ) tolerance limit, $U(X'_1, \cdots, X'_n) = \theta^* + b\phi^*$.

Balakrishnan and Chandramouleeswaran (1994a) have also used the BLUEs, θ^* and ϕ^*, to propose a natural estimator for the reliability of X at time t as

$$R_X^*(t) = 1 - F_X(t;\theta^*,\phi^*) = 1 - \tfrac{1}{2}e^{(t-\theta^*)/\phi^*}, \qquad t \leq \phi^*$$
$$= \tfrac{1}{2}e^{-(t-\theta^*)/\phi^*}, \qquad t \geq \phi^*. \qquad (24.64)$$

They have examined the bias and variance of this estimator for complete as well as Type-I1 right-censored samples at various choices of t. They observed that the estimator in (24.64) is almost unbiased at all the levels of reliability examined even for sample sizes as small as 5. These authors have also explained how the tables of tolerance factors b could be used successfully to determine the lower $100\gamma\%$ confidence limit for the reliability $R_X(t)$.

Balakrishnan and Chandramouleeswaran (1994b) used the BLUEs of 8 and ϕ, based on Type-I1 right-censored samples, to develop prediction intervals. Specifically, by taking X'_1 I X'_2 I . · · I X'_{n-s} to be the available Type-I1 censored sample where the largest s observations have been censored, they have discussed the prediction of X'_{n-s+1} and X'_n through the pivotal quantities

$$Q_1 = \frac{X'_{n-s+1} - X'_{n-s}}{\phi^*} \quad \text{and} \quad Q_2 = \frac{X'_n - X'_{n-s}}{\phi^*}, \qquad (24.65)$$

respectively. Balakrishnan and Chandramouleeswaran (1994b) have presented necessary tables of percentage points of Q, and Q_2 (determined through Monte Carlo simulations) for various choices of n and s. They have in addition discussed the prediction of observations from a future sample of size m (particularly for Y'_1 and Y'_m) based on the pivotal quantities

$$Q_3 = \frac{Y'_1 - o^*}{\phi^*} \quad \text{and} \quad Q_4 = \frac{Y'_m - \theta^*}{\phi^*}, \qquad (24.66)$$

and presented some necessary tables of percentage points of Q_3 and Q_4. Ling (1977) and Ling and Lim (1978) have discussed a Bayesian approach to these prediction problems.

6 RELATED DISTRIBUTIONS

If X has probability density function (24.1), then $|X - \theta|$ is distributed exponentially, in fact, as $\frac{1}{2}\phi \times (\chi^2$ with two degrees of freedom). In particular, if $0 = 0$, then $|X|$ is so distributed. For this reason, if $X_1, X_2, \cdots, X_n$ are independent random variables, each having probability density function (24.1) with $\theta = 0$, then the distribution of any statistic depending only on the absolute values $|X_1|, |X_2|, \cdots, |X_n|$ can be derived from an initial joint distribution of independent multiples of χ^2 variables. For example, $|X_1|/|X_2|$ is distributed as F with 2,2 degrees of freedom (see Chapter 27).

An interesting connection between the normal and Laplace distributions has been established by Nyquist, Rice, and Riordan (1954). They showed that if U_1, U_2, U_3, and U_4 are independent unit normal variables, then the probability density function of

$$D = \begin{vmatrix} U_1 & U_2 \\ U_3 & U_4 \end{vmatrix} = U_1U_4 - U_2U_3$$

is of form (24.1) with $\theta = 0, \phi = 2$. It may be noted here that $U_1U_4 - U_2U_3$ and $U_1U_4 + U_2U_3$ have the same distribution. [The case when the expected values of the U's are not equal to zero leads to a more complicated distribution, and it was considered by Nicholson (1958).] Missiakoulis and Darton (1985) and Mantel (1987) have made some additional remarks on this result.

Mantel and Pasternack (1966) gave a heuristic demonstration that $Y = U_1U_4 + U_2U_3$ follows the Laplace distribution; see also Mantel (1969). A simple proof of this result through the use of characteristic functions has been provided by Mantel (1970). First, the characteristic function of Y is given by

$$E[e^{itY}] = E[e^{it(U_1U_4 + U_2U_3)}] = \{E[e^{itU_1U_4}]\}^2, \tag{24.67}$$

since U_1U_4 and U_2U_3 are independently and identically distributed. Now $E[e^{itU_1U_4}]$ can be evaluated in two stages. The conditional expectation

$$E\left[e^{itU_1U_4}|U_4\right] = e^{-U_4^2t^2/2}, \tag{24.68}$$

since U_1 is $N(0,1)$ (Chapter 13). Next, the unconditional expectation

$$E[e^{itU_1U_4}] = E\left[e^{-U_4^2t^2/2}\right] = \frac{1}{\sqrt{1+t^2}} \tag{24.69}$$

since U_4^2 is χ_1^2 (Chapter 18). Using (24.69) in (24.67), we get

$$E[e^{itY}] = \frac{1}{1+t^2}$$

which agrees with the characteristic function in (24.3), thus proving that Y is distributed as Laplace.

The two reciprocal Fourier integrals

$$\tfrac{1}{2}\int_{-\infty}^{\infty} \exp(itx - |x|)\, dx = (1 + t^2)^{-1}$$

and

$$\pi^{-1}\int_{-\infty}^{\infty} (1 + t^2)^{-1} \exp(itx)\, dx = e^{-|x|}$$

represent a formal connection between the Cauchy and Laplace distributions [see also (24.2), (24.3) and Chapter 161.

Transformed forms of the Laplace distribution have been discussed by Johnson (1954). He considered (by analogy with the lognormal, S_U and S_B systems; see Chapters 14 and 12) distributions of a random variable Y when (with 6 > 0)

$$X = \begin{cases} \gamma + 6 \log Y & (S'_L \text{ system}), \\ \gamma + \delta \sinh^{-1} Y & (S'_U \text{ system}), \\ \gamma + \delta \log\left(\dfrac{Y}{1 - Y}\right) & (S'_B \text{ system}), \end{cases}$$

and X has the standard Laplace distribution (24.2).

The (β_1, β_2) points of the S'_L system lie on the line with parametric equations

$$\beta_1(Y) = \frac{4(\delta^2 - 4)(15\delta^4 + 7\delta^2 + 2)^2}{\delta^2(\delta^2 - 9)^2(2\delta^2 + 1)^3}, \qquad \delta > 3, \tag{24.70}$$

$$\beta_2(Y) = \frac{3(\delta^2 - 4)(8\delta^8 + 212\delta^6 + 95\delta^4 + 33\delta^2 + 12)}{\delta^2(\delta^2 - 9)(\delta^2 - 16)(2\delta^2 + 1)^2}, \qquad \delta > 4. \tag{24.71}$$

The (β_1, β_2) points of S'_U lie "below" this line (i.e., larger values of β_2 for given β_1); those of S'_B lie above it. All possible values of (β_1, β_2) are covered by these three systems combined. For both S'_L and S'_U the rth moment is infinite if $r \geq 6$.

The S'_L system of distributions is referred as the log-Laplace distributions (analogous to the lognormal and log-logistic distributions). Kotz, Johnson, and Read (1985) provide a brief review of this system of distributions, and Uppuluri (1981) discuss some of its properties.

Asymmetrical Laplace distributions, with probability density functions of form

$$p_X(x) = \begin{cases} (2\phi_1)^{-1}\exp\left[-\dfrac{|x-\theta|}{\phi_1}\right], & x \ge \theta, \\ (2\phi_2)^{-1}\exp\left[-\dfrac{|x-\theta|}{\phi_2}\right], & x < \theta, \end{cases} \tag{24.72}$$

where $\phi_1 \neq \phi_2$ and $\phi_1, \phi_2 > 0$, are sometimes used [see McGill (1962)]. Lingappaiah (1988), terming this as two-piece double exponential distribution, has discussed some properties.

Another form of asymmetrical Laplace distribution has probability density function

$$p_X(x) = \begin{cases} p\phi^{-1}\exp\left[-\dfrac{|x-\theta|}{\phi}\right], & x \ge \theta, \\ (1-p)\phi^{-1}\exp\left[-\dfrac{|x-\theta|}{\phi}\right], & x < \theta, \end{cases} \tag{24.73}$$

with $0 < p < 1$. Holla and Bhattacharya (1968) have used this distribution as the compounding distribution of the expected value of a normal distribution. The characteristic function of the resulting compound normal distribution is

$$(1+t^2\phi^2)^{-1}\{1+(2p-1)it\phi\}\exp\left[it\theta - \tfrac{1}{2}t^2\sigma^2\right] \tag{24.74}$$

where σ^2 is the variance of the compounded normal distribution. The probability density function (with argument y) is

$$(\phi\sqrt{2\pi})^{-1}e^{\sigma^2/(2\phi^2)}\left[pe^{-(y-\theta)/\phi}\left\{\sqrt{\frac{\pi}{2}} - S_1 M\left(\frac{1}{2};\frac{3}{2};-\frac{1}{2}S_1^2\right)\right\} + (1-p)e^{(y-\theta)/\phi}\left\{\sqrt{\frac{\pi}{2}} - S_2 M\left(\frac{1}{2};\frac{3}{2}-\frac{1}{2}S_2^2\right)\right\}\right], \tag{24.75}$$

where

$$S_j = (\sigma/\phi) + (-1)^j\left\{\frac{x-\theta}{\sigma}\right\}, \qquad j = 1,2,$$

and $M(\cdot)$ is the confluent hypergeometric function (Chapter 1).

For the particular case of p = $\frac{1}{2}$ we have the distribution

$$\text{Normal } (\xi, \text{a}) \bigwedge_{\xi} \text{Laplace } (\theta, \phi).$$

This distribution is symmetrical, with mean 8. The variance is (a$^2 + 2\phi^2$) and the moment ratio a, ($\equiv \beta_2$) is

$$3 + 12\phi^4(\sigma^2 + 2\phi^2)^{-2}.$$

Holla and Bhattacharya obtained an expression for the distribution of the sum of n independent random variables each having this distribution. They also obtained the following formula for the cumulative distribution function (argument y):

$$\Phi\left(\frac{y-\theta}{\sigma}\right) - \frac{1}{2}e^{\sigma^2/(2\phi^2)}\sinh\left(\frac{y-\theta}{\sigma}\right) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{\sigma^2/(2\phi^2)}$$
$$\times\left\{e^{(y-\theta)/\phi}S_2M\left(\frac{1}{2};\frac{3}{2};-\frac{1}{2}S_2^2\right) - e^{-(y-\theta)/\phi}S_1M\left(\frac{1}{2};\frac{3}{2};-\frac{1}{2}S_1^2\right)\right\}. \tag{24.76}$$

Among compound *Laplace* distributions, we note, first,

$$\text{Laplace } (\theta, \phi) \bigwedge_{0} \text{Normal } (\xi, \text{a}).$$

The probability density function is

$$p_X(x) = \frac{1}{2\phi}\left\{\exp\left(\frac{1}{2}\left(\frac{\sigma}{\phi}\right)^2\right)\right\}\left[\Phi\left(\frac{x-\xi}{\sigma} - \frac{\sigma}{\phi}\right)\exp\left(-\frac{x-\xi}{\phi}\right)\right.$$
$$\left. + \Phi\left(-\frac{x-\xi}{\sigma} - \frac{\sigma}{\phi}\right)\exp\left(\frac{x-\xi}{\phi}\right)\right] \tag{24.77}$$

where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x}\exp(-\frac{1}{2}t^2)\,dt$. Second,

$$\text{Laplace } (\theta, \phi) \bigwedge_{\phi^{-1}} \text{Gamma } (\alpha, \beta)$$

(The gamma distribution is as given in Chapter 17.)

The probability density function is

$$p_X(x) = \tfrac{1}{2}\alpha\beta[1 + |x - \theta|\beta]^{-(\alpha+1)}. \tag{24.78}$$

The relation between distributions of form (24.78) and **Laplace** distributions is rather similar to that between **Pearson** Type VII and normal distributions (Chapter 28). We note, for example, that as β tends to zero and α to infinity, with $\alpha\beta = 1$, $p_X(x) \to \frac{1}{2}\exp(-|x-\theta|)$.

Distribution (24.78) is symmetrical about θ. Moments of order α or greater do not exist. For r even and less than α

$$\mu_r = \alpha\beta^r \sum_{j=0}^{r} (-1)^j \binom{r}{j} (\alpha + j - r)^{-1}. \tag{24.79}$$

In particular, the variance is

$$\sigma^2 = \frac{2\beta^2}{(\alpha-1)(\alpha-2)}, \qquad \alpha > 2, \tag{24.80}$$

and also

$$\beta_2 = \frac{6(\alpha-1)(\alpha-2)}{(\alpha-3)(\alpha-4)}, \qquad \alpha > 4. \tag{24.81}$$

The mean deviation is

$$\nu_1' = \frac{\beta}{\alpha - 1}. \tag{24.82}$$

Thus

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \sqrt{\frac{2(\alpha-2)}{\alpha-1}}.$$

The cumulative distribution function is of very simple form:

$$F_X(x) = \begin{cases} \frac{1}{2}\{1 + |x-\theta|\beta\}^{-\alpha}, & x \le \theta, \\ 1 - \frac{1}{2}\{1 + |x-\theta|\beta\}^{-\alpha}, & x \ge \theta. \end{cases}$$

Subbotin (1923), on the basis of certain broad requirements for "error distributions," obtained the class

$$p_X(x) = \left[2^{(\delta/2)+1}\Gamma\left(\frac{1}{2}\delta + 1\right)\right]^{-1} \phi^{-1}\exp\left[-\frac{1}{2}\left|\frac{x-\theta}{\phi}\right|^{2/\delta}\right],$$

$$\delta, \phi > 0. \tag{24.83}$$

[See also **Fréchet (1924)**, where **Subbotin's** arguments are criticized.] This class of distributions includes **Laplace** ($\delta = 2$), normal ($\delta = 1$), and, as a limiting ($\delta \to 0$) case, rectangular distributions. It is symmetrical about θ, and has finite moments of all positive orders. The rth central moment is

$$\mu_r = \begin{cases} 0, & r \text{ odd}, \\ \dfrac{\phi^r 2^{r\delta/2}\Gamma((r+1)\delta/2)}{\Gamma(\delta/2)}, & r \text{ even}. \end{cases} \tag{24.84}$$

The variance is

$$\sigma^2 = \frac{2^{\delta}\Gamma(3\delta/2)}{\Gamma(\delta/2)}\phi^2 \tag{24.85}$$

and the mean deviation is

$$\nu_1' = \frac{2^{\delta/2}\Gamma(\delta)}{\Gamma(\delta/2)}\phi. \tag{24.86}$$

Thus

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{\Gamma(\delta)}{[\Gamma(\delta/2)\Gamma(3\delta/2)]^{1/2}}. \tag{24.87}$$

Also

$$\beta_2 = \frac{\Gamma(5\delta/2)\Gamma(\delta/2)}{[\Gamma(3\delta/2)]^2}. \tag{24.88}$$

Some values of β_2 and the ratio (24.87) are given in Table 24.10. The cumulative distribution function corresponding to (24.83) can be expressed in terms of incomplete gamma functions. Maximum likelihood estimation of the parameters was discussed by Diananda (1949); see also Turner (1960).

The classes of distributions (24.78) and (24.83) were used by Box and Tiao (1962) as prior distributions for certain Bayesian statistical analyses. Distribution (24.83) provides a convenient set of alternatives to normality, if symmetry can be assumed, and hence has been used often in robustness studies. Tiao and Lund (1970), for example, discussed the use of linear minimum variance unbiased estimators in inference robustness studies of the location parameter of the distribution (24.83). Order statistics from this distribution and their properties have also been discussed by these authors. With $\theta = 0$ and 6 assumed to known, Jakuszenkow (1979) discussed the estimation of variance of the distribution (24.83) which is a multiple of ϕ^2. Sharma (1984) presented an improved estimator in this case. Zeckhauser and Thompson

Table 24.10 Ratio of mean deviation to standard deviation and β_2 for Subbotin distributions

δ	$\frac{\text{Mean Deviation}}{\text{Standard Deviation}}$	β_2
0 (uniform)	0.866	1.800
0.25	0.858	1.923
0.5	0.841	2.188
0.75	0.815	2.548
1 (normal)	0.798	3.000
1.5	0.757	4.222
2 (Laplace)	0.707	6.000
3	0.623	12.257
4	0.548	25.200
5	0.481	51.951

(1970) discussed linear regression with errors having the Subbotin density in (24.83). Specifically these authors investigated the model

$$y_i = a + bx_i + e_i, \qquad i = 1,2,\dots,n,$$

where the manifest observations are the (x_i, y_i) pairs, and the error random variables are i.i.d. with density (24.83) with $\theta = 0$. The parameters of the model are a, b, ϕ, and δ. The likelihood function based on the entire sample is

$$L(a,b,\phi,\delta) = ce^{-S/\phi^{2/\delta}},$$

where

$$S = \sum_{i=1}^{n} |y_i - a - bx_i|^{2/\delta}.$$

It is then clear that ϕ has no effect on the maximum likelihood estimation of parameters of the regression line and that the maximum likelihood estimator of ϕ is

$$\hat{\phi} = \left(\frac{2S}{n\delta}\right)^{\delta/2}.$$

Zeckhauser and Thompson (1970) then examined the maximum likelihood estimation of the parameters a, b, and S. Krysicki (1966) gave formulas for estimating the parameters in a mixture of two Laplace distributions, each having $\theta = 0$. Sródka (1966) discussed the distributions obtained if ϕ^{-1} is supposed to have a generalized gamma distribution (as defined in Chapter 17). Kanji (1985) and Jones and McLachlan (1990) have discussed the ***Laplace-normal mixture distribution*** with density function

$$p_X(x;\theta,\phi,\sigma,p) = \frac{p}{2\phi}e^{-|(x-\theta)/\phi|} + \frac{1-p}{\sqrt{2\pi}\,\sigma}e^{-(x-\theta)^2/2\sigma^2}, \qquad -\infty < x < \infty, \tag{24.89}$$

and applied the distribution to fit wind shear data. Maximum likelihood estimation of parameters of the distribution (24.89) has been discussed recently by Scallan (1992).

The ***reflected gamma distribution*** with density function

$$p_X(x) = \frac{1}{2\phi\Gamma(\alpha)}\left|\frac{x-\theta}{\phi}\right|^{\alpha-1} e^{-|(x-\theta)/\phi|}, \qquad -\infty < x < \infty, \alpha, \phi > 0, \tag{24.90}$$

introduced by Borghi (1965), includes the Laplace distribution as a special case when the shape parameter $\alpha = 1$. Kantam and Narasimham (1991), while studying the best linear unbiased estimator and some other linear estimators of 8, observed that the median (unlike in the Laplace case) becomes an inefficient estimator of 8 when a becomes large. Harvey (1967) introduced a more general form of the reflected gamma distribution in (24.90) which naturally becomes a four-parameter generalization of the Laplace distribution (24.1). Unlike in the case of (24.901, Harvey's generalized density will not, in general, be zero at $x = 8$.

The ***double Weibull distribution*** with density function

$$p_X(x) = \frac{c}{2\phi}\left|\frac{x-\theta}{\phi}\right|^{c-1} e^{-|(x-\theta)/\phi|^c}, \qquad -\infty < x < \infty, c, \phi > 0, \tag{24.91}$$

introduced by Balakrishnan and Kocherlakota (1985), is a symmetric family of distributions which includes the Laplace distribution as a special case when the shape parameter c = 1. Balakrishnan and Kocherlakota (1985) and Dattatreya Rao and Narasimham (1989) derived the BLUEs of θ and ϕ, assuming c to be known, based on complete and Type-I1 censored samples, respectively. Assuming 8 to be known, Vasudeva Rao, Dattatreya Rao, and Narasimham (1991) discussed optimal linear estimators of θ based on the values of $|X_i' - \theta|$. [Observe that the MLE of ϕ in the case of Laplace distribution in (24.29) is a linear form in $|X_i' - \theta|$.] An interesting relationship between the logistic and the Laplace distributions was brought out by George and Rousseau (1987) while discussing the distribution of the midrange in a sample from a logistic distribution (see Chapter 23).

Along the lines of skew-normal distributions (see Chapters *12* and 13) introduced by Azzalini (1985), Balakrishnan and Arnbagaspitiya (1994) considered the ***skew-Laplace distribution*** with density function

$$2p_X(x)F_X(\lambda x), \tag{24.92}$$

where $p_X(x)$ is the two-parameter Laplace density in *(24.1)* and $F_X(\cdot)$ is the corresponding distribution function *(24.11)*. Balakrishnan and Ambagaspitiya *(1994)* discussed various properties of this distribution, which becomes the Laplace distribution when $A = 0$ and the two-parameter exponential distribution with $A \to \infty$. These authors also studied order statistics from this distribution and derived BLUEs of the parameters θ and 4, assuming that the shape parameter A is known.

The distribution with characteristic function

$$\phi_X(t) = (1 + |t|^{\alpha})^{-1}, \qquad -\infty < t < \infty, 0 < \alpha \le 2 \tag{24.93}$$

is called an ***a-Laplace distribution*** since a = *2* corresponds to the Laplace characteristic function *(24.3)*. This has been shown to be a unimodal distribution by Linnik and Laha; see, for example, Lukacs (1970). Pillai *(1985)* has introduced a larger class of distributions, termed ***semi-a-Laplace distributions,*** of which a-Laplace distribution is a special case. Let $\phi(t)$ be a characteristic function, which is never zero, defined by

$$\phi(t) = (1 + f(t))^{-1}. \tag{24.94}$$

Then, from the properties of $\phi(t)$, it is clear that $f(0) = 0$ and $f(t)$ is continuous. A distribution function f is called ***semi-a-Laplace*** if $f(t)$ in (24.94) has the property

$$f(t) = af(bt) \qquad \text{for} 0 < b < 1, \tag{24.95}$$

where a is the unique solution of the equation

$$ab'' = 1, \qquad 0 < \alpha\ I\ 2. \tag{24.96}$$

The numbers b and a are termed ***order*** and ***exponent*** of the semi-a-Laplace distribution, respectively. Pillai (1985) has also proved the following characterization of this distribution:

"$F(x)$ is semi-a-Laplace distribution of some variable X with order b iff $F(x)$ satisfies the equation

$$F(x) = pF_1(x) + (1 - p)F_2(x) \tag{24.97}$$

for some $p \in (0, 1)$ where $F_1(x)$ is the distribution of bX and $F_2(x) = F * F_{,}$."

More properties of this distribution have been discussed by Pillai (1985) and Divanji (1988).

The distribution corresponding to the characteristic function (24.93) has also been referred to as ***Linnik's distribution*** [see Devroye (1990)]. By making use of the fact that for $\alpha \le 1, \phi_X(t)$ in (24.93) is a Pólya characteristic function (convex on the positive halfline), Devroye (1986) presented a simple algorithm for generating pseudorandom observations from this distribution. Subsequently Devroye (1990) presented an algorithm for all values of α which is based on the following observation: Suppose that S_α is a symmetric stable random variable with characteristic function $e^{-|t|^\alpha}$ and that V_β is an independent random variable with density

$$\frac{e^{-v^\beta}}{\Gamma\left(1+\dfrac{1}{\beta}\right)}, \qquad v > 0.$$

Then $X = S_\alpha V_\beta^{\beta/\alpha}$ has characteristic function

$$\begin{aligned}\phi_X(t) &= E[e^{itX}] = E\left[e^{-|t|^\alpha V_\beta^\beta}\right]\\ &= \frac{1}{\Gamma(1+1/\beta)}\int_0^\infty e^{-v^\beta - |t|^\alpha v^\beta}\,dv\\ &= (1+|t|^\alpha)^{-1/\beta}.\end{aligned} \tag{24.98}$$

For the case $\beta = 1$ the characteristic function (24.98) is that of the Linnik distribution in (24.93). This result, in addition to showing that the characteristic function in (24.93) is the same as that of $S_\alpha V_1^{1/\alpha}$, where V_1 is exponentially distributed, provides a short proof of the validity of the Linnik's characteristic function (24.93). Lin (1994) has recently investigated some basic properties such as self-decomposability and also established two characterizations.

Kotz and Ostrovskii (1994) have recently given a mixture representation of the Linnik distribution. Specifically, with X_α and X_β denoting two random variables possessing the Linnik distribution (24.93) with parameters α and β, respectively $(0 < \alpha < \beta \le 2)$ and $Y_{\alpha\beta}$ a nonnegative random variable (independent of X_β) with density function

$$g(s;\alpha,\beta) = \left(\frac{\beta}{\pi}\sin\frac{\pi\alpha}{\beta}\right)\cdot\frac{s^{\alpha-1}}{1+s^{2\alpha}+2s^\alpha\cos(\pi\alpha/\beta)}, \qquad 0 < s < \infty. \tag{24.99}$$

Kotz and Ostrovskii (1994) have shown that

$$X_\alpha \overset{d}{=} X_\beta Y_{\alpha\beta}. \tag{24.100}$$

From this representation the infinite divisibility of mixtures of Linnik distributions with respect to the parameter α and a scale parameter follows easily.

Kotz, Ostrovskii, and Hayfavi (1994) have presented convergent asymptotic series expansions for the Linnik density function. The analytic structure of the density depends substantially on the arithmetic nature of the parameter α. For example, when $\alpha = 1$, the density is given by

$$p_1(x) = \frac{1}{\pi}(\cos x)\log\frac{1}{|x|} + \frac{1}{2}\sin|x| + \frac{1}{\pi}\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma'(2k+1)}{\Gamma^2(2k+1)}x^{2k}. \qquad (24.101)$$

7 APPLICATIONS

As mentioned already in Section 2, the Laplace distribution (being heavier tailed than the normal) has been used quite commonly as an alternative to the normal distribution in robustness studies; for example, see Andrews et al. (1972) and Hoaglin, Mosteller, and Tukey (1985).

In addition the Laplace distribution has found some interesting applications on its own. Manly (1976) gave some examples of fitness functions based on the double exponential distribution. Easterling (1978) considered a model for steam generator inspection as exponential responses with double exponential measurement error. Hsu (1979), while discussing the use of long-tailed distributions for position errors in navigation, suggested the Laplace distribution. Okubo and Narita (1980) used the double exponential for the distribution of extreme winds expected in Japan. As mentioned in the last section, the Laplace-normal mixture distribution in (24.89) has been used to fit some wind shear data by Kanji (1985) and Jones and McLachlan (1990). Bagchi, Hayya, and Ord (1983) used the Laplace distribution while modeling demand during lead time for slow-moving items. Dadi and Marks (1987) discussed detector relative efficiencies in the presence of Laplace noise. A few more applications of the distribution were indicated in Sections 4 and 6.

REFERENCES

Ahsanullah, M., and Rahim, M. A. (1973). Simplified estimates of the parameters of the double exponential distribution based on optimum order statistics from a middle-censored sample, *Naval Research Logistics Quarterly*, **20**, 745–751.

Akahira, M. (1987). Second order asymptotic comparison of estimators of a common parameter in the double exponential case, *Annals of the Institute of Statistical Mathematics*, *39*, 25–36.

Akahira, M. (1990). Second order asymptotic comparison of the discretized likelihood estimator with asymptotically efficient estimators in the double exponential case, *Metron,* 48, 5–17.

Akahira, M., and Takeuchi, K. (1990). Loss of information associated with the order statistics and related estimators in the double exponential distribution case, *Australian Journal of Statistics,* 32, 281–291.

Akahira, M., and Takeuchi, K. (1993). Second order asymptotic bound for the variance of estimators for the double exponential distribution, In *Statistical Science & Data Analysis* (ed., K. Matusita et al.), Amsterdam, Netherlands: VSP Publishers, pp. 375–382.

Ali, M. Masoom, Umbach, D., and Hassanein, K. M. (1981). Estimation of quantiles of exponential and double exponential distributions based on two order statistics, *Communications in Statistics—Theory and Methods,* 10, 1921–1932.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location,* Princeton, NJ: Princeton University Press.

Asrabadi, B. R. (1985). The exact confidence interval for the scale parameter and the MVUE of the Laplace distribution, *Communications in Statistics—Theory and Methods,* 14, 713–733.

Awad, A. M., and Fayoumi, M. (1985). Estimation of $P(Y < X)$ in case of the double exponential distribution, *Proceedings of 7th Conference on Probability Theory, Brasov, Romania,* 527–531.

Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics,* 12, 171–178.

Bagchi, U., Hayya, J. C., and Ord, J. K. (1983). The Hermite distribution as a model of demand during lead time for slow-moving items, *Decision Sciences,* 14,447–466.

Bain, L. J., and Engelhardt, M. (1973). Interval estimation for the two-parameter double exponential distribution, *Technometrics,* **15**, 875–887.

Balakrishnan, N. (1988). Recurrence relations among moments of order statistics from two related outlier models, *Biometrical Journal,* 30, 741–746.

Balakrishnan, N. (1989). Recurrence relations among moments of order statistics from two related sets of independent and non-identically distributed random variables, *Annals of the Institute of Statistical Mathematics,* 41, 323–329.

Balakrishnan, N., and Ambagaspitiya, R. S. (1988). Relationships among moments of order statistics in samples from two related outlier models and some applications, *Communications in Statistics—Theory and Methods, 17,* 2327–2341.

Balakrishnan, N., and Ambagaspitiya, R. S. (1994). On skew-Laplace distributions, *Report,* McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N., and Chandramouleeswaran, M. P. (1994a). Reliability estimation and tolerance limits for Laplace distribution based on censored samples, *Report,* McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N., and Chandramouleeswaran, M. P. (1994b). Prediction intervals for Laplace distribution based on censored samples, *Report,* McMaster University, Hamilton. Ontario. Canada.

Balakrishnan, N., Chandramouleeswaran, M. P., and Ambagaspitiya, R. S. (1994). BLUE's of location and scale parameters of Laplace distribution based on Type-II censored samples and associated inference, *Report,* McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N., Chandramouleeswaran, M. P., and Govindarajulu, Z. (1994). Inference on parameters of the Laplace distribution based on Type-II censored samples using Edgeworth approximation, *Report,* McMaster University, Hamilton, Ontario, Canada.

Balakrishnan, N., and Cohen, A. C. (1991). *Order Statistics and Inference: Estimation Methods,* San Diego, CA: Academic Press.

Balakrishnan, N., and Cutler, C. D. (1994). Maximum likelihood estimation of the Laplace parameters based on Type-II censored samples, In H. *A. Dacid Festschrift Volume* (eds, D. F. Morrison, H. N. Nagaraja, and P. K. Sen), New York: Springer-Verlag (to appear).

Balakrishnan, N., Govindarajulu, Z., and Balasubramanian, K. (1993). Relationships between moments of two related sets of order statistics and some extensions, *Annals of the Institute of Statistical Mathematics,* 45, 243–247.

Balakrishnan, N., and Kocherlakota, S. (1985). On the double Weibull distribution: Order statistics and estimation, *Sankhyā, Series B,* 47, 161–178.

Balakrishnan, N., and Kocherlakota, S. (1986). Effects of nonnormality on $\bar{X}$ charts: Single assignable cause model, *Sankhyii, Series B,* 48, 439–444.

Balanda, K. P. (1987). Kurtosis comparisons of the Cauchy and double exponential distributions, *Communications in Statistics—Theory and Methods,* 16, 579–592.

Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data,* 3d ed., Chichester, England: Wiley.

Borghi, O. (1965). Sobre una distribución de frecuencias, *Trabajos de Estadistica,* 16, 171–192.

Box, G. E. P., and Tiao, G. C. (1962). A further look at robustness via Bayes's theorem, *Biometrika,* 49, 419–432.

Brown, B. M., and Resnick, S. I. (1977). Extreme values of independent stochastic processes, *Journal of Applied Probability,* 14, 732–739.

Chan, L. K., and Chan, N. N. (1969). Estimates of the parameters of the double exponential distribution based on selected order statistics, *Bulletin of the Institute of Statistical Research and Training,* 3, 21–40.

Cheng, S. W. (1978). Linear quantile estimation of parameters of the double exponential distribution, *Soochow Journal of Mathematics,* 4, 39–50.

Chu, J. T., and Hotelling, H. (1955). The moments of the sample median, *Annals of Mathematical Statistics,* 26, 593–606.

Craig, A. T. (1932). On the distribution of certain statistics, *American Journal of Mathematics,* 54, 353–366.

Dadi, M. I., and Marks, R. J., II (1987). Detector relative efficiencies in the presence of Laplace noise, *IEEE Transactions on Aerospace and Electronic Systems,* 23, 568–582.

Damsleth, E., and El-Shaarawi, A. H. (1989). ARMA models with double-exponentially distributed noise, *Journal of the Royal Statistical Society, Series B,* 51, 61–69.

Dattatreya Rao, A. V., and Narasimham, V. L. (1989). Linear estimation in double Weibull distribution, *Sankhyā, Series* B, 51, 24–64.

David, H. A. (1981). *Order Statistics,* 2d ed., New York: Wiley.

Davis, R., and Resnick, S. 1. (1988). Extremes of moving averages of random variables from the domain of attraction of the double exponential distribution, *Stochastic Processes,* 30, 41–68.

Devroye, L. (1986). *Non-uniform Random Variate Generation,* New York: Springer-Verlag.

Devroye, L. (1990). A note on Linnik's distribution, *Statistics & Probability Letters, 9,* 305–306.

Diananda, P. H. (1949). Note on some properties of maximum likelihood estimates, *Proceedings of the Cambridge Philosophical Society,* 45, 536–544.

Divanji, G. (1988). On semi a-Laplace distributions, *Journal of the Indian Statistical Association,* 26, 31–38.

Dobrogowski, A. (1976). A property of the double exponential distribution, *Matematyka Stosowana, Polskie Towarzystwo Matematycmego, 6,* 49–52. (In Polish.)

Dwinas, S. (1948). A deduction of the Laplace-Gauss law of errors, *Revista Matemática Hispano-Americana,* **8(4),** 12–18. (In Spanish.)

Easterling, R. G. (1978). Exponential responses with double exponential measurement error—A model for steam generator inspection, *Proceedings of the DOE Statistical Symposium, U.S. Department of Energy,* pp. 90–110.

Edwards, A. W. F. (1974). Letter to the Editor, *Technometrics,* 16, 641–642.

Edwards, L. (1948). The use of normal significance limits when the parent population is of Laplace form, *Journal of the Institute of Actuaries Students' Society,* **8,** 87–99.

Efron, B. (1986). Double exponential families and their use in generalized linear regression, *Journal of the American Statistical Association,* 81, 709–721.

Farison, J. B. (1965). On calculating moments for some common probability laws, *IEEE Transactions on Information Theory, 11,* 586–589.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications,* vol. 2, New York: Wiley.

Findeisen, P. (1981). Characterization of the bilateral exponential distribution, *Metrika, 29,* 95–102. (In German.)

Fréchet, M. (1924). Sur la loi des erreurs d'observation, *Matematicheskii Sbornik,* 32, 1–8.

Fréchet, M. (1928). Sur l'hypothèse de I'additivite des erreurs partielles, *Bulletin des Sciences et Mathématiques, Paris,* 63, 203–206.

Gallo, F. (1979). On the Laplace first law: Sample distribution of the sum of values and the sum of absolute values of the errors; distribution of the related *T, Statistica,* 39, 443–454.

George, E. O., and Rousseau, C. C. (1987). On the logistic midrange, *Annals of the Institute of Statistical Mathematics, 39,* 627–635.

Govindarajulu, Z. (1963). Relationships among moments of order statistics in samples from two related populations, *Technometrics,* 5, 514–518.

Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population, ***Journal of the American Statistical Association*, 61,** 248–258. (Correction: **71,** 255.)

Greenwood, J. A., Olkin, I., and Savage, I. R. (1962). Index to ***Annals of Mathematical Statistics*,** Volumes 1–31, 1930–1960. University of Minnesota, Minneapolis St. Paul: North Central Publishing.

Grice, J. V., Bain, L. J., and Engelhardt, M. (1978). Comparison of conditional and unconditional confidence intervals for the double exponential distribution, ***Communications in Statistics—Simulation and Computation*, 7,** 515–524.

Hall, D. L., and Joiner, B. L. (1983). Asymptotic relative efficiencies of R-estimators of location, ***Communications in Statistics—Theory and Methods*, 12,** 739–763.

Harter, H. L., Moore, A. H., and Curry, T. F. (1979). Adaptive robust estimation of location and scale parameters of symmetric populations, ***Communications in Statistics—Theory and Methods*,** 8, 1473–1492.

Harvey, H. (1967). ***A Family of Awrages in Statistics*,** Morrisville, PA: Annals Press.

Hausdorff, F. (1901). Beiträge zur Wahrscheinlichkeitsrechnung, ***Verhandlungen der Konigliche Siichsischen Gesellschaft der Wissenschaften, Leipzig, Mathematisch-Physische Classe*, 53,** 152–178.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (eds.) (1983). ***Understanding Robust and Exploratory Data Analysis*,** New York: Wiley.

Holla, M. S., and Bhattacharya, S. K. (1968). On a compound Gaussian distribution, ***Annals of the Institute of Statistical Mathematics*, 20,** 331–336.

Hombas, V. C. (1986). The double exponential distribution: Using calculus to find a maximum likelihood estimator, ***The American Statistician*, 40,** 178.

Horn, P. S. (1983). A measure of peakedness, ***The American Statistician*, 37,** 55–56.

Hsu, D. A. (1979). Long-tailed distributions for position errors in navigation, ***Applied Statistics*, 28,** 62–72.

Hwang, J. T., and Chen, J. (1986). Improved confidence sets for the coefficients of a linear model with spherically symmetric errors, ***Annals of Statistics*, 14,** 444–460.

Iliescu, D. V., and Vodă, V. Gh. (1973). Proportion-p estimators for certain distribution, ***Statistica*, 33,** 309–321.

Jakuszenkow, H. (1979). Estimation of the variance in the generalized Laplace distribution with quadratic loss function, ***Demonstratio Mathematica*, 12,** 581–591.

Johnson, N. L. (1954). Systems of frequency curves derived from the first law of Laplace, ***Trabajos de Estadística*, 5,** 283–291.

Jones, P. N., and McLachlan, G. J. (1990). Laplace-normal mixtures fitted to wind shear data, ***Journal of Applied Statistics*, 17,** 271–276.

Joshi, S. N. (1984). Expansion of Bayes risk in the case of double exponential family, ***Sankhyā*, *Series A*, 46,** 64–74.

Kacki, K., and Krysicki, W. (1967). Die Parameterschatzung einer Mischung von zwei Laplaceschen Verteilungen (im allgemeinen Fall), ***Roczniki Polskiego Towarzystwa Matematycznego, Seria I: Prace Matematyczne*, 11,** 23–31.

Kanji, G. K. (1985). A mixture model for wind shear data, ***Journal of Applied Statistics*. 12.** 49–58.

Kantam, R. R. L., and Narasimham, V. L. (1991). Linear estimation in reflected gamma distribution, ***Sankhyii, Series B*, *53*,** 25–47.

Kappenman, R. F. (1975). Conditional confidence intervals for double exponential distribution parameters, ***Technometrics*, 17,** 233–236.

Kappenman, R. F. (1977). Tolerance intervals for the double exponential distribution, ***Journal of the American Statistical Association*, 72,** 908–909.

Karst, O. J., and Polowy, H. (1963). Sampling properties of the median of a Laplace distribution, ***American Mathematical Monthly*, 70,** 628–636.

Keynes, J. M. (1911). The principal averages and the laws of error which lead to them, ***Journal of the Royal Statistical Society, Series A*, 74,** 322–328.

Khan, A. H., and Khan, R. V. (1987). Relations among moments of order statistics in samples from doubly truncated Laplace and exponential distributions, ***Journal of Statistical Research*, *21*,** 35–44.

Kotz, S., Johnson, N. L., and Read, C. B. (1985). Log-Laplace distribution, In ***Encyclopedia of Statistical Sciences*,** vol. 5 (eds., S. Kotz, N. L. Johnson, and C. B. Read), New York: Wiley, pp. 133–134.

Kotz, S., and Ostrovskii, I. (1994). A mixture representation of the Linnik distribution, ***Statistics & Probability Letters*** (to appear).

Kotz, S., Ostrovskii, I., and Hayfavi, A. (1994). Analytic and asymptotic properties of Linnik's probability densities, ***Submitted for publication.***

Krysicki, W. (1966). Zastosowani metody momentów do estymacji parametrów mieszaniny dwóch rozkladów Laplace'a, ***Zeszyty Naukowe Politechniki Lódzkiej*,** 59, 5–13.

Laplace, P. S. (1774). Mémoire sur la probabilitk des causes par les èvénemens, ***Mémoires de Mathématique et de Physique*, *6*,** 621–656.

Lien, D. H. D., Balakrishnan, N., and Balasubramanian, K. (1992). Moments of order statistics from a non-overlapping mixture model with applications to truncated Laplace distribution, ***Communications in Statistics—Theory and Methods*, 21,** 1909–1928.

Lin, G. D. (1994). Characterizations of the Laplace and related distributions via geometric compound, ***Sankhyā*, *Series A*, *56*,** 1–9.

Lin, P. K. H., Richards, D. O., Long, D. R., Myers, M. D., and Taylor, J. A. (1983). Tables for computing shortest confidence intervals involving the F-distribution, ***Communications in Statistics—Simulation and Computation*, 12,** 711–725.

Ling, K. D. (1977). Bayesian predictive distribution for sample from double exponential distribution, ***Nanta Mathematica*, 10,** 13–19.

Ling, K. D., and Lim, S. K. (1978). On Bayesian predictive distribution for samples from double exponential distribution based on grouped data, ***Nanta Mathematica*, 11,** 192–201.

Lingappaiah, G. S. (1988). On two-piece double exponential distributions, ***Journal of the Korean Statistical Society*, 17,** 46–55.

Loh, W.-Y. (1984). Random quotients and robust estimation, ***Communications in Statistcs—Theory and Methods*, 13,** 2757–2769.

Lukacs, E. (1970). ***Characteristic Functions*,** Second edition, London: Griffin.

Manly, B. F. J. (1976). Some examples of double exponential fitness functions, ***Heredity*, *36*,** 229–234.

Mantel, N. (1969). More light bulb statistics, *The American Statistician*, 23, 21–23.

Mantel, N. (1970). A characteristic function exercise, *The American Statistician*, 24(4), 50.

Mantel, N. (1987). The Laplace distribution and 2 by 2 unit normal determinants, *The American Statistician*, 41, 88.

Mantel, N., and Pasternack, B. S. (1966). Light bulb statistics, *Journal of the American Statistical Association*, 61, 633–639.

McGill, W. J. (1962). Random fluctuations of response rate, *Psychometrika*, 27, 3–17.

Missiakoulis, S., and Darton, R. (1985). The distribution of 2 by 2 unit normal determinants, *The American Statistician*, 39, 241.

Nicholson, W. L. (1958). On the distribution of 2 × 2 random normal determinants, *Annals of Mathematical Statistics*, 29, 575–580.

Norton, R. M. (1984). The double exponential distribution: Using Calculus to find a maximum likelihood estimator, *The American Statistician*, 38, 135–136.

Nyquist, H., Rice, S. O., and Riordan, J. (1954). The distribution of random determinants, *Quarterly of Applied Mathematics*, 42, 97–104.

Ogawa, J. (1951). Contribution to the theory of systematic statistics, I, *Osaka Mathematical Journal*, 3, 175–213.

Okubo, T., and Narita, N. (1980). On the distribution of extreme winds expected in Japan, *National Bureau of Standards Special Publication 560-1*, 12 pp.

Ord, J. K. (1983). Laplace distribution, In *Encyclopedia of Statistical Sciences*, vol. 4 (eds., S. Kotz, N. L. Johnson, and C. B. Read), New York: Wiley, pp 473–475.

Parker, I. (1988). Transformations and influential observations in minimum sum of absolute errors regression, *Technometrics*, 30, 215–220.

Patel, S. R. (1986). On estimation of finite mixtures of distributions, *Journal of the Indian Statistical Association*, 24, 53–63.

Pillai, R. N. (1985). Semi-α-Laplace distributions, *Communications in Statistics—Theory and Methods*, 14, 991–1000.

Raghunandanan, K., and Srinivasan, R. (1971). Simplified estimation of parameters in a double exponential distribution, *Technometrics*, 13, 689–691.

Ramsey, F. L. (1971). Small sample power functions for nonparametric tests of location in the double exponential family, *Journal of the American Statistical Association*, 66, 149–151. (Correction, 72, 703.)

Rosenberger, J. L., and Gasko, M. (1983). Comparing location estimators: Trimmed means, medians and trimean, In *Understanding Robust and Exploratory Data Analysis*, (eds., D. C. Hoaglin, F. Mosteller, and J. W. Tukey), New York: Wiley, pp. 297–338.

Saleh, A. K. Md. E., Ali, M. Masoom, and Umbach, D. (1983). Estimating the quantile function of a location-scale family of distributions based on few selected order statistics, *Journal of Statistical Planning and Inference*, 8, 75–86.

Sansing, R. C. (1976). The t-statistic for a double exponential distribution, *SIAM Journal of Applied Mathematics*, 31, 634–645.

Sarhan, A. E. (1954). Estimation of the mean and standard deviation by order statistics, Part I, *Annals of Mathematical Statistics*, 25, 317–328.

Sarhan, A. E. (1955). *Ibid.*, Part *III*, *Annals of Mathematical Statistics*, **26**, 576–592.

Sassa, H. (1968). The probability density of a certain statistic in one sample from the double exponential population, ***Bulletin, Tokyo Gakugei University***, **19**, 85–89. (In Japanese.)

Scallan, A. J. (1992). Maximum likelihood estimation for a normal/Laplace mixture distribution, ***The Statistician***, **41**, 227–231.

Schlittgen, R. (1979). The median test and inhomogeneity of slopes, ***Biometrical Journal***, **21**, 287–292.

Shamma, S. E., Amin, R. W., and Shamma, A. K. (1991). A double exponentially weighted moving average control procedure with variable sampling intervals, ***Communications in Statistics—Simulation and Computation***, **20**, 511–528.

Sharma, D. (1984). On estimating the variance of a generalized Laplace distribution, ***Metrika***, **31**, 85–88.

Shyu, J.-C., and Owen, D. B. (1986a). One-sided tolerance intervals for the two-parameter double exponential distribution, ***Communications in Statistics—Simulation and Computation***, **15**, 101–119.

Shyu, J.-C., and Owen, D. B. (1986b). Two sided tolerance intervals for the two-parameter double exponential distribution, ***Communications in Statistics—Simulation and Computation***, **15**, 479–495.

Shyu, J.-C., and Owen, D. B. (1987). β-expectation tolerance intervals for the double exponential distribution, ***Communications in Statistics—Simulation and Computation***, **16**, 129–239.

Smith, J. H. (1947). Estimation of linear functions of cell proportions, ***Annals of Mathematical Statistics***, **18**, 231–254.

Srinivasan, R., and Wharton, R. M. (1982). Confidence bands for the Laplace distribution, ***Journal of Statistical Computation and Simulation***, **14**, 89–99.

Śródka, T. (1964). Estymatory i przedzialy ufności odchylenia standardowego w rozkladzie Laplace'a, ***Zeszyty Naukowe Politechniki Lódzkiej***, **57**, 5–9.

Śródka, T. (1966). Złożenie rozkladu Laplace'a z pewnym uogólnionym rozkladem gamma, Maxwella i Weibulla, ***Zeszyty Naukowe Politechniki Lbdzkiej***, **59**, 21–28.

Stigler, S. M. (1975). Napoleonic statistics: The work of Laplace, ***Biometrika***, **62**, 503–517.

Subbotin, M. T. (1923). On the law of frequency of errors, ***Mathematicheskii Sbornik***, **31**, 296–301.

Sugiura, N., and Naing, M. T. (1989). Improved estimators for the location of double exponential distribution, ***Communications in Statistics—Theory and Methods***, **18**, 541–554.

Takano, K. (1988). On the Lévy representation of the characteristic function of the probability distribution $ce^{-|x|}dx$, ***Bulletin of the Faculty of Science, Ibaraki University***, **20**, 61–65.

Tiao, G. C., and Lund, D. R. (1970). The use of OLUMV estimators in inference robustness studies of the location parameter of a class of symmetric distributions, ***Journal of the American Statistical Association***, **65**, 370–386.

Turner. M. C. (1960). On heuristic estimation methods, ***Biometrics***, **16**, 299–301.

Ulrich, G., and Chen, C.-C. (1987). A bivariate double exponential distribution and its generalization, *ASA Proceedings on Statistical Computing,* 127–129.

Umbach, D., Ali, M. Masoom, and Saleh, A. K. Md. E. (1984). Hypothesis testing for the double exponential distribution based on optimal spacing, *Soochow Journal of Mathematics,* **10,** 133–143.

Uppuluri, V. R. R. (1981). Some properties of the log-Laplace distribution, In *Statistical Distributions in Scientific Work,* vol. 4, (eds., G. P. Patil, C. Taillie and B. Baldessari), Dordrecht: Reidel, pp. 105–110.

Uthoff, V. A. (1973). The most powerful scale and location invariance test of the normal versus the double exponential, *Annals of Statistics,* **1,** 170–174.

van Zwet, W. R. (1964). *Concex Transformations of Random Variables,* Amsterdam: Mathematical Centre Tracts 7, Mathematisch Centrum.

Vasudeva Rao, A., Dattatreya Rao, A. V., and Narasimham, V. L. (1991). Optimum linear unbiased estimation of the scale parameter by absolute values of order statistics in the double exponential and the double Weibull distributions, *Communications in Statistics—Simulation and Computation,* **20,** 1139–1158.

Weida, F. M. (1935). On certain distribution functions when the law of the universe is Poisson's first law of error, *Annals of Mathematical Statistics,* **6,** 102–110.

Wilson, E. B. (1923). First and second law of errors, *Journal of the American Statistical Association,* **18,** 841–851.

Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution, *Journal of Mathematical Psychology,* **15,** 109–144.

Yen, V. C., and Moore, A. H. (1988). Modified goodness-of-fit test for the Laplace distribution, *Communications in Statistics—Simulation and computation,* **17,** 275–281.

Zeckhauser, R., and Thompson, M. (1970). Linear regression with non-normal error terms, *Reciew of Economics and Statistics,* **52,** 280–286.

CHAPTER 25

Beta Distributions

1 DEFINITION

The family of beta distributions is composed of all distributions with probability density functions of form:

$$p_Y(y) = \frac{1}{B(p,q)} \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}, \qquad a \le y \le b, \quad (25.1)$$

with $p > 0$, $q > 0$. It is denoted beta (p, q). This will be recognized as a Pearson Type I or II distribution (see Chapter 12, Section 4.1). If $q = 1$, the distribution is sometimes called a ***power-function*** distribution.

If we make the transformation

$$X = \frac{Y-a}{b-a},$$

we obtain the probability density function

$$p_X(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \qquad 0 \le x \le 1. \qquad (25.2)$$

This is the ***standard form*** of the beta distribution with ***parameters*** p, q. It is the form that will be used in most of this chapter. The ***standard power-function*** density is

$$p_X(x) = px^{p-1}, \qquad 0 \le x \le 1. \qquad (25.2)'$$

Harter (1978) introduced the family of symmetric $(p = q)$ ***standardized*** beta variables with the density function

$$p_X(x) = \left[\frac{\Gamma(2p)}{\Gamma^2(p)\left(2\sqrt{2p+1}\right)^{2p-1}}\right](2p+1-x^2)^{p-1}, \qquad (25.3)$$

$$-\sqrt{2p+1} \le x \le \sqrt{2p+1}.$$

Of course $E[X] = 0$ and $\text{var}(X) = 1$. For $p = 1.5(0.5)4.0$ he provides explicit formulas for the cdfs. The simplest, for $p = 2$, is

$$F_X(x) = \left(\frac{3\sqrt{5}}{100}\right)\left(5x - \frac{1}{3}x^3\right) + \frac{1}{2}, \qquad -\sqrt{5} \le x \le \sqrt{5}. \qquad (25.4)$$

The probability density function of a symmetric beta distribution with parameter p, mean μ, and standard deviation σ is

$$p_X(x) = \frac{\Gamma(2p)}{\sigma\{\Gamma(p)\}^2\left(2\sqrt{2p+1}\right)^{2p-1}}\left[2p+1-\left(\frac{x-\mu}{\sigma}\right)^2\right]^{p-1}, \qquad (25.5)$$

$$\mu - \sigma\sqrt{2p+1} \le x \le \mu + \sigma\sqrt{2p+1}.$$

The probability integral of the distribution (25.2) up to x is the ***incomplete beta function ratio,*** and it is denoted by $I_x(p,q)$ so that

$$I_x(p,q) = \frac{1}{B(p,q)}\int_0^x t^{p-1}(1-t)^{q-1}\,dt \qquad (25.6)$$

The word "ratio," which distinguishes *(25.6)* from the ***incomplete beta function***,

$$B_x(p,q) = \int_0^x t^{p-1}(1-t)^{q-1}\,dt, \qquad (25.7a)$$

is often omitted. A description of the properties of $I_x(p,q)$ is contained in Chapter 1 (Section A5) and in Chapter 3 (Section *6).*

Dutka *(1981)* provides a detailed account of the history of $B_x(p,q)$ and $I_x(p,q)$, tracing it back to *1676,* in a letter from Isaac Newton to Henry Oldenberg. The formula given is the special case of

$$B_x(p,q) = p^{-1}x^p(1-x)^q {}_2F_1(p+q,1;p+1;x), \qquad (25.7b)$$

where ${}_2F_1(\cdot)$ denotes the Gaussian hypergeometric function defined in *Eq. (1.104)* of Chapter *1.*

2 GENESIS AND RANDOM NUMBER GENERATION

in "normal theory" the beta distribution arises naturally as the distribution of $V^2 = X_1^2/(X_1^2 + X_2^2)$, where X_1^2, X_2^2 are independent random variables, and X_j^2 is distributed as χ^2 with ν_j degrees of freedom $(j = 1,2)$ (see Chapter 18). The distribution of V^2 is then a standard beta distribution, as in (25.2), with $p = \frac{1}{2}\nu_1$, $q = \frac{1}{2}\nu_2$. Generally $Y = W_1/(W_1 + W_2)$ has a standard beta distribution with parameters p_1 and p_2 if W_j has the gamma density (see Chapter 17) with parameters (p_j, β) $(j = 1,2)$ (for any $\beta > 0$).

Notice that V^2 and $(X_1^2 + X_2^2)$ are mutually independent. An extension of this result is that if $X_1^2, X_2^2, \cdots, X_k^2$ are mutually independent with X_j^2 distributed as χ^2 with ν_j degrees of freedom $(j = 1,2,\dots,k)$ (see Chapter 18), then

$$V_1^2 = \frac{X_1^2}{X_1^2 + X_2^2}$$

$$V_2^2 = \frac{X_1^2 + X_2^2}{X_1^2 + X_2^2 + X_3^2}$$

$$\vdots$$

$$V_{k-1}^2 = \frac{X_1^2 + \cdots + X_{k-1}^2}{X_1^2 + \cdots + X_k^2}$$

are mutually independent random variables, each with a beta distribution, the values of p, q for V_j^2 being $\frac{1}{2}\sum_{i=1}^{j}\nu_i$, $\frac{1}{2}\nu_{j+1}$, respectively. Under these conditions the product of any consecutive set of V_j^2's also has a beta distribution [see Jambunathan (1954) and Section 8]. This property also holds when the v's are any positive numbers (not necessarily integers). Kotlarski (1962) has investigated general conditions under which products of independent variables have a beta distribution.

The special standard beta distribution with $p = q = \frac{1}{2}$ [known as the arc-sine distribution because $\Pr[X \le x] = (2/\pi)\sin^{-1}\sqrt{x}$ for $0 \le x \le 1$] arises in an interesting way in the theory of "random walks." Suppose that a particle moves along the real line by steps of unit length, starting from zero, it being equally likely that a step will be to the left (decreasing) or right (increasing). Let the random variable T_{2n} denote the number of times in the first 2n steps for which the point is in the interval 0 to 2n inclusive at the conclusion of a step. Then

$$\Pr[T_{2n} = 2k] = \binom{2k}{k}\binom{2n-2k}{n-k}2^{-2n}, \qquad k = 0,1,\dots,n.$$

The ratio $T_{2n}/(2n)$ can be regarded as the fraction of time spent on the positive part of the real line. As n tends to infinity, the limiting distribution of $T_{2n}/(2n)$ is the arc-sine distribution:

$$\lim_{n\to\infty}\left\{\sum_{k\le nx}\Pr[T_{2n}=2k]\right\}=\frac{1}{\pi}\int_0^x t^{-1/2}(1-t)^{-1/2}\,dt=\left(\frac{2}{\pi}\right)\sin^{-1}\sqrt{x}\,. \tag{25.8}$$

Standard beta distributions with $p+q=1$, but $p\neq\frac{1}{2}$, are sometimes called generalized arc-sine distributions. For more details on the arc-sine distribution, see Section 7.

A beta distribution can also be obtained as the limiting distribution of eigenvalues in a sequence of random matrices. Suppose that A_n to be a symmetric $n\times n$ matrix whose elements a_{ij} $(i\le j)$ are independent random variables, all a_{ij}'s with $i\neq j$ having a common distribution, and all a_{ii}'s another common distribution, both distributions being symmetrical about zero with variance σ^2 and with all absolute moments finite. Under these conditions Wigner (1958) has shown that the proportion of eigenvalues of the "normalized" matrix $(2\sigma\sqrt{n})^{-1}\mathbf{A}_n$, which are less than x, tends to the limit

$$2\pi^{-1}\int_{-1}^{x}\sqrt{1-t^2}\,dt$$

as $n\to\infty$. This is of form (25.1) with $a=-1$, $b=1$, $p=q=3/2$. Arnold (1967) has shown that this result holds under much weaker conditions on the distributions of the a_{ij}'s.

A class of distributions that includes the beta $(\frac{1}{2},\frac{1}{2})$ and beta $(2,2)$ distributions, can be generated by the following procedure: Starting with the interval $(0,1)$, observe the value of a random variable X_1 distributed uniformly over $(0,1)$ [i.e., as beta $(1,1)$]. Then choose one of the two subintervals $(0,X_1),(X_1,1)$, with probabilities p, $1-p$ of choosing the longer or shorter one, respectively. Denoting the chosen interval by (L_1,U_1), then observe the value of a random variable X_2, uniformly distributed over (L_1,U_1), and choose as (L_2,U_2), the longer or shorter of the intervals $(L_1,X_2),(X_2,U_1)$ with probabilities p, $1-p$, respectively. Continue in this way, choosing (L_{n+1},U_{n+1}) as the longer or shorter of $(L_n,X_{n+1}),(X_{n+1},U_n)$ with probabilities p, $1-p$, respectively. It is easy to see that as $n\to\infty$, the interval length (U_n-L_n) tends to zero with probability one, and there is a limiting value Y_p, say, to which L_n and U_n tend.

The distribution of $Y_{1/2}$ is beta $(\frac{1}{2},\frac{1}{2})$ [Chen, Lin, and Zame (1981)], and the distribution of Y_1 is beta $(2,2)$ [Chen, Goodman, and Zame (1984)]. It is natural to conjecture that Y_p has an approximate (but not exact) beta distribution for values of p other than $\frac{1}{2}$ or 1. Johnson and Kotz (1994) show that

$$\operatorname{var}(Y_p) = \frac{2(7-6p)}{4(11-6p)} \tag{25.9}$$

and that

$$\beta_2(Y_p) = \frac{3(11-6p)(151-204p+60p^2)}{(7-6p)^2(79-30p)}. \tag{25.10}$$

If Y_p had a beta (α, α) distribution, the value of α giving the correct value for $\operatorname{var}(Y_p)$ would be $2(7-6p)^{-1}$. This would result in a "nominal" value

$$\beta_2 = 3(11-6p)(25-18p)^{-1}. \tag{25.11}$$

Table 25.1 compares values of β_2 from (25.10) (actual) and (25.11) (nominal), for selected values of p.

The agreement between actual and nominal values supports the conjecture that beta $[2(7-6p)^{-1}, 2(7-6p)^{-1}]$ would be a good approximation to the distribution of Y_p. O'Connor, Hook, and O'Connor (1985) came to the same conclusion on the basis of simulations.

Another procedure leading to limiting beta distributions has been described by Kennedy (1988). Values of k independent variables $Z_{n1}, \dots, Z_{nk}$ each uniformly distributed over (L_n, U_n) are observed. The interval (L_{n+1}, U_{n+1}) is then chosen as $(L_n, \max(Z_{n1}, \cdots, Z_{nk}))$,

$$(\min(Z_{n1}, \cdots, Z_{nk}), U_n), \quad \text{or} \quad (\min(Z_{n1}, \cdots, Z_{nk}), \max(Z_{n1}, \cdots, Z_{nk}))$$

with probabilities p, q, r, respectively $(p + q + r = 1)$. Kennedy (1988) showed that if the initial interval is $(0,1)$, the limit to which both L_n and U_n converge (with probability 1) is distributed as beta $(k(p+r), k(q+r))$ over $(0,1)$. [Of course, if the initial interval is (A, B), the limit distribution is beta $(k(p+r), k(q+r))$ over (A, B).] There is an alternative proof, based on moment calculations, in Johnson and Kotz (1993).

Yet another way in which a beta distribution arises is as the distribution of an ordered variable from a rectangular distribution (Chapter 26). If $Y_1, Y_2, \dots, Y_n$ are independent random variables each having the standard

Table 25.1 Actual and Nominal Values of β_2

P	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Actual β_2	1.287	1.315	1.348	1.388	1.438	1.500	1.580	1.687	1.831	2.019	2.143
Nominal β_2	1.320	1.345	1.374	1.408	1.449	1.500	1.563	1.645	1.754	1.909	2.143

rectangular distribution so that

$$p_{Y_j}(y) = 1, \qquad 0 \le y \le 1, \tag{25.12}$$

and the corresponding order statistics are $Y_1' \le Y_2' \le \cdots \le Y_n'$, the sth-order statistic Y_s' has the beta distribution

$$p_{Y_s'}(y) = [B(s, n - s + 1)]^{-1} y^{s-1}(1 - y)^{n-s}, \qquad 0 \le y \le 1. \tag{25.13}$$

Fox (1963) suggested that this result may be used to generate beta-distributed random variables from standard rectangularly distributed variables. By this method only integer values can be obtained for n and n – s. A method applicable for fractional values of n and n – s was constructed by Johnk (1964). He showed that if X and Y are independent standard rectangular variables, then the conditional distribution of $X^{1/n}$, given that $X^{1/n} + Y^{1/r} \le 1$, is a standard beta distribution with parameters n, r + 1, and the conditional distribution of $Y^{1/r}$ is beta with parameters n + 1 and r.

This process involves the calculation of $X^{1/n}$ and $Y^{1/r}$, which may be awkward. If n and/or r are large, then a large number of pairs of values (X, Y) is likely to be needed to ensure $X^{1/n} + Y^{1/r} \le 1$, as pointed out by Pekh and Marchenko (1992). [In fact $\Pr[X^{1/n} + Y^{1/r} \le 1] < 1 - \Pr[X^{1/n} > \frac{1}{2}]\Pr[Y^{1/r} > \frac{1}{2}] < 2^{-n} + 2^{-r}$. Hence, if $\min(n, r) \ge 11$, $\Pr[X^{1/n} + Y^{1/r} \le 1]$ is less than 0.001.] Bánkövi (1964) has suggested a method whereby these calculations may be avoided if n and r are both rational. This consists of selecting integers $a_1, a_2, \cdots, a_M$, $b_1, b_2, \cdots, b_N$ such that

$$n = \sum_{j=1}^{M} a_j^{-1}, \qquad r = \sum_{j=1}^{N} b_j^{-1}.$$

Then using the fact that if $X_1, X_2, \cdots, X_M, Y_1, \cdots, Y_N$ are independent standard rectangular variables, $\max(X_1^{a_1}, X_2^{a_2}, \cdots, X_M^{a_M})$ and $\max(Y_1^{b_1}, Y_2^{b_2}, \cdots, Y_N^{b_N})$ are distributed as $X^{1/n}$, $Y^{1/r}$, respectively.

If n (or r) is not a rational fraction, it may be approximated as closely as desired by such a fraction. Bánkövi has investigated the effects of such approximation on the desired beta variates. The GR method is based on the property that $X = Y/(Y + Z)$ has a beta (p, q) distribution if Y and Z are independent gamma variables with shape parameters p and q, respectively (see the beginning of this section).

Generation of beta random variables based on acceptance/rejection algorithms was studied by Ahrens and Dieter (1974) and Atkinson and Pearce (1976), among others. The latter authors recommend the Forsythe (1972) method, which was originally applied to generate random normal deviates. Chen (1978) proposed a modified algorithm BA: $(p, q > 0)$.

Initialization: Set $\alpha = p + q$. If $\min(p, q) \le 1$, set $\beta = \max(p^{-1}, q^{-1})$; otherwise set $\beta = \sqrt{\{(\alpha - 2)/(2pq - \alpha)\}}$. Set $\gamma = p + \beta^{-1}$.

1. Generate uniform (0, 1) random numbers U_1, U_2, and set $V = \beta \log\{U_1/(1 - U_1)\}$, $W = pe^V$.
2. If $\alpha \log\{\alpha/(q + W)\} + \gamma V - 1.3862944 < \log(U_1^2 U_2)$, go to 1.
3. Deliver $X = W/(q + W)$.

This algorithm is reasonably fast for values of p and q down to about 0.5. More complicated versions (*BB*)(*BC*), also described by Chen (1978), cover all a, b > 0 and offer quicker variate generation speed. Here is

Algorithm BB ($\min(p_0, q_0) > 1$)
Initialization: Set $p = \min(p_0, q_0)$, $q = \max(p_0, q_0)$, $\alpha = p + q$, $\beta = \sqrt{[(\alpha - 2)/(2pq - \alpha)]}$; $\gamma = p + \beta^{-1}$.

1. Generate uniform (0, 1) random numbers U_1, U_2, and set $V = \beta \log\{U_1/(1 - U_1)\}$, $W = pe^V$, $Z = U_1^2 U_2$, $R = \gamma V - 1.3862944$, $S = p + R - W$.
2. If $S + 2.609438 \ge 5Z$, go to 5.
3. Set $T = \log Z$. If $S \ge T$, go to 5.
4. If $R + \alpha \log\{\alpha/(q + W)\} < T$, go to 1.
5. If $p = p_0$, deliver $X = W/(q + W)$; otherwise deliver $X = q/(q + W)$.

Schmeiser and Shalaby (1980) developed three exact methods applicable for $\min(p, q) > 1$ (corresponding to Chen's BB algorithm). One of the methods is a minor modification of the Ahrens and Dieter (1974) algorithm: BNM. All the methods use the property that points of inflexion of the beta density are at

$$x = \frac{(p - 1) \pm [(p - 1)(q - 1)]/(p + q - 3)^{1/2}}{p + q - 2}$$

if these values lie between zero and one, and are real.

A detailed comparison of the various methods, carried out by Schmeiser and Shalaby (1980), shows that BB is the fastest for heavily skewed distributions but yields to BNM for heavy-tailed symmetric distributions. No algorithm does better than BB for the following values of the parameters:

$p = 1.01$	$q = 1.01, 1.50, 2.00, 5.00, 10.00, 100.00$
$p = 1.50$	$q = 1.50, 2.00, 5.00, 10.00, 100.00$
$p = 2.00$	$q = 2.00$
$p = 5.00$	$q = 5.00$
$p = 10.00$	$q = 10.00$
$p = 100.00$	$q = 100.00$

Devroye (1986) contains summaries of methods of generating random variables with beta distributions.

3 PROPERTIES

If X has the standard beta distribution (25.2), its rth moment about zero is

$$\mu'_r = \frac{B(p+r,q)}{B(p,q)} = \frac{\Gamma(p+r)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+r)}$$

$$= \frac{}{(p+q)^{[r]}} \quad \text{(if } r \text{ is an integer)}, \qquad (25.14)$$

where $y^{[r]} = y(y+1)\cdots(y+r-1)$ is the ascending factorial. In particular

$$E[X] = \frac{p}{p+q}, \qquad (25.15a)$$

$$\operatorname{var}(X) = pq(p+q)^{-2}(p+q+1)^{-1}, \qquad (25.15b)$$

$$\alpha_3(X) = \sqrt{\beta_1(X)}$$
$$= 2(q-p)\sqrt{p^{-1}+q^{-1}+(pq)^{-1}}\cdot(p+q+2)^{-1}, \qquad (25.15c)$$

$$\alpha_4(X) = \beta_2(X) = 3(p+q+1)\{2(p+q)^2 + pq(p+q-6)\}$$
$$\times[pq(p+q+2)(p+q+3)]^{-1}, \qquad (25.15d)$$

$$E[X^{-1}] = (p+q-1)(p-1)^{-1}, \qquad (25.15e)$$

$$E[(1-X)^{-1}] = (p+q-1)(q-1)^{-1}. \qquad (25.15f)$$

Pham-Gia (1994) has recently established some simple bounds for $\operatorname{var}(X)$. Specifically, he has shown that $\operatorname{var}(X) < 1/4$, and if the density of X is unimodal (i.e., $p > 1$ and $q > 1$) then $\operatorname{var}(X) < 1/12$; further, if the density of X is U-shaped (i.e., $p < 1$ and $q < 1$), then he has proved that $\operatorname{var}(X) > 1/12$.

Writing $\lambda = (p+q)^{-1}$ and $\theta = p(p+q)^{-1}$, we have the following recurrence relation among the central moments of the standard beta distribution:

$$\mu_{s+1} = -\frac{s\lambda}{1+s\lambda}\mu_s + \theta\sum_{j=1}^{s}\binom{s}{j}\frac{\lambda^j(1-\theta)^j j!}{(1+s\lambda)\cdots(1+[s-j]\lambda)}\mu_{s-j}, \qquad (25.16)$$

with $\mu_0 = 1$, μ_1 $(= E[X - E[X]]) = 0$, $\mu_2 = \lambda\theta(1 - \theta)/(1 + \lambda)$,

$$\mu_3 = \frac{2\lambda^2\theta(1-\theta)}{(1+\lambda)(1+2\lambda)}(1 - 2\theta) \qquad \text{[Mühlbach (1972)]}.$$

The moment-generating function can be expressed as a confluent **hypergeo**metric function [Eq. (1.1211, Chapter 1]:

$$E[e^{tX}] = M(p; p + q; t) \tag{25.17}$$

and, of course, the characteristic function is $M(p; p + q; it)$.

The moment-generating function of $(-\log X)$, where X is a standard beta, is

$$M(t) = E[\exp(-t \log X)] = \frac{B(p-t,q)}{B(p,q)}, \tag{25.17$'$}$$

and the corresponding cumulant-generating function is

$$K(t) = \log\left[\frac{\Gamma(p+q)}{\Gamma(p)}\right] - \log\left[\frac{\Gamma(p+q-t)}{\Gamma(p-t)}\right].$$

The cumulants are

$$\kappa_r = (r-1)!\sum_{j=0}^{q-1}(p+j)^{-r}, \qquad r = 1,2,\dots, \tag{25.17$''$}$$

if q is an integer. In the general case

$$\kappa_r = (-1)^r\left[\psi^{(r-1)}(p) - \psi^{(r-1)}(p+q)\right], \tag{25.17$'''$}$$

where $\psi^{(r-1)}(x) = (d^r/dx^r)\log\Gamma(x)$ is the $(r+1)$-gamma function (see Chapter 1, Section A2).

The mean deviation of X is

$$\delta_1(X) = E(|X - E[X]|) = \frac{2}{B(p,q)}\frac{p^pq^q}{(p+q)^{p+q+1}}. \tag{25.18a}$$

If p = q, the expression reduces to

$$\tilde{\delta}_1(X) = \left[B(p,p)p2^{2p}\right]^{-1}. \tag{25.18b}$$

The authors thank Dr. T. Pham-Gia for pointing out an error in the expression for $\delta_1(X)$ which appeared in the first edition of this volume. [See also Pham-Gia and Turkkan (1992).]

For p and q large, using Stirling's approximation to the gamma function, the mean deviation is approximately

$$\sqrt{\frac{2pq}{\pi(p+q)}}\cdot\frac{1}{p+q}\left\{1+\frac{1}{12}(p+q)^{-1}-\frac{1}{12}p^{-1}-\frac{1}{12}q^{-1}\right\}, \quad (25.19)$$

and

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} \doteqdot \sqrt{\frac{2}{\pi}}\left\{1+\frac{7}{12}(p+q)^{-1}-\frac{1}{12}p^{-1}-\frac{1}{12}q^{-1}\right\}.$$

The mean deviation about the median (m) is

$$\frac{2m^p(1-m)^q}{(p+q)B(p,q)} = 2\operatorname{var}(X)\left\{\frac{1}{B(p+1,q+1)}m^p(1-m)^q\right\}. \quad (25.20)$$

If $p>1$ and $q>1$, then $p_X(x)\to 0$ as $x\to 0$ or $x\to 1$; if $0<p<1$, $p_X(x)\to\infty$ as $x\to 0$; and if $0<q<1$, $p_X(x)\to\infty$ as $x\to 1$. If $p=1$ $(q=1)$, $p_X(x)$ tends to a finite nonzero value as $x\to 0$ (1).

If $p>1$ and $q>1$, the density function has a single mode at $x=(p-1)/(p+q-2)$. If $p<1$ and $q<1$, there is an **antimode** (minimum value) of $p_X(x)$ at this value of x. Such distributions are called U-shaped beta (or Type I or II) distributions. If $(p-1)(q-1)$ is not positive, the probability density function does not have a mode or an **antimode** for $0<x<1$. Such distributions are called J-shaped or reverse J-shaped beta (or Type I) distributions. [**Peleg** and **Normand** (1986) advocate using the reparametrization $am=p-1$, $m=q-1$ so that the mode is at $a/(a+1)$ and does not depend on m. Although they call this a *modified* beta distribution, it is in fact just a regular beta distribution that is differently parametrized.] If $p=q$, the distribution is symmetrical about $x=\frac{1}{2}$.

For all positive values of p and q, there are points of inflexion at

$$\frac{p-1}{p+q-2}\pm\frac{1}{p+q-2}\sqrt{\frac{(p-1)(q-1)}{p+q-3}} \quad (25.21)$$

provided these values are real and lie between 0 and 1. Note that as for all **Pearson** curves, the points of inflexion are equidistant from the modes.

The expected value $p/(p+q)$ depends on the ratio p/q. If this ratio is kept constant, but p and q are both increased, the variance decreases, and the (standardized) distribution tends to the unit normal distribution. Some of the properties of beta distributions described in this section are shown in Figures 25.1*a*, *b*. Note that if the values of p and q are interchanged, the distribution is "reflected" about $x=\frac{1}{2}$.

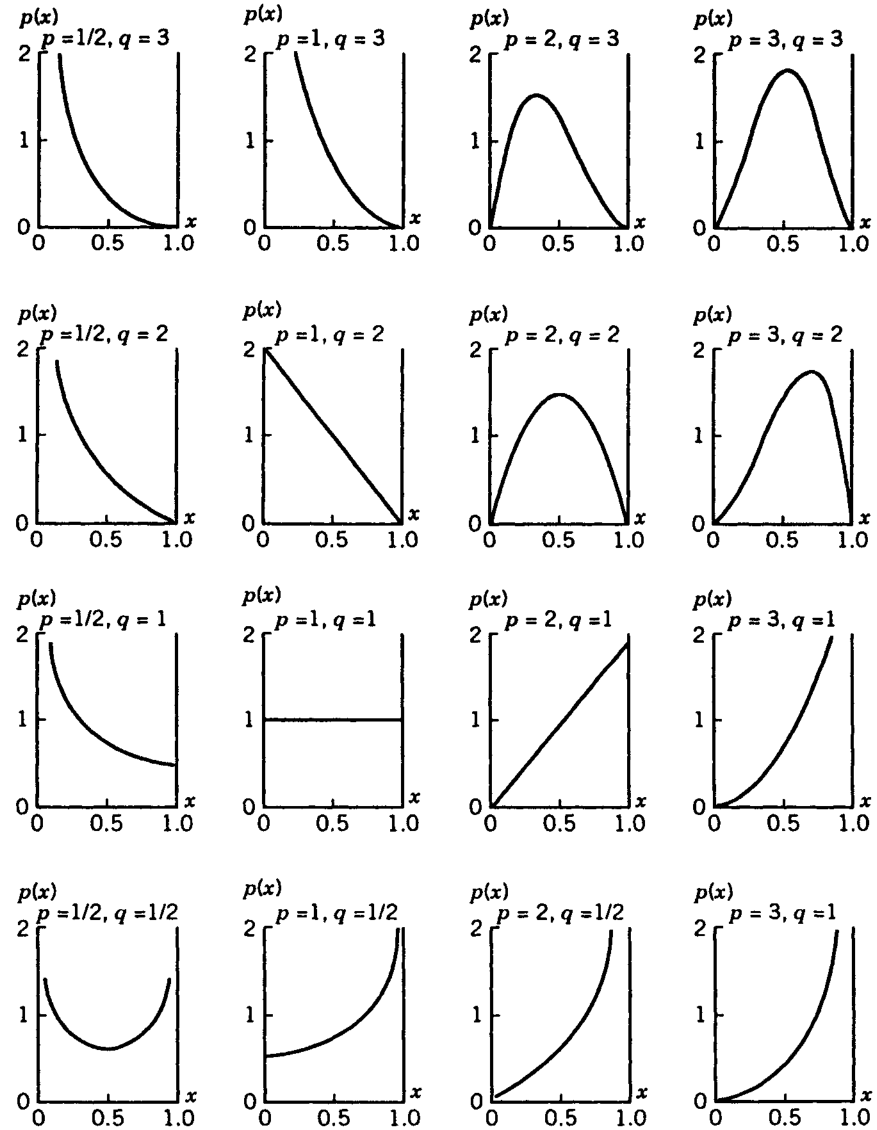


Figure 25.1a Beta density functions

The Lorenz curve [see Chapter 12, Eq. (12.1611 has coordinates

$$[I_x(p,q), I_x(p+1,q)],$$

and the Gini index [Chapter 12, Eq. (12.911 is

$$\frac{2B(2p,2q)}{p[B(p,q)]^2}. \tag{25.22}$$

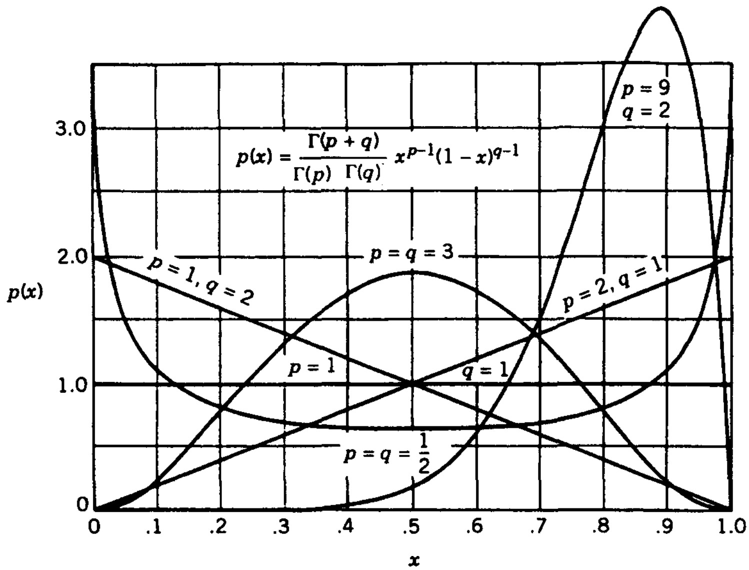


Figure 25.1b Beta density functions

4 ESTIMATION

Discussion of parameter estimation for beta distributions goes back to Pearson's classical paper of 1895 where the method of moments was introduced. Direct algebraic solution of the ML equations cannot be obtained for beta distributions. Koshal (1933, 1935) tackled the ML estimation of four-parameter beta distributions, approximating the actual ML parameter estimates by an interactive method using estimates derived by the method of moments as initial values.

Estimation of all four parameters in distribution (25.1) can be effected by equating sample and population values of the first four moments. Calculation of a, b, p, and q from the mean μ'_1 and central moments μ_2, μ_3, μ_4 uses the following formulas [Elderton and Johnson (1969)]. Putting

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{6 + 3/31 - 2/32},$$

then

$$p, q = \tfrac{1}{2} r \left\{ 1 \pm (r+2) \sqrt{\beta_1 \{ (r+2)^2 \beta_1 + 16(r+1) \}^{-1}} \right\} \qquad (25.23)$$

with $p \lessgtr q$ according as $\alpha_3 = \sqrt{\beta_1} \gtrless 0$. Also

$$\frac{p-1}{q-1} = \frac{\text{mode}(Y) - a}{b - \text{mode}(Y)} \tag{25.24}$$

[where mode $(Y) = a + (b-a)(p-1)/(p+q-2)$] and

$$b - a = \tfrac{1}{2}\sqrt{\mu_2}\sqrt{(r+2)^2\beta_2 + 16(r+1)}\,. \tag{25.25}$$

If the values of a and b are known, then only the first and second moments need to be used, giving

$$\mu_1' = a + (b-a)p/(p+q),$$

$$\mu_2 = (b-a)^2 pq(p+q)^{-2}(p+q+1)^{-1},$$

whence

$$\frac{\mu_1' - a}{b-a} = \frac{p}{p+q}, \tag{25.26}$$

$$\frac{\mu_2}{(b-a)^2} = \frac{p}{p+q}\left(1 - \frac{p}{p+q}\right)\frac{1}{p+q+1}. \tag{25.27}$$

Thus

$$p + q = \frac{[(\mu_1' - a)/(b-a)][1 - (\mu_1' - a)/(b-a)]}{(\mu_2/(b-a)^2)} - 1, \tag{25.28}$$

$$p = \left(\frac{\mu_1' - a}{b-a}\right)^2\left(1 - \frac{\mu_1' - a}{b-a}\right)\left(\frac{\mu_2}{(b-a)^2}\right)^{-1} - \frac{\mu_1' - a}{b-a}. \tag{25.29}$$

The existence, consistency, and asymptotic normality and efficiency of a root of the likelihood equations are usually proved under conditions similar to those given by Cramér (1946) or Kulldorff (1957), which, among other things, allow Taylor expansion of the derivative of the log-likelihood function in a fixed neighborhood of the true parameter value.

When it is necessary to estimate at least one of the end points (a or b) of the four-parameter beta distribution, no such fixed neighborhood of Taylor expansion validity exists. But if the true shape parameters (p and q) are large enough (> 2, regular case), Whitby (1971) has shown that the conditions can be weakened to allow Taylor expansion in a sequence of shrinking neighborhoods, and the usual asymptotic results, with $n^{1/2}$ normalization, can be proved.

If a and b are known and $Y_1, Y_2, \cdots, Y_n$ are independent random variables each having distribution (25.1), the maximum likelihood equations for estimators $\hat{p}, \hat{q}$ of p, q, respectively are

$$\psi(\hat{p}) - \psi(\hat{p} + \hat{q}) = n^{-1} \sum_{j=1}^{n} \log\left(\frac{Y_j - a}{b - a}\right), \tag{25.30a}$$

$$\psi(\hat{q}) - \psi(\hat{p} + 4) = n^{-1} \sum_{j=1}^{n} \log\left(\frac{b - Y_j}{b - a}\right), \tag{25.30b}$$

where $\psi(\cdot)$ is the **digamma** function [Eq. (1.371, Chapter 1]. The **Cramér** and **Kulldorff conditions** cover this case and Eqs. (25.30a) and (25.30b) have to be solved by trial and error. If $\hat{p}$ and $\hat{q}$ are not too small, the approximation

$$\psi(t) \doteqdot \log(t - \tfrac{1}{2})$$

may be used. Then approximate values of $(\hat{p} - \frac{1}{2})/(\hat{p} + \hat{q} - \frac{1}{2})$ and $(\hat{q} - \frac{1}{2})/(\hat{p} + \hat{q} - \$1$ can be obtained from (25.30a) and (25.30b), whence follow, as first approximations to p and q,

$$\hat{p} \doteqdot \frac{\frac{1}{2}\left\{1 - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}\right\}}{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - \text{a})\right)^{\text{''''}} - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}}, \tag{25.31a}$$

$$\hat{q} \doteqdot \frac{\frac{1}{2}\left\{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - a)\right)^{1/n}\right\}}{1 - \Pi_{j=1}^{n}\left((Y_j - a)/(b - \text{a})\right)^{1/n} - \Pi_{j=1}^{n}\left((b - Y_j)/(b - a)\right)^{1/n}}. \tag{25.31b}$$

Starting from these values, solutions of (25.30a) and (25.30b) can be obtained by an iterative process. Gnanadesikan, **Pinkham**, and Hughes (1967) give exact numerical solutions for a few cases.

The asymptotic covariance matrix of $\sqrt{n}\,\hat{p}$ and $\sqrt{n}\,\hat{q}$ (as $n \to \infty$) is

$$[\psi'(p)\psi'(q) - \psi'(p+q)\{\psi'(p) + \psi'(q)\}]^{-1} \times \begin{pmatrix} \psi'(q) - \psi'(p+q) & \psi'(p+q) \\ \psi'(p+q) & \psi'(p) - \psi'(p+q) \end{pmatrix}. \tag{25.32}$$

Introducing approximations for $\psi'(\cdot)$, we have for large values of p and q,

$$\operatorname{var}(\hat{p}) \doteq p(2p-1)n^{-1},$$

$$\operatorname{var}(\hat{q}) \doteq q(2q-1)n^{-1},$$

$$\operatorname{corr}(\hat{p},\hat{q}) \doteq \sqrt{(1-2p^{-1})(1-2q^{-1})}\,. \tag{25.33}$$

Fielitz and Myers (1975, 1976) and Romesburg (1976), in brief communications, discuss the comparative advantages and disadvantages of the method of moments versus the maximum likelihood method for estimating parameters p and q. The difficulties involved in the maximum likelihood method are related to employing efficient search procedures to maximize the likelihood function. The Newton-Raphson method is extremely sensitive to the initial values of $\tilde{p}$ and $\tilde{q}$, and there is no guarantee that convergence will be achieved. Fielitz and Myers (1976) point out that for the sample problem considered by Gnanadesikan, Pinkham, and Hughes (1967), the method of moments yield more accurate estimates of p and q than does the method of maximum likelihood, possibly due to bias introduced by the computational method used in determining the ML estimators.

Beckman and Tietjen (1978) have shown that the equations (25.30a) and (25.30b) can be reduced to a single equation for $\hat{q}$ alone:

$$\psi(\hat{q}) - \psi\{[\psi^{-1}[\log G_1 - \log G_2 + \psi(\hat{q})]] + \hat{q}\} - \log G_2 = 0, \tag{25.34a}$$

where

$$G_1 = \prod_{j=1}^{n}\left(\frac{Y_j - a}{b - a}\right)^{1/n},$$

$$G_2 = \prod_{j=1}^{n}\left(\frac{b - Y_j}{b - a}\right)^{1/n}.$$

Having solved (25.34a) for $\hat{q}$, the estimator, $\hat{p}$, of p is calculated from

$$\hat{p} = \psi^{-1}[\log G_1 - \log G_2 + \psi(\hat{q})]. \tag{25.34b}$$

Lau and Lau (1991) provide a detailed investigation of methods of calculating good *initial* estimators p,, q, of p and q, respectively.

For $G_1 + G_2 = G_T \le 0.95$ they recommend

$$\log p_e = -3.929 + 10.523G_2 - 3.026G_1^3 + 1.757\exp\left(G_2\sqrt{G_1}\right) \tag{25.35a}$$

and

$$\log q_e = -3.895 + 1.222\sqrt{G_2} - 6.9056G_1^3 + 39.057G_1^2G_1^3 + 1.5318\exp(G_T). \quad (25.35b)$$

But for $0.95 \leq G_T \leq 0.999$ they suggest

$$\log p_e = 110706.79 + 3.0842\sqrt{G_1} + 110934.01\log G_T + 6.3908\exp(G_1G_2^2) - 233851.3G_T + 45300.7\exp(G_T) \quad (25.35c)$$

and

$$\log q_e = 113753.4 - 2.1G_1^2 + 113979.94\log G_T + 2.154G_1G_2^6 - 240149.9G_T + 46500.7\exp(G_T). \quad (25.35d)$$

They also study the sampling distribution of the ML estimators $\hat{p}$ and $\hat{q}$ and provide a table of the sample values of percentage bias d = 100 × $(m - p)/p$, where $m = \Sigma p_e/K$ [computed for K = 1000 values of n, p, q (n = 30, 60, 100, $p(=q)$ = 2, 6, 10, 20, and 40)], skewness $K^{-1}\Sigma(p_e - m)^3/S^3$ = a, and kurtosis $b_2 = K^{-1}\Sigma(p_e - m)^4/S^4$, where $S^2 = \Sigma(p_e - m)^2/K$.

For p = q = 10 representative values are

n =	30	100
d	11.1%	3.1%
a_1	1.17	0.59
b_2	5.6	3.7

The same authors also provide a procedure for estimating a confidence interval for p,, using Bowman and Shenton's (1979a, 1979b) method for calculating fractiles of distributions belonging to Pearson's system.

If a and b are unknown, and maximum likelihood estimators of a, b, p, and q are required, the procedure based on (25.31a) and (25.31b) can be repeated using a succession of trial values of a and b, until the pair (a, b), for which the maximized likelihood (given a and b) is as great as possible, is attained.

Carnahan (1989) investigated in detail maximum likelihood estimation for four-parameter beta distributions. He adds to (25.31a) and (25.31b) the additional ML equations

$$\frac{1}{n(p-1)} \cdot \frac{\partial \log L}{\partial a} = \frac{p+q-1}{p-1} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b-a}{Y_i-a}\right) = 0, \quad (25.31c)$$

and

$$-\frac{1}{n(q-1)} \cdot \frac{\partial \log L}{\partial b} = \frac{p+q-1}{q-1} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b-a}{b-Y_i}\right) = 0. \quad (25.31d)$$

(Note that these are essentially "method of moments" expressions, relating the sample values of the harmonic means $E[(Y-a)^{-1}]^{-1}, E[(b-Y)^{-1}]^{-1}$ to the corresponding theoretical values; see Section 2.) Unfortunately, the likelihood function for the distribution is unbounded and has a global maximum that is infinite, so values of a "near" to Y_1', and of b "near" to Y_n' must be excluded. There is also a possibility of local maxima, which may not be well defined for small sample sizes and which plague various numerical schemes for maximizing likelihood. The ML estimators are asymptotically normal and unbiased with variances asymptotically equal to the Cramér-Rao lower bounds provided that $\min(p,q) > 2$. However, a numerical study by Carnahan indicates that only for very large sample sizes ($n \geq 500$) does the bias become small and the Cramér-Rao bound become a good approximation to the variance. The author recommends employing the least and greatest order statistics to improve the estimates of the end points.

The information matrix, from which the asymptotic variances and covariances of ML estimates can be obtained, (in the regular case of $p, q > 2$) is

$$\mathbf{I} = n\begin{pmatrix} \frac{q(p+q-1)}{(p-2)(b-a)^2} & \frac{(p+q-1)}{(b-a)^2} & \frac{q}{(p-1)(b-a)} & -\frac{1}{b-a} \\ \frac{p+q-1}{(b-a)^2} & \frac{p(p+q-1)}{(q-2)(b-a)^2} & \frac{1}{b-a} & -\frac{p}{(q-1)(b-a)} \\ \frac{q}{(p-1)(b-a)} & \frac{1}{b-a} & -\psi'(p+q)+\psi'(p) & -\psi'(p+q) \\ -\frac{1}{b-a} & \frac{-p}{(q-1)(b-a)} & -\psi'(p+q) & \psi'(p+q)+\psi'(q) \end{pmatrix}. \quad (25.36)$$

The diagonal elements of $\mathbf{I}^{-1}$ are the asymptotic variances of the parameter estimates. Explicit inversion has not been attempted. Carnahan (1989) provides numerical results.

AbouRizk, Halpin, and Wilson (1993) [see also AbouRizk, Halpin, and Wilson (1991)], using their program "Beta Fit," compare several methods for estimating the parameters of four-parameter beta distributions as in (25.1) (which they term "generalized beta distributions"). Among these were the following:

1. Moment methods. Equating first four sample and population moments; next taking $a = Y_1'$ and $b = Y_n'$ and using only the first two moments [e.g., as suggested by Riggs (1989)].
2. "Feasibility moment matching" method. Minimizing the (unweighted) sum of squares of differences between sample and population means, variances, skewness ($\sqrt{b_1}$ and $\sqrt{\beta_1}$) and kurtosis (b_2 and β_2), subject to $a < Y_1'$ and $b > Y_n'$, and possibly other restrictions (e.g., $a > 0$, $b > 0$).
3. Maximium likelihood method. Maximizing with arbitrary values of a and b [as described formally in (25.30)]; any variation of a and b is not considered.
4. "Regression-based" methods [see Swain, Venkataraman, and Wilson (1988)]. Using order statistics and the relationships [see Chapter 12, Eq. (12.20)]

$$E\left[F_Y(Y_j')\right] = \frac{j}{n+1},$$

$$\operatorname{var}\left(F_Y(Y_j')\right) = \frac{j(n-j+1)}{(n+1)^2(n+2)},$$

$$\operatorname{cov}\left(F_Y(Y_j'), F_Y(Y_k')\right) = \frac{j(n-k+1)}{(n+1)^2(n+2)}, \qquad j < k.$$

Two variants of a least-squares method are used in minimizing $\sum_{j=1}^{n} w_j\{F_Y(Y_j') - j/(n+1)\}^2$ with respect to a, b, p, and q.

Case 1. $w_j = 1$ for all j ("ordinary least squares"),
Case 2. $w_j = \{\operatorname{var}(F_Y(Y_j'))\}^{-1}$, $j = 1, \dots, n$ ("diagonally weighted least squares").

In each case minimization is subject to the restrictions $a < Y_1'$, $b > Y_n'$ ($a > 0$, $b > 0$), as in the third method mentioned above.

Dishon and Weiss (1980) provide small sample comparisons of maximum likelihood and moment estimators for standard beta distributions (25.2) ($a = 0$ and $b = 1$): The maximum likelihood estimators $\hat{p}$ and $\hat{q}$, which in this case are solutions of the equations,

$$\psi(\hat{p} + \hat{q} + 2) - \psi(\hat{p} + 1) = \frac{1}{n} \sum \log\left(\frac{1}{X_i}\right) \tag{25.37a}$$

and

$$\psi(\hat{p} + 4 + 2) - \psi(\hat{q} + 1) = \frac{1}{n} \sum \log\left(\frac{1}{1 - X_i}\right) \tag{25.37b}$$

are compared with the moment estimators

$$\tilde{p} = \frac{\tilde{\mu}'_1(\tilde{\mu}'_1 - \tilde{\mu}'_2)}{\tilde{\mu}'_2 - (\tilde{\mu}'_1)^2} - 1 \tag{25.38a}$$

and

$$\tilde{q} = \frac{(1 - \tilde{\mu}'_1)(\tilde{\mu}'_1 - \tilde{\mu}'_2)}{\tilde{\mu}'_2 - (\tilde{\mu}'_1)^2} - 1, \tag{25.38b}$$

where $\tilde{\mu}'_1$ and $\tilde{\mu}'_2$ are estimators of the first and second moments, respectively.

The results are summarized in Table 25.2. The authors generated a number of beta variables with known values of p and q and calculated, for different values of sample size n with 100 replications, $\hat{p}$, $\hat{q}$, $\tilde{p}$, and $\tilde{q}$ by means of the equations given above. (Both estimators tend to produce errors of the same sign.) They defined

$$R_p = \frac{\sum_{j=1}^{100}(\hat{p}_j - p)^2}{\sum_{j=1}^{100}(\tilde{p}_j - p)^2}$$

(with an analogous definition for R_q), where $\hat{p}_j$ is the ML estimator of p and $\tilde{p}_j$ is the moment estimator of p in the jth replication. [The authors also develop efficient procedures for computing $\psi(z)$ using the expansion

$$\psi(1 + z) = -\gamma + \sum_{n=1}^{\infty} \frac{z}{n(n + z)}, \qquad z \neq -1, -2, -3, \dots,$$

where $\gamma = 0.57722\dots$ is Euler's constant, as defined in Chapter 1, Eq. (1.19), and the Euler-Maclaurin summation formula.] The data in the table show that when n is low, the ML estimator is usually more accurate than the moment estimator (with notable exception when $p = q$).

Table 25.2 Comparison of estimates obtained from ML and moment estimates for a univariate beta distribution. Each row is for 100 replications

Parameter Values		Sample Size	N^a			
p	q	n	p	q	R_p	R_q
$-\frac{1}{2}$	$-\frac{1}{2}$	25	58	56	0.935	0.888
		50	58	64	0.911	0.799
		100	53	57	0.805	0.847
$-\frac{1}{2}$	1	25	64	61	0.793	0.802
		50	70	57	0.765	0.953
		100	62	56	0.646	0.829
1	1	25	42	44	1.020	1.020
		50	48	51	1.004	0.977
		100	51	50	0.962	0.975
$-\frac{1}{2}$	5	25	75	66	0.663	0.778
		50	66	63	0.564	0.706
		100	61	59	0.728	0.758
5	1	25	57	56	0.984	0.984
		50	57	59	0.932	0.912
		100	55	51	0.961	0.940
5	5	25	44	46	1.007	1.000
		50	41	42	1.017	1.021
		100	58	63	0.980	0.970
10	5	25	54	58	1.000	0.996
		50	57	58	0.996	0.989
		100	51	59	0.984	0.981
$-\frac{1}{2}$	100	25	64	67	0.806	0.852
		50	70	67	0.777	0.840
		100	76	68	0.693	0.801
1	100	25	56	61	0.915	0.889
		50	70	70	0.837	0.833
		100	62	64	0.914	0.896
50	100	25	53	54	0.996	0.996
		50	55	56	0.992	0.993
		100	50	50	1.000	1.000
100	100	25	57	55	0.999	0.999
		50	43	47	1.000	1.000
		100	57	57	1.000	1.000

[a] N = number of cases in which the MLE is closer to the true value of $p, q,$ than the moment estimator.

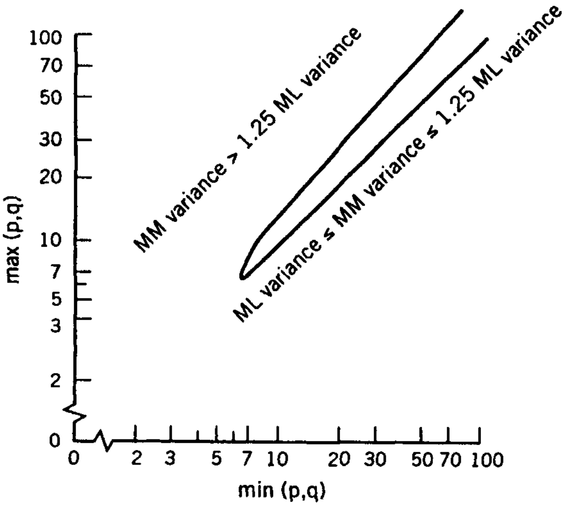


Figure 25.2 Comparison of the variances associated with MM and ML estimators

As Figure **25.2** from Kottes and Lau (1978) indicates, when p and q are small or their difference is large, the (asymptotic) method of moments variance exceeds the (asymptotic) maximum likelihood variance by at least **25%.** These are the situations when the need to fit beta distributions is the greatest. Fortunately, in many cases a and b, or at least one of these parameters, can be assigned known values.

If only the r smallest values $X_1', X_2', \dots, X_r'$ are available, the maximum likelihood equations are

$$\frac{r}{n}\log\left[\left(\prod_{j=1}^{r} X_j'\right)^{1/r}\right] = \psi(\hat{p}) - \psi(\hat{p}+\hat{q}) - \left(1-\frac{r}{n}\right)\frac{\partial}{\partial\hat{p}}\log\left[\int_{X_r'}^{1} t^{\hat{p}-1}(1-t)^{\hat{q}-1}\,dt\right] \quad (25.39a)$$

and

$$\frac{r}{n}\log\left[\left(\prod_{j=1}^{r}(1-X_j')\right)^{1/r}\right] = \psi(\hat{q}) - \psi(\hat{p}+\hat{q}) - \left(1-\frac{r}{n}\right)\frac{\partial}{\partial\hat{q}}\log\left[\int_{X_r'}^{1} t^{\hat{p}-1}(1-t)^{\hat{q}-1}\,dt\right] \quad (25.39b)$$

[Gnanadesikan, Pinkham, and Hughes (1967)].

Fang and Yuan (1990) apply the sequential algorithm for optimization by number theoretic methods (SNTO) proposed by Fang and Wang (1989) to obtain ML estimators of parameters of standard beta distributions. The method is superior to the Newton-Raphson method. It does not require unimodality or existence of derivatives (only continuity of the likelihood) and is not sensitive to the initial values. For the data provided by Gnanadesikan, Pinkham, and Hughes (1967), this method yields more accurate values than those of moment estimators or Gnanadesikan, Pinkham, and Hughes's (1967) estimators.

If one of the values p and q is known, the equations are much simpler to solve. In particular, for the standard power-function distribution ($q = 1$), the maximum likelihood estimator of p is

$$\hat{p} = \left[n^{-1} \sum_{j=1}^{n} \log X_j \right]^{-1}, \tag{25.40}$$

and we have

$$n \operatorname{var} \hat{p} \doteqdot p^2. \tag{25.41}$$

A moment estimator of p in this case is

$$\tilde{p} = \bar{X}(1 - \bar{X})^{-1}, \tag{25.42}$$

for which

$$n \operatorname{var} \tilde{p} \doteqdot p(p+1)^2(p+2)^{-1}. \tag{25.43}$$

Note that $(\operatorname{var} \hat{p})/(\operatorname{var} \tilde{p}) = p(p+2)(p+1)^{-2}$. The asymptotic relative efficiency of $\tilde{p}$ increases with p; it is as high as 75% for p = 1, tends to 100% as $p \to \infty$, but tends to zero as $p \to 0$. There is further discussion of power-function distributions in Chapter 20, Section 8.

Interestingly Guenther (1967) has shown that for the special case of the standard power-function distribution, with the pdf

$$p_{X_j}(x) = px^{p-1}, \qquad 0 < x < 1, j = 1, \dots, n, \tag{25.44}$$

the minimum variance unbiased estimator of p is $-(n-1)[\sum_{j=1}^{n} \log X_j]^{-1}$. Its variance is $p^2(n-2)^{-1}$, while the Cramér-Rao lower bound (Chapter 1, Section B15) is $p^2 n^{-1}$.

In operations research applications (especially in connection with PERT) it is often ***assumed*** that the standard deviation ***must*** be one-sixth of the range of variation (!). Thus, for a standard beta (p, q) distribution (range 0 to 1), it is assumed that

$$\sigma(X) = \tfrac{1}{6}, \tag{25.45}$$

while for the more general distribution (25.1),

$$\sigma(X) = \tfrac{1}{6}(b - a). \tag{25.46}$$

This assumption is used in fitting a beta distribution on the basis of "least possible" (a^*), "greatest possible" (b^*), and "most likely" (m^*) values as estimated from engineers' experience of a process. These are used as estimates of a, b and the modal value

$$m = a + \frac{p-1}{p+q-2}(b-a) \qquad [\min(p,q) > 1], \tag{25.47}$$

respectively [Hillier and Lieberman (1980)].

Values of estimates p^*, q^* of p, q, respectively, can be obtained from the simultaneous equations

$$\frac{p^*q^*}{(p^*+q^*)^2(p^*+q^*+1)} = \frac{1}{36} \qquad \text{(cf. (25.46))} \tag{25.48a}$$

and

$$\frac{p^*-1}{p^*+q^*-2} = \frac{m^*-a^*}{b^*-a^*}. \qquad \text{(cf. (25.47))} \tag{25.48b}$$

It would be, perhaps, more natural to use an estimated expected value, $\bar{X}$, say, and equate that to the population value, leading to

$$\frac{p^*}{p^*+q^*} = \frac{\bar{X}-a^*}{b^*-a^*} \tag{25.48c}$$

in place of (25.48b). In fact it appears to be customary to use an equation like (25.48c) but with $\bar{X}$ replaced by an estimate of the expected value

$$a^* + \frac{1}{6}\left\{\frac{4(m^*-a^*)}{b^*-a^*} + 1\right\}(b^*-a^*) = a^* + \frac{1}{6}(4m^* + b^* - 5a^*),$$

leading to

$$\frac{p^*}{p^* + q^*} = \frac{1}{6}\left\{4\left(\frac{m^* - a^*}{b^* - a^*}\right) + 1\right\}. \qquad (25.48d)$$

From (25.48a), (25.48c), or (25.48d), $(p^* + q^*)$ can be expressed in terms of p*, a*, and b*. Inserting this expression for $(p^* + q^*)$ in (25.48a), we obtain an equation in p*. For example, using (28.48c),

$$p^* + q^* = \left(\frac{b^* - a^*}{\bar{X} - a^*}\right)p^*$$

and

$$\frac{p^*q^*}{(p^* + q^*)^2} = \left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right),$$

whence (25.48a) becomes

$$\left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right) = \frac{1}{36}\left\{\left(\frac{b^* - a^*}{\bar{X} - a^*}\right)p^* + 1\right\},$$

that is,

$$p^* = \frac{\bar{X} - a^*}{b^* - a^*}\left\{36\left(\frac{\bar{X} - a^*}{b^* - a^*}\right)\left(1 - \frac{\bar{X} - a^*}{b^* - a^*}\right) - 1\right\}. \qquad (25.49)$$

Using (28.48d) would also lead to a simple explicit value for p*, but (28.48b) would lead to a cubic equation for p*.

Farnum and Stanton (1987) carried out a critical investigation of the accuracy of the assumption that for a standard beta (p, q) variable

$$\text{Expected value} = \tfrac{1}{6}\{4(\text{mode}) + 1\},$$

that is,

$$\frac{p}{p + q} \doteqdot \frac{1}{6}\left(\frac{4(p - 1)}{p + q - 2} + 1\right) \qquad (25.50)$$

[presumably when (25.48a) is satisfied]. They found the approximation is correct to within 0.02 when the mode is between 0.13 and 0.87, and they suggest using improved approximations

$$\frac{2}{2 + (\text{mode})^{-1}} \quad \text{for mode} < 0.13, \tag{25.51a}$$

and

$$\{3 - 2(\text{mode})\}^{-1} \quad \text{for mode} > 0.87. \tag{25.51b}$$

Moitra (1990) suggested that some allowance should be made for "skewness" (which he measures by $E[(X - E[X])^3]$, rather than the shape factor $\sqrt{\beta_1}$, which would be $6\sqrt{6}E[(X - E[X])^3]$ if $\sigma(X) = 1/6$). Moitra noted that the "traditional" assumptions can be expressed as

$$E[X] = \frac{a + b + k(\text{mode})}{k + 2} \tag{25.52a}$$

and

$$\sigma(X) = c^{-1}(b - a), \tag{25.52b}$$

with $k = 4$, $c = 6$. He found that $c = 6$ is not "optimal for values of k other than 4 or 5, and $k = 4$ is not optimal for values of c other than 6."

Moitra made the following recommendations: "If the skewness is judged or known to be high, p would be between 2 and 3, and since we are estimating subjective distributions, we can set $p = 2.5$." But "if the skewness is judged to be moderate, then we can see from the graphs that p is very likely to be between 3 and 4, and so we can similarly set $p = 3.5$. Finally, if the skewness is considered to be only a little, we set $p = 4.5$." He also provided the "best" combinations of values for k and c, which are given in Table 25.3, and an analysis appropriate to triangular distributions (see Chapter 26, Section 9) for which

$$E[X] = \tfrac{1}{3}(a + b + m). \tag{25.53}$$

Table 25.3 Best combinations of k and c

	k					
c	**1**	**2**	**3**	**4**	**5**	**6**
3	Best					
4		Best	Good			
5			Good	Best		
6				Good	Best	
7						Best
8						Best

In the case a = 0, b = 1, we have

$$E[X] = \tfrac{1}{4}(1 + m), \qquad (25.54a)$$

$$[\sigma(X)]^2 = \tfrac{1}{16}(1 - m + m^2). \qquad (25.54b)$$

The values of $\sigma(X)$ varies from 0.25 (for m = 0 or 1) to 0.22 for m = 0.5. These values are quite close to each other, and somewhat greater than the traditional value (used in connection with beta distributions) of 1/6. Setting k = 1 and c = 4 or 4.5, estimators can be approximated using a triangular distribution. "The advantage of this procedure is that by invoking the triangular distribution we do not have to make any further assumptions, nor do we need any additional information." However, as Table 25.3 indicates, the choice of k = 2 (rather than 1), with c = 4 or 4.5, may be more appropriate.

5 APPLICATIONS

Beta distributions are very versatile (see Figures 25.1*a*, *b*) and a variety of uncertainties can be usefully modeled by them. This flexibility encourages its empirical use in a wide range of applications [see, e.g., Morgan and Henrion (1990)].

The beta distributions are among the most frequently employed to model theoretical distributions. Usually the range of variation (a, *b*) of such distributions is known, and fitting is effected by equating the first and second moments of the theoretical and fitted curve. No random sample values enter into this calculation so that maximum likelihood methods are inapplicable, and a fortiori, arguments based on asymptotic efficiency are irrelevant.

An example of some importance is the use of beta distributions to fit the distributions of certain criteria used in statistical likelihood ratio tests. Usually the range of variation of the likelihood ratio is known to be from zero to one, and that of any monotonic function of the likelihood ratio can be derived from this knowledge. If the likelihood ratio is based on n independent identically distributed random variables, it is often found that a usefully good fit can be obtained by supposing

$$(\text{likelihood ratio})^{2/n}$$

to have a beta distribution with a = 0, b = 1. Use of the power $2n^{-1}$ is suggested by Wilks's theorem that under certain fairly broad conditions, $-2n^{-1}$ log (likelihood ratio) has an asymptotic χ^2 distribution (as $n \to \infty$). (See also Chapter 29, Section 9, where some additional cases are discussed.) Of course a general power c might be used, and c, as well as p and q, if fitted a substantially improved fit could be expected by using this method.

This would be equivalent to fitting a "generalized" beta distribution, to be described in Section 25.7.

A standard Type I distribution has been found to give good approximation when fitted (by equation of first two moments) to relative frequencies of a binomial distribution [Benedetti (1956)]. If the binomial parameters are N, ω, then the approximate value for the probability that the binomial variable is less than r is

$$I_{(r-1/2)}((N-1)\omega, (N-1)(1-\omega)), \qquad (25.55a)$$

as compared with the exact value

$$I_{1-\omega}(N-r+1, r). \qquad (25.55b)$$

[See also Eq. (3.34), Chapter 3.] Numerical comparisons given in Benedetti (1956), and also in Johnson (1960), show that except in the "tails" (probabilities between 0.05 and 0.95) a good practical approximation is obtained for $N \geq 50$ and $0.1 \leq \omega \leq 0.9$.

For many years a fashionable use for beta distributions has been as "prior" distributions for binomial proportions (see Section 2.2, Chapter 6). While this leads to some conveniently simple mathematics, and beta distributions are often referred to as "natural" prior distributions for the binomial parameter p (in that the posterior distribution obtained by its use is also of the same form), there seems to be little definite evidence in their favor. This was pointed out at an early date by Barnard (1957), in the course of discussion of Horsnell (1957). Much more recently a similar remark was made by Ganter (1990), commenting on Hart (1990). Shaw (1991) proposes use of beta priors to reduce the number of tests needed in routine checks on reliability, apparently without initial examination of validity of assumptions. There has been, however, some attention to the need for care in selection of prior distributions. For example, Chaloner and Duncan (1983) describe methods of "elicitation" of parameter values for beta priors, though the basis for using beta distributions in this context is not addressed. Pham-Gia (1994) has recently studied the information gain or loss by considering changes in the reciprocal of the expected posterior variance. Specifically, he has shown that the ratio of the expected values of the posterior variances of two beta distributions provides a convenient criterion that is consistent with many Bayesian results and, in addition, enables the determination of the least informative beta prior distribution.

In recent years beta distributions have been used in modeling distributions of hydrologic variables [Janardan and Padmanabhan (1986)], logarithm of aerosol sizes [Bunz et al. (1987); Van Dingenan, Raes, and Vanmarcke (1987)], activity time in PERT analysis [Golenko-Ginzburg (1988)], errors in forms [Yang, Li, and Li (1988)], isolation data in photovoltaic system analysis [Rahman, Khallat, and Salameh (1988)], porosity/void ratio of soil [Harrop-

Williams (1989)], phase derivatives in communication theory [Andersen, Lauritzen, and Thommesen (1990); Lauritzen, Thommesen, and Andersen (1990)], conductance of catfish retinal zones [Haynes and Yau (1990)], variables affecting reproductivity of cows [McNally (1990)], size of progeny in *Escherchia* Coli [Koppes and Grover (1992)], ventilation from inert gas washout distributions [Meyer, Groebe, and Thews (1990)], dissipation rate in breakage models [Yamazaki (1990)], proportions in gas mixtures [Agrawal and Yang (1991)], sea-state reflectivity [Delignon, Garello, and Hillion (1991)], atmospheric transmittance and related indices for solar radiation [Graham and Hollands (1990); Milyutin and Yaromenko (1991)], clutter and power of radar signals [Maffett and Wackerman (1991); Sopel'nik and Lerchenko (1991), respectively], acoustic emissions in evaluation of chip forms [Sukvittayawong and Inasaki (1991)], traffic flow [Ressel (1991)], construction duration [AbouRizk and Halpin (1992); AbouRizk, Halpin, and Wilson (1991)], particle size [Boss (1992a, b); Popplewell and Peleg (1992)], gas absorption [Karavias and Myers (1992)], and tool wear [Wang and Dornfeld (1992)].

Wiley, Herschokoru, and Padiau (1989) developed a model to estimate the probability of transmission of HIV virus during a sexual contact between an infected and a susceptible individual. Let β be the per contact infectivity associated with sexual contact between an infected and a susceptible individual. The authors model each sexual encounter as an independent event where HIV is transmitted with probability β. Thus, if the couple has n sexual encounters, the probability of T, the infection of the seronegative partner, is

$$\Pr(T|\beta) = 1 - (1 - \beta)^n.$$

To allow for heterogeneity of infectivity among couples, they consider β as a random variable with the beta density function

$$p(\beta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\beta^{a-1}(1-\beta)^{b-1}, 0 < \beta < 1.$$

The marginal distribution of T is then

$$\Pr(T) = E[\Pr(T|\beta)] = 1 - \prod_{j=0}^{n-1}\left(\frac{b+j}{a+b+j}\right).$$

Using data on the number of contacts and seroconversion of couples, they fit this model by the maximum likelihood method. The estimates indicate that heterogeneity is indeed quite extreme.

Thompson (1990) describes applications in the analysis of probabilistic informational retrieval in "expert systems." Treacy et al. (1991) use truncated beta distributions in automated tolerance analysis of assemblies. Beta distributions are widely used in many areas of operations research. Moitra (1990) provides a number of examples in risk analysis for strategic planning, finance and marketing, engineering systems simulation, and decision theory.

Pham and Turkkan (1994) discussed recently the exact distribution of the sum of two independent beta variables and applied it to a standby system with beta-distributed component lives. They show that this enables the calculation of the exact reliability of such a system when the exact values of the parameters are known, and also greatly improves the available approximate methods of computing the reliability.

6 APPROXIMATIONS AND TABLES

6.1 Approximations

Several approximations to the incomplete beta function ratio $I_x(p,q)$ have been described in Section 6.1 of Chapter 3. Relevant references, given at the end of this chapter for convenience, include Aroian (1941, 1950), Cadwell (1952), Hartley and Fitch (1951), Nair (1948), Pearson and Pearson (1935), Thomson (1947), and Wise (1950, 1960). Here we add only several further approximations which came to our attention subsequent to the completion of the first edition of ***Discrete Distributions.*** The first approximation is one of a number proposed in Peizer and Pratt (1968) and Pratt (1968). For the incomplete beta function ratio, their approximations are

$$I_x(p,q) = \Phi(z),$$

where

$$z = \frac{d}{\left|q - \frac{1}{2} - n(1-x)\right|}\left\{\frac{2}{1+(6n)^{-1}}\left[\left(q-\frac{1}{2}\right)\log\left\{\frac{q-}{n(1-x)}\right\} + \left(p-\frac{1}{2}\right)\log\left\{\frac{p-(1/2)}{nx}\right\}\right]\right\}^{1/2}, \tag{25.56}$$

with $n = p + q - 1$. The value of d is either

$$d = q - \tfrac{1}{3} - \left(n+\tfrac{1}{3}\right)(1-x) \qquad \text{(case 1)}$$

or

$$d = q - \frac{1}{3} - \left(n + \frac{1}{3}\right)(1-x) + \overline{5}\ \bar{q} \quad - \frac{1-x}{P} + \frac{x-(1/2)}{p+q} \qquad \text{(case 2)}$$

Case 2 generally gives the more accurate results. Using this value of d, the error of $I_x(p,q)$ is less than 0.001 for $p, q \geq 2$ and less than 0.01 for

$p, q \geq 1$. Limits on the ***proportional*** error are

$$\frac{|\Phi(z) - I_x(p,q)|}{I_x(p,q)} < \begin{cases} 0.01 & \text{if } p,q \geq 3 \text{ and } 0.2 \leq R \leq 5.0, \\ 0.02 & \text{if } p,q \geq 1.75 \text{ and } 0.125 \leq R \leq 8, \\ 0.03 & \text{if } p,q \geq 1.5 \text{ and } 0.1 \leq R \leq 10, \end{cases} \tag{25.57}$$

where

$$R = \frac{(q - (1/2))x}{(p - (1/2))(1 - x)}.$$

Mudholkar and Chaubey *(1976)* provide Patnaik-type, Pearson-type, and Sankaran-type approximations to $I_x(p,q)$ based on the distribution of $-\log X$ when X has a standard beta (p,q) distribution. We have of course

$$I_x(p,q) = \Pr[-\log X > -\log x].$$

The sth cumulant of $-\log X$ is

$$\kappa_r(-\log X) = (-1)^r\{\psi^{(r-1)}(p) - \psi^{(r-1)}(p+q)) \tag{25.58}$$

[see (25.17)‴].

1. ***Patnaik-type approximation.*** Approximating $-\log X$ by $c\chi_\nu^2$, where c and ν are chosen to give the correct first two moments,

$$c = \frac{1}{2}\frac{\psi'(p) - \psi'(p+q)}{\psi(p+q) - \psi(p)}, \quad \nu = \frac{2\{\psi(p+q) - \psi(p)\}^2}{\psi'(p) - \psi'(p+q)},$$

we obtain

$$I_x(p,q) \doteqdot \Pr\left[\chi_\nu^2 \geq \frac{-\log x}{c}\right]. \tag{25.59a}$$

Using the Wilson-Hilferty approximation [Chapter 18, Eq. (18.26)], we obtain

$$I_x(p,q) \doteqdot 1 - \Phi(z), \tag{25.59b}$$

where

$$z = \left[\left(\frac{-\log x}{c\nu}\right)^{1/3} - \left(1 - \frac{2}{9\nu}\right)\right]\left(\frac{2}{9\nu}\right)^{-1/2}.$$

2. Pearson-type approximation. Approximating $-\log X$ by $(c'\chi^2_{\nu'} + b)$ with c', ν', and b chosen to give the correct first three moments, we obtain (25.59a) with c, ν replaced by c', ν', respectively, and $-\log x$ replaced by $-(\log x + b)$.
3. Sankaran-type approximation. We determine h so that the leading term in the third cumulant of $\{(-\log X)/\kappa_1\}^h$ vanishes. This value of h is $\{1 - \kappa_1\kappa_3/(3\kappa_2^2)\}$. Approximating $\{(-\log X)/\kappa_1)\}^h$ by a normal variable with

$$\text{Expected value} = 1 - \frac{h\kappa_3}{6\kappa_1\kappa_2} = \mu$$

and

$$\text{Variance} = \frac{h^2\kappa_2}{\kappa_1^2} = \sigma^2,$$

we obtain (25.59b) with

$$z = \frac{\{(-\log x)/\kappa_1\}^h - \mu}{\sigma}. \tag{25.60}$$

The next approximation is of a more general nature. Consider a standard beta $(kn,$ In) distribution. Keeping k and l fixed, let $n \to \infty$. It is known that the standardized density function tends to a unit normal $[N(0,1)]$ density. Królikowska (1966) investigated the behavior of the leading term of the absolute difference between the standardized beta and the unit normal distribution as $n \to \infty$. She found that it is of order $n^{-1/2}$, except when k = I, in which case it is of order n^{-1}.

Volodin (1970) obtained the approximate formula

$$I_x(p,q) \doteqdot \frac{qx}{p+q}\left\{{}_2F_1(1,1;1+p;1-x) + \frac{p}{1+q^2}\,{}_2F_1(1,1;2+q;x)\right\}, \tag{25.61}$$

where ${}_2F_1(a, b; c; x)$ is a hypergeometric function (see Chapter 1, Section A6). It is very accurate for small values of p and q. The approximation is based on the observation that the characteristic function of the random variable

$$W = (p+q)\log\left\{\frac{X}{1-X}\right\}, \tag{25.62}$$

where X has distribution (25.1), is

$$\frac{\Gamma(1+p+[p+q]it)\Gamma(1+q-[p+q]it)}{\{1+it(p+q)/p\}\{1-it(p+q)/q\}/\{\Gamma(1+p)\Gamma(1+q)\}}. \quad (25.63)$$

This shows that W can be regarded as the sum of three independent variables, W_1, W_2, and W_3, where W_1 and W_2 have exponential distributions (see Chapter 19) with parameters $p/(p+q)$ and $q/(p+q)$, respectively, and W_3 has the standard beta $(p+1, q+1)$ distribution. The approximation (25.61) is obtained by replacing W_3 with a random variable having a uniform distribution over (0, 1).

If W_3 is neglected in this representation (as might be possible if q is small compared with p), the following approximation is obtained:

$$I_x(p,q) \doteqdot J_x(p,q) = \begin{cases} \dfrac{q}{p+q}\left(\dfrac{x}{1-x}\right)^p & \text{for } 0 \le x \le \dfrac{1}{2}, \\ 1 - \dfrac{p}{p+q}\left(\dfrac{1-x}{x}\right)^q & \text{for } \dfrac{1}{2} \le x \le 1. \end{cases} \quad (25.64)$$

This approximation is also very good for small values of p and q. In fact, if $p+q<1$,

$$\max_{0\le x\le 1}|I_x(p,q) - J_x(p,q)| \le \frac{p}{1+q}\left[\exp\left\{\frac{\pi^2}{24}(p+q)\right\} - 1\right]^{1+q}$$

$$\le \frac{p(p+q)}{1+q} \le p(p+q). \quad (25.65)$$

Molina (1932) obtained the following approximation to the incomplete beta function:

$$I_x(p,q) \doteqdot \sum_{j=0}^{6} \frac{A_j}{j!}\left(\frac{z}{N}\right)^{q+j} D(q+j,z), \quad (25.66)$$

where

$$N = p + \tfrac{1}{2}q - \tfrac{1}{2}, \quad z = -N\log x,$$

$$A_0 = 1, \quad A_1 = A_3 = A_5 = 0,$$

$$A_2 = \tfrac{1}{12}(q-1), \quad A_4 = \tfrac{1}{240}(q-1)(5q-7),$$

$$A_6 = \tfrac{1}{4032}(q-1)(35q^2 - 112q + 93),$$

$$D(a,b) = \int_0^1 t^{a-1}e^{-bt}\,dt = b^{-a}\Gamma_b(a),$$

where $\Gamma_b(a)$ is an incomplete gamma function, as defined in Chapter 1, Section A5. [See also appendix in Gnanadesikan, Pinkham, and Hughes (1967), where more computational details are given.]

Woods and Posten (1968), have constructed computer programs based on the Fourier expansion

$$I_x(p,q) = 1 - \theta\pi^{-1} - \sum_{j=1}^{\infty} b_j \sin(j\theta), \tag{25.67}$$

where

$$\theta = \cos^{-1}(2x-1),$$
$$b_1 = 2\pi^{-1}(p-q)(p+q)^{-1},$$
$$b_2 = \pi^{-1}(p+q)^{-1}(p+q+1)^{-1} \times \{2(p-q)^2 - (p+q)(p+q-1)\},$$
$$(j+2)b_{j+2} = (j+p+q+1)^{-1} \times \{2(p-q)(j+1)b_{j+1} + (j+1-p-q)jb_j\}.$$

They found that (for m sufficiently large) if the infinite series be terminated at the term containing b_m, the error is less than

$$\tfrac{1}{2}m|b_m|\{\min(p,q)\}^{-1} \quad \text{if } p \neq q,$$
$$\tfrac{1}{4}m|b_m|p^{-1} \quad \text{if } p = q,\ m \text{ is even}.$$

In some special cases there are simple explicit formulas for the b's:

Case 1. $p = q$:

$$b_{2j-1} = 0,$$
$$b_{2j} = (j\pi)^{-1}\prod_{i=1}^{j}\left(\frac{2j-1-2p}{2j-1+2p}\right).$$

Case 2. $q = 1/2$:

$$b_j = 2(j\pi)^{-1}\prod_{i=1}^{j}\left(\frac{2p-2j+1}{2p+2j-1}\right).$$

If both p and q have fractional part equal to one-half, then $b_j = 0$ for

$j \geq p + q$. An attractive feature of this method is that

$$\sum_{j=0}^{N} b_j \sin(j\theta)$$

can be evaluated without using trigonometric functions as

$$2u_1(x - x^2)^{1/2},$$

where

$$x = \tfrac{1}{2}(1 + \cos\theta),$$

and u_1 is obtained by application of the backward recurrence relation

$$u_j = b_j + 2(2x - 1)u_{j+1} - u_{j+2}$$

$$\text{for } j = N, N-1, \ldots, 1 \text{ with } u_{N+1} = u_{N+2} = 0.$$

Kalinin (1968) has obtained the expansion

$$\begin{aligned}
&[B(p,q)]^{-1} x^{p-1}(1-x)^{q-1} \\
&\quad = (pq)^{-1/2}(p+q)^{3/2}\phi(y)\exp\left[\sum_{j=1}^{\mu-1} W_j p^{-(1/4)j} + R_\mu\right], \qquad (25.68)
\end{aligned}$$

where

$$\begin{aligned}
y &= (p+q)(p^{-1}+q^{-1})^{1/2}\left\{x - \frac{p}{p+q}\right\}, \\
R_\mu &= O(p^{-(1/2)\mu}), \\
\phi(y) &= (\sqrt{2\pi})^{-1}\exp(-\tfrac{1}{2}y^2), \\
W_j &= \frac{y^j}{j}\left(\frac{p}{p+q}\right)^{j/2}\left[\left(\frac{p}{q}\right)^{j/2} - \left(\frac{q}{p}\right)^{j/2}\right] \\
&\quad - \frac{y^{j+2}}{j+2}\left(\frac{p}{p+q}\right)^{(j/2)+1}\left[\left(\frac{p}{q}\right)^{j/2} - \left(\frac{q}{p}\right)^{j/2+1}\right], \qquad \text{for } j \text{ odd}, \\
W_{2k} &= \frac{y^{2k}}{2k}\left(\frac{p}{p+q}\right)^{k}\left[\left(\frac{p}{q}\right)^{k} + \left(\frac{q}{p}\right)^{k}\right] \\
&\quad - \frac{y^{2k+2}}{2k+2}\left(\frac{p}{p+q}\right)^{k+1}\left[\left(\frac{p}{q}\right)^{k} + \left(\frac{q}{p}\right)^{k+1}\right] \\
&\quad + \frac{B_{k+1}}{k(k+1)}\left[\left(\frac{p}{p+q}\right)^{k} - \left(\frac{p}{q}\right)^{k} - 1\right].
\end{aligned}$$

Kalinin also gives similar (but rather more complicated) expansions for the density functions of gamma, F-, and t-distributions. Because of their complexity we have not reproduced these in the appropriate chapters.

Frankl and Maehara *(1990)* obtain the following inequalities for the tail areas of a standard beta (p,q) distribution $[\mu = E[X] = p/(p+q)]$:

$$\Pr[|X-\mu| > \varepsilon\mu] < 2a_\varepsilon^{-1}(2\pi q\mu)^{-1/2}\exp(-p\varepsilon a_\varepsilon), \qquad (25.69a)$$

where

$$a_\varepsilon = \tfrac{1}{2}\varepsilon - \tfrac{1}{3}\varepsilon^2,$$

$$\Pr[|X-\mu| > \varepsilon\mu] < 2(\varepsilon a_\varepsilon)^{-1}\log(2q). \qquad (25.69b)$$

6.2 Tables

The first edition (1934) of Pearson's tables included values of $I_x(p,q)$ to seven decimal places for

$$p,q = 0.5(0.5)11.0(1)50 \qquad \text{with } p \ge q,$$
$$x = 0.00(0.01)1.00.$$

The second edition (1968) also includes values of $I_x(p,q)$ to seven decimal places for

$$p = 11.5(1.0)14.5 \qquad \text{with } q = 0.5,$$
$$x = 0.00(0.01)1.00,$$

and to eight decimal places for $p = 0.5(0.5)11.0(1)16$; $q = 0.5$, $x = 0.988(0.0005)0.9985$, $0.9988(0.0001)0.9999$, and for $q = 1.0(0.5)3.0$, $x = 0.988(0.001)0.999$. Values to seven decimal places are also given for $x = 0.975, 0.985$.

Further values have been calculated by Osborn and Madey *(1968)*. These cover values of p, q in a region where interpolation using Pearson's tables is difficult. Values of $B_x(p,q)$ and $I_x(p,q)$ are to five significant figures for

$$p,q = 0.50(0.05)2.00,$$
$$x = 0.10(0.01)1.00.$$

The formulas used for calculation were

$$B_x(p,q) = x^p\left[\frac{1}{p} + \frac{1-q}{p+1}x + \frac{(1-q)(2-q)}{2!(p+2)}x^2 + \cdots\right] \qquad (25.70a)$$

for $0 < x \le \frac{1}{2}$, and

$$B_x(p,q) = B_{0.5}(p,q) + \frac{1-w^q}{(q)2^q} + \frac{(1-p)(1-w^{q+1})}{1!(q+1)2^{q+1}} + \frac{(1-p)(2-p)(1-w^{q+2})}{2!(q+2)2^{q+2}} + \cdots, \qquad (25.70b)$$

with $w = 2(1-x)$ for $\frac{1}{2} < x < 1$.

Percentage points of the beta distribution have been tabulated by Thompson (1941), Clark (1953), Harter (1964), and Vogler (1964). Thompson gave values of $X(P;p;q)$, where

$$I_{X(P;p;q)}(p,q) = P$$

to five significant figures for

$$p = 0.5(0.5)15.0, 20, 30, 60,$$
$$q = 0.5(0.5)5.0, 6, 7.5, 10, 12, 15, 20, 30, 60,$$
$$P = 0.50, 0.25, 0.10, 0.05, 0.025, 0.01, 0.005.$$

These tables are included in Pearson and Hartley (1954), the third edition (1966) of which contains also values for $P = 0.0025$ and 0.001, calculated by Amos (1963). Harter (1964) gives $X(P;p;q)$ to seven significant figures for $p,q = 1(1)40$; $P = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1(0.1)0.5$. Vogler (1964) gives $X(P;p;q)$, and also $B(p,q)$, to six significant figures for

$$p = 0.50(0.05)1.00, 1.1, 1.25(0.25)2.50(0.5)5.0, 6, 7.5, 10, 12, 15, 20, 30, 60,$$
$$q = 0.5(0.5)5.0, 6, 7.5, 10, 12, 15, 20, 30, 60,$$
$$P = 0.0001, 0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5.$$

Bouver and Bargmann (1975) used a continued fraction expression

$$I_x(p,q) = \{B(p,q)\}^{-1} p^{-1} x^p (1-x)^q \left\{ \frac{1}{1-}\,\frac{c_1}{1+}\,\frac{d_1}{1-}\,\frac{c_2}{1+}\,\frac{d_2}{1-} \cdots \right\}, \qquad (25.71)$$

where

$$c_j = \frac{(p+j-1)(p+q+j-1)x}{(p+2j-2)(p+2j-1)},$$

$$d_j = \frac{j(q-j)}{(p+2j-1)(p+2j)}.$$

This formula was obtained earlier by Aroian (1941, 1950), and it is referred to in Abramowitz and Stegun (1964), Boardman (1975), Tretter and Walster (1979), and Kennedy and Gentle (1980). Bouver and Bargmann (1975) recommend this formula for use when p and q are both within the (very broad) range 10^{-8} to 70,000, and especially when either or both of p and q are less than 1. [See Posten (1986) for further details.] However, Kennedy and Gentle (1980) prefer the IMSL (1977) subroutine for calculation of $I_x(p,q)$ based on Bosten and Battiste (1974).

Recurrence relations, such as Eq. (1.95) of Chapter 1, can be used in calculating $I_x(p,q)$ from available values for smaller p and/or q. Other useful relations are

$$I_x(p+1,q) = I_x(p,q) - \{pB(p,q)\}^{-1}x^p(1-x)^q, \qquad (25.72a)$$

$$I_x(p,q+1) = I_x(p,q) + \{qB(p,q)\}^{-1}x^p(1-x)^q. \qquad (25.72b)$$

Combining (25.72a) and (25.72b), we obtain

$$I_x(p+1,q+1) = I_x(p,q) + \{B(p,q)\}^{-1} \times\left\{q^{-1}x^{p+1}(1-x)^q - p^{-1}x^p(1-x)^{q+1}\right\} \qquad (25.73)$$

[Soper (1921); Gleissner (1984)]. Bosten and Battiste (1974) use (25.73), and give explicit details of application in a computer routine. [See also Lee (1989, 1992a, b).]

For the case where q is less than 1, we get by direct expansion of $(1-x)^{-(1-q)}$ and term-by-term integration

$$I_x(p,q) = \frac{1}{B(p,q)}\sum_{j=0}^{\infty}\frac{(1-q)^{[j]}}{j!}\frac{x^{p+j}}{p+j}$$

$$= \frac{x^p}{B(p,q)}\sum_{j=0}^{\infty}\left\{\prod_{i=1}^{j}(1-qi^{-1})\right\}\frac{x^j}{p+j} \qquad (25.74)$$

[with $(1-q)^{[j]} = (1-q)(2-q)\cdots(j-q)$; $(1-q)^{[0]} = 1$]. The expression

$$I_x(p,q) = \frac{x^p}{pB(p,q)}\sum_{j=0}^{\infty}\frac{(1-q^*)^{[j]}}{p+j}\frac{x^j}{j!}$$

$$+\frac{x^p(1-x)^q}{qB(p,q)}\sum_{j=1}^{[q]}\frac{q^{(j)}}{(p+q)^{(j)}}(1-x)^{-j}, \qquad (25.75)$$

where

$$q^* = \begin{cases} 1 & \text{if } q \text{ is an integer} \\ q - [q] & \text{otherwise} \end{cases}$$

and $a^{(j)} = a(a-1)\cdots(a-j+1)$ was used by Ludwig (1963) and improved for computation purposes by Bosten and Battiste (1974).

Majumder and Bhattacharjee (1973a, b) applied (25.72a) recursively, obtaining the formula

$$\begin{aligned} I_x(p+j+t, q-1) &= I_x(p,q) - \{pB(p,q)\}^{-1} x^p (1-x)^{q-j} \\ &\times \left[1 + \frac{q-1}{p+1}\frac{x}{1-x}\left[1 + \frac{q-2}{p+2}\frac{x}{1-x}\right.\right. \\ &\times \left[1 + \frac{q-j+1}{p+j-1}\frac{x}{1-x}\left[1 + \frac{q-j}{p+j}x\right.\right. \\ &\times \left.\left.\left[1 + \cdots \left[1 + \frac{p+q+t-2}{p+j+t-1}x\right]\cdots\right]\right]. \end{aligned} \qquad (25.76)$$

Lee (1992a) carried out a comparison of computing time taken in applying

Aroian's (1941, 1950) continued fraction (25.71),
Bosten and Battiste's (1974) formula (25.75),
Lee's (1989, 1992b) formula (25.73),
Majumder and Bhattacharjee's formula (25.76).

Lee found that the last of these occupied considerably less CPU time (from tests on an IBM 3090 VM/CMS system) than the other three formulas. He remarked, however, that the Bosten and Battiste formula (25.75) is more accessible in the form used in an IMSL (1985) package.

7 RELATED DISTRIBUTIONS

The distribution of $-\log X$, when X has the standard beta (p,q) distribution (25.2) has been discussed in Section 25.3. Goldfarb and Gentry (1979) [see also Barrett, Normand, and Peleg (1991)] suggest the possible use of log-beta distributions in place of lognormal distributions when fitting data which come from parent distributions which may be positively or negatively skew. As the name indicates, Y has a log-beta distribution if $\log Y$ has a beta distribution (i.e., Y is distributed as e^X, where X has a beta distribution).

Clearly the support of Y must be over a finite positive interval, $0 \le \eta_1 \le Y \le \eta_2$. Then $X = \log Y$ has a beta (p, q) distribution over the interval (η_1, η_2), and $(\log Y - \eta_1)/(\eta_2 - \eta_1) = U$ has a standard beta (p, q) distribution. The moments can be expressed as

$$\mu'_r(Y) = \frac{e^{r\eta_1}}{B(p,q)} \int_0^1 u^{p-1}(1-u)^{q-1} \exp\{r(\eta_2 - \eta_1)u\}\, du$$

$$= e^{r\eta_1} {}_1F_1(p; p+q; r(\eta_2 - \eta_1)) \tag{25.77}$$

[where ${}_1F_1(\cdot)$ is a confluent hypergeometric function; see Chapter 1, Eq. (1.125)]. Further references include Bunz et al. (1987), Chang et al. (1988), Han et al. (1989), and Runyan et al. (1988).

If X has the beta distribution (25.2), then by the transformation

$$T = \frac{X}{1-X} \tag{25.78}$$

we obtain a distribution with probability density function

$$p_T(t) = \frac{1}{B(p,q)} \left(\frac{t}{1+t}\right)^{p-1} \left(\frac{1}{1+t}\right)^{q-1} \frac{1}{(1+t)^2}$$

$$= \frac{1}{B(p,q)} \frac{t^{p-1}}{(1+t)^{p+q}}, \qquad t > 0. \tag{25.79}$$

This is a standard form of Pearson Type VI distribution, sometimes called a beta-prime distribution [Keeping (1962)].

It is also known as a beta distribution of the second kind (while the ordinary beta distributions discussed in this chapter are referred to as beta distributions *of* the first kind.) This distribution and its generalizations are discussed in Chapter 27, Section 6, dealing with the F-distribution. The relationship between the Type VI and beta distributions is exploited in Chapter 27 to express the probability integral of a central F-distribution in terms of an incomplete beta function ratio.

Little work has been done on "Weibullized" beta distributions, obtained by supposing a random variable Z to be such that (for some c) Z^c has a standard beta distribution. It is of course easy to write down the moments of such a distribution, since

$$\mu'_r(Z) = E[Z^r] = E\left[(Z^c)^{r/c}\right],$$

and so $\mu'_r(Z)$ is the (r/c)th moment of the corresponding beta distribution.

If X has a power-function distribution (Section 25.1), then X^{-1} has a Pareto distribution (Chapter 20). ***Compound beta*** distributions may be formed by ascribing distributions to some or all of the parameters p, q, a and b of distribution (25.1). However, such distributions have not been used much in applied statistical work. Continuous distributions for p and q usually present analytical difficulties, owing to the presence of the beta function $B(p,q)$ in (25.1) [or (25.2)]. As a matter of interest, we may note that if we suppose that p and q are positive integers and that for $p+q=s$ (≥ 2) fixed, p is equally likely to take values $1,2,\dots,(s-1)$, then the probability density function of X, given $p+q=s$, is

$$p_X(x|s)=\frac{(s-1)\Sigma_{p-1=0}^{s-2}\binom{s-2}{p-1}x^{p-1}(1-x)^{s-2-(p-1)}}{s-1}=1, \quad (25.80)$$

that is, the distribution is rectangular (Chapter 26).

It follows that ***whateuer the distribution of*** $(p+q)$, the compound distribution is rectangular if the conditional distribution of p, given $(p+q)$, is discrete rectangular as described above. The ***length-biased*** distribution corresponding to standard beta (p,q) is standard beta $(p-1,q)$. [Lingappaiah (1988) contains a discussion.]

Roy, Roy, and Ali (1993) have introduced the binomial mixture of beta distributions (of the first kind) with density function

$$p_X(x|n,p,a,b)=\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}\frac{x^{(a/2)+r-1}(1-x)^{(b/2)-1}}{B((a/2)+r,b/2)},$$

$0<x<1$. The kth moment about zero of X is given by

$$E[X^k]=\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}\frac{\Gamma((a/2)+r+k)\Gamma((a+b)/2+r)}{\Gamma((a+b)/2+r+k)\Gamma((a/2)+r)}$$

from which, in particular, we obtain the mean and variance of X to be

$$E[X]=\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}(a+2r)/(a+b+2r)$$

and

$$\begin{aligned}\operatorname{var}(X)=&\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}(a+2r)(a+2r+2)\\&\times\{(a+b+2r)(a+b+2r+2)\}^{-1}\\&-\left\{\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}(a+2r)/(a+b+2r)\right\}^2.\end{aligned}$$

Similarly, these authors also introduced the binomial mixture of beta distributions (of the second kind) with density function

$$p_X(x|n,p,a,b) = \sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} \frac{x^{(a/2)+r-1}}{B((a/2)+r,(b/2))(1+x)^{((a+b)/2)+r}}$$

$$0 < x < \infty.$$

The k^{th} moment about zero of X is given by

$$E[X^k] = \sum_{r=0}^{n} \binom{n}{r} p^r (1-p)^{n-r} \Gamma\left(\frac{a}{2}+r+k\right)\Gamma\left(\frac{b}{2}-k\right) \Big/ \left\{\Gamma\left(\frac{a}{2}+r\right)\Gamma\left(\frac{b}{2}\right)\right\};$$

the mean and variance are, in particular, given by

$$E[X] = (a+2np)/(b-2)$$

and

$$\operatorname{var}(X) = \frac{2a(a+b-2)}{(b-2)^2(b-4)} + \frac{4[p\{2(na+nb-2)+np(2na-b-4n)\}]}{(b-2)^2(b-4)}$$

Johnson (1949) has considered the distribution of $\log[X/(1-X)]$ when X has distribution (25.2). The moment-generating function of $\log[X/(1-X)]$ is

$$E\left[X^t(1-X)^{-t}\right] = \frac{B(p+t,q-t)}{B(p,q)} = \frac{\Gamma(p+t)\Gamma(q-t)}{\Gamma(p)\Gamma(q)} \quad \text{[cf. (25.63)]}, \tag{25.81}$$

whence the rth cumulant is

$$\kappa_r = \psi^{(r-1)}(p) + (-1)^r \psi^{(r-1)}(q). \tag{25.82}$$

Compare with (25.17)″ and (25.17)‴ in Section 25.3.

Approximating the polygamma functions we obtain (for p, q large)

$$\beta_1 \doteqdot p^{-1} + q^{-1} - 4(p+q)^{-1}, \tag{25.83}$$

$$\beta_2 \doteqdot 3 + 2p^{-1} + 2q^{-1} - 6(p+q)^{-1}.$$

These may be compared with the approximations [derived from (25.15c) and (25.15d)] for the moment ratios of X:

$$\beta_1(X) \doteqdot 4(p^{-1} + q^{-1}) - 16(p+q)^{-1}, \qquad (25.84)$$

$$\beta_2(X) \doteqdot 3 + 6(p^{-1} + q^{-1}) - 30(p+q)^{-1}.$$

Ratnaparkhi and Mosimann (1990) provide tables of β_1 and β_2 for all combinations of p, q = 0.1(0.2)0.5, 1, 3, 5(5)20 with addition of q = 25. We have already remarked (Section 25.2) that beta distributions can be generated as the distributions of ratios $X_1/(X_1 + X_2)$ where X_1, X_2 are independent random variables having chi-squared distributions.

If one or both of X_1, X_2 have ***noncentral*** χ^2 distributions (Chapter 29), the distribution of the ratio is called a ***noncentral beta*** distribution [Hodges (1955); Seber (1963)]. These distributions are evidently related to singly or doubly noncentral F-distributions (and will be discussed in Chapter 30) in the same way as beta distributions are related to central F-distributions (see the earlier part of this section).

Pham-Gia and Duong (1989) investigate the generalized three-parameter beta distribution [G3B(α_1, a,; λ)] with pdf,

$$p_Y(y|\alpha_1, \alpha_2; \lambda) = \frac{\lambda^{\alpha_1} y^{\alpha_1 - 1}(1-y)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)[1 - (1-\lambda)y]^{\alpha_1 + \alpha_2}},$$

$$0 \le y \le 1; \alpha_1, \alpha_2, \lambda > 0, \qquad (25.85)$$

which is the distribution of the ratio $X_1/(X_1 + X_2)$ where X_i $(i = 1, 2)$ are independent two-parameter gamma (α_i, β_i) random variables (and A = β_1/β_2). It reduces to the standard beta distribution when A = 1. If Y ~ G3B(α_1, a,; λ) then $(1 - Y)$ ~ G3B(α_2, a, ; λ^{-1}), which is a property similar to the one enjoyed by standard beta distributions. Libby and Novick (1982) studied these distributions in a multivariate setting and used them for fitting utility functions. Chen and Novick (1984) used them as priors for binomial sampling models.

The presence of the parameter A allows G3B to take a wider variety of shapes than the standard beta distributions. For example, G3B(α, a ; A) can be positively or negatively skewed depending on the value of A. In general, for 0 < A < 1, the pdf of G3B(α_1, a,; λ) is below that of the corresponding standard beta near zero but crosses the latter to become the greater at $y_0 = \{1 - \lambda^{\alpha_1/(\alpha_1 + \alpha_2)}\}^{-1} - (1-\lambda)^{-1}$. For A > 1 the reverse is true, with the same crossing point.

Figure 25.3 shows density functions of G3B for some selected values of a,, a,, and A. It can be seen that G3B(α_1, α_2; A) and G3B($\alpha_2, \alpha_1; \lambda^{-1}$) are symmetrical about y = 0.50. For a, = a, = $\frac{1}{2}$ and A = 2.5, the G3B is U-shaped with **antimode** at y,, = $\frac{2}{3}$ [y,, = $\frac{1}{2}$ for **beta**($\frac{1}{2}, \frac{1}{2}$)]. Also **G3B(1, 1**; 2.5)

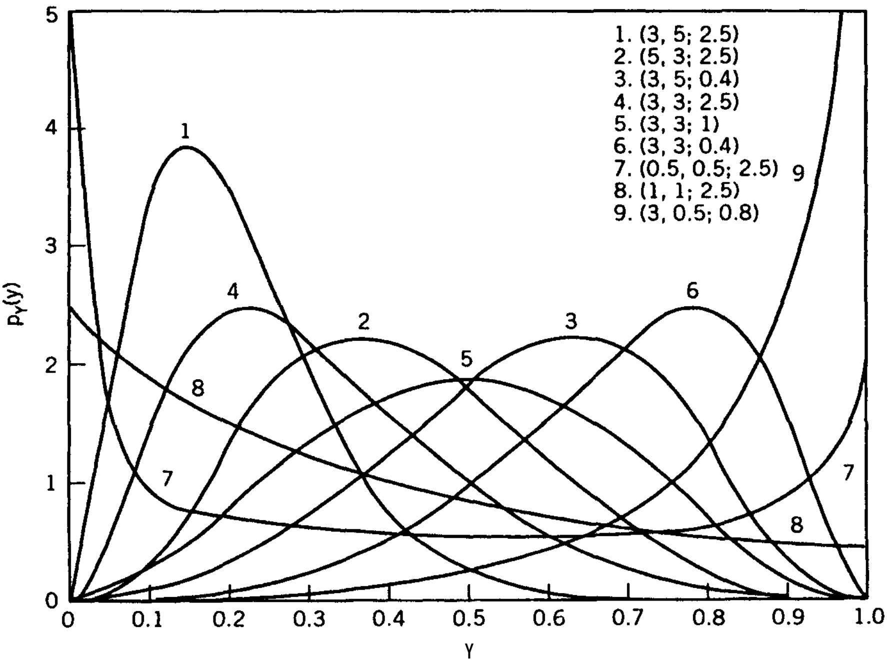


Figure 25.3 Densities of generalized beta distributions $p_Y(y|\alpha_1, \alpha_2: A)$. (cf. Figs. 25.1a, b)

is strictly decreasing while beta $(1, 1)$ is the uniform distribution on $(0, 1)$. Finally G3B(3, 0.5; 0.8) is J-shaped like beta(3, 0.5) but crosses the latter from below near $y_0 = 0.8704$.

Volodin (1994) has considered some generalized forms of the beta distribution. The generalized beta random variable X has its cumulative distribution function to be

$$F_X(x) = \frac{1}{B(\alpha, \beta + \gamma)} \int_0^x z^{\alpha-1}(x - z)^{\beta}(1 - z)^{\gamma-1}\, dz$$

and its probability density function as

$$p_X(x) = \frac{\beta}{B(\alpha, \beta + y)} \int_0^x z^{\alpha-1}(x - z)^{\beta-1}(1 - z)^{\gamma-1}\, \mathrm{dz}.$$

The k^{th} moment of X (about 0) is

$$E[X^k] = \frac{\beta}{B(\alpha, \beta + \gamma)} \sum_{i=0}^{k} \binom{k}{i} B(\alpha + k - i, \beta + \gamma + i)/(\beta + i).$$

If we define $Y = 1 - X$, then we have

$$F_Y(y) = 1 - \frac{1}{B(\alpha+\beta,\gamma)} \int_Y^1 z^{\gamma-1}(z-y)^{\beta}(1-z)^{\alpha-1}\,dz,$$

$$p_Y(y) = \frac{\beta}{B(\alpha+\beta,\gamma)} \int_y^1 z^{\gamma-1}(z-y)^{\beta-1}(1-z)^{\alpha-1}\,dz,$$

and

$$E[Y^k] = \frac{\beta}{B(\alpha+\beta,\gamma)} B(\alpha+\beta+k,\gamma)B(k+1,\beta).$$

Note that in the particular case when $\beta = 0$, the above distribution becomes a beta distribution.

The following generalization of the beta distribution was suggested by Armero and Bayarri (1994) in connection with marginal prior/posterior distribution for parameter p $(0 < p < 1)$ representing traffic intensity in a $M/M/1$ queue (or, equivalently, the parameter of a geometric distribution $\Pr[N = n|\rho] = (1-\rho)\rho^n$, $n = 0, 1, 2, \dots$). A random variable X has a Gauss ***hypergeometric distribution*** with parameters a, β, γ and z $(\alpha > 0, \beta > 0)$ if it has a continuous distribution with probability density function

$$p_X(x|\alpha,\beta,\gamma,z) = Cx^{\alpha-1}(1-x)^{\beta-1}/(1+zx)^{\gamma}, 0 \le x \le 1,$$

where the normalizing constant C is given by

$$C^{-1} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} F(\gamma,\alpha;\alpha+\beta;-z).$$

Here, F represents the Gauss hypergeometric function (see Chapter 1). The k^{th} moment (about zero) of X is

$$E[X^k] = \frac{B(k+\alpha,\beta)}{B(\alpha,\beta)} \cdot \frac{F(\gamma,\alpha+k;\alpha+\beta+k;-z)}{F(\gamma,\alpha;\alpha+\beta;-z)}.$$

The above given Gauss hypergeometric distribution becomes the beta(α,β) distribution when either γ or z equals zero. Some plots of the Gauss hypergeometric density are shown in Fig. 25.4.

The arc-sine distribution [Eq. (25.8)] has been studied by Norton (1975, 1978), Shantaram (1981), and Arnold and Groeneveld (1980), among others.

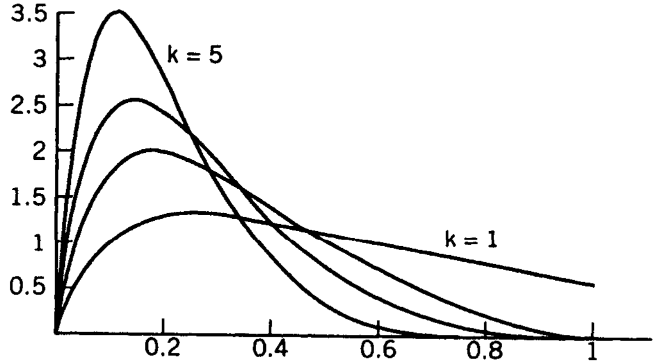


Figure 25.4 Gauss hypergeometric density $p_X(x|2, k, 3, 2)$ for $k = 1, 2, 3, 5$

The central moments of the standard arc-sine density,

$$p_U(u) = \frac{(1-u^2)^{-1/2}}{\pi}, \qquad -1 \le u \le 1,$$

are

$$\mu_{2j} = \binom{2j}{j}\left(\frac{1}{2}\right)^{2j}, \qquad j = 1, 2, 3, \ldots . \tag{25.86}$$

Also for the distribution with the pdf,

$$p_U(u|b) = \begin{cases} \dfrac{1}{\pi\sqrt{\left(\dfrac{2}{b}\right)^2 - x^2}}, & |x| < \left|\dfrac{2}{b}\right| \\ 0, & |x| \ge \left|\dfrac{2}{b}\right| \end{cases}, \quad b \ne 0, \tag{25.87}$$

we have [Norton (1975)]

$$\mu_{2j} = b^{-2j}\binom{2j}{j},$$

and these moments characterize the distribution. If U and V are i.i.d. uniform $(-\pi, \pi)$ random variables, then $\sin U$, $\sin 2U$, $-\cos 2U$, $\sin(U+V)$, and $\sin(U-V)$ all have arc-sine distributions.

The following characterizations of the arc-sine distribution are based on the fact that the moments characterize this distribution (because its support is finite):

1. If X is a symmetric random variable, the random variables X^2 and $(1 + X)/2$ are identically distributed if and only if X has the arc-sine distribution (25.8).
2. If X is a symmetric random variable and X^2 and $1 - X^2$ are identically distributed, then X and $2X\sqrt{1 - X^2}$ are identically distributed if and only if X has the arc-sine distribution (25.8).
3. If X, and X_2 are symmetric i.i.d. random variables and X_i^2 and $1 - X_i^2$ are identically distributed, then $X_1^2 - X_2^2$ and X_1X_2 are identically distributed if and only if X_i have the arc-sine distribution (25.8).

Characterization (1) can be expressed in terms of the fraction of time, W, spent by a symmetric random walk on the positive side of zero as follows: From (25.8), X − 2W − 1 has density (25.1) Seller (1966) noted that W is more likely to be near 0 or 1 than to be near $\frac{1}{2}$. Indeed, since X^2 and $W = (1 + X)/2$ are identically distributed, it follows that $4(W - \frac{1}{2})^2$ and W are identically distributed. Thus we have for $t \in (0, 1)$,

$$\Pr[W \leq 1 - t] = \Pr\left[\left|W - \tfrac{1}{2}\right| \leq \tfrac{1}{2}\sqrt{t}\right]. \tag{25.88}$$

For example, Pr[*W* 5 0.011 = Pr[*W* ≥ 0.991 = Pr[0.45 ≤ W I 0.551. Shantaram (1981) proved that if X and Y are i.i.d. random variables, then X + Y and *XY* are identically distributed if and only if X has an arc-sine distribution or is one of a certain class of discrete random variables, consisting of two degenerate random variables and for each K ≥ 2 a unique random variable whose support consists of K mass points. A similar result was obtained by Norton (1978).

Patel and Khatri (1978) studied the Lagrangian-beta distribution with the cdf,

$$F_T(t|n, a, \beta, r) = 1 - \sum_{j=0}^{r-1} \frac{\;}{n + \beta j}\binom{n + \beta j}{j}(a\ t) \sim (\alpha t)^{n+\beta j-j} \tag{25.89}$$

and the pdf,

$$p_T(t|n, \alpha, \beta, r) = \sum_{j=0}^{r-1} n(n + \beta j)^{-1}\binom{n + \beta j}{j}\alpha(\alpha t)^j(1 - \alpha t)^{n+\beta j-j-1},$$

$$0 \leq \alpha t \leq 1. \tag{25.90}$$

For $\beta = 1$ we have

$$p_T(t|n, \mathrm{a}, 1, r) = \frac{\mathrm{a}}{B(n,r)}(\alpha t)^{r-1}(1-\alpha t)^{n-1}. \tag{25.91}$$

The motivation is as follows. The time between occurrences of a classical negative binomial process (see Chapter 5) follows a beta distribution. Suppose that the process is controlled by the generalized negative binomial (GNB) distribution

$$\Pr[X = x|n, \mathrm{a}, \mathrm{t}] = \frac{\mathrm{n}}{x!}\frac{\Gamma(n+\beta x)}{\Gamma(n+\beta x - x + 1)}(\alpha t)^x(1-\alpha t)^{n+\beta x - x},$$

$$x = 0, 1, \ldots . \tag{25.92}$$

[Jain and Consul (1971), Chapter 6, Section 1] where $t \geq 0$ is the time between occurrences of the GNB process, $0 \leq \mathrm{a\,t} \leq 1$. Then the cdf of the random variable T, which is the time until the rth occurrence of the GNB process, is given by (25.89).

Patel and Khatri (1978) calculated the moments of this distribution. In particular,

$$E[T] = \frac{n}{\alpha}\sum_{x=0}^{r-1}\frac{1}{(n+\beta x)(n+\beta x+1)} \tag{25.93a}$$

and

$$E[T^2] = \frac{2}{n\alpha}E[T] + \frac{2}{\alpha^2}\sum_{x=0}^{r-1}\frac{x(n-\beta)-2}{(n+\beta x)(n+\beta x+1)(n+\beta x+2)}. \tag{25.93b}$$

($E[T]$ decreases as β increases.) They also provide graphical representation of the distribution.

Recently, Fosam and Sapatinas (1994) have presented some regression type characterization results for the Pareto and power-function distributions, basing them on beta random variables, thus relating these distributions.

8 PRODUCTS, QUOTIENTS, AND DIFFERENCES OF INDEPENDENT BETA VARIABLES

Early results on the distribution of products of independent beta variables were obtained by Kotlarski (1962). Since then a small 'cottage industry' has developed in statistical literature on these topics. A few examples are presented here.

Since the product of k independent random variables with standard beta (p_i, q_i) distributions $(i = 1, 2, \ldots, k)$ has finite range of variation (from 0 to 1), its distribution is determined by its moments. With $k = 2$, $Y = X_1 X_2$ has a beta distribution if and only if

$$\mu'_r = \frac{\Gamma(p_1 + r)\Gamma(p_1 + q_1)}{\Gamma(p_1)\Gamma(p_1 + q_1 + r)} \cdot \frac{\Gamma(p_2 + r)\Gamma(p_2 + q_2)}{\Gamma(p_2)\Gamma(p_2 + q_2 + r)}, \qquad r = 1, 2, \ldots, \tag{25.94a}$$

can be expressed in the form

$$\mu'_r = \frac{\Gamma(p + r)\Gamma(p + q)}{\Gamma(p)\Gamma(p + q + r)} \qquad \text{for all } r \tag{25.94b}$$

For the requirement of cancelling of terms in the numerator and denominator of (25.94a) means that either $p_1 = p_2 + q_2$ or $p_2 = p_1 + q_1$. In the first case we have $p = p_2$, $q = q_1 + q_2$ in (24.94b); in the second case $p = p_1$, $q = q_1 + q_2$. Generally the rth moment of $Y = \prod_{i=1}^k X_i$ is

$$\mu'_r(Y) = \prod_{i=1}^{k} \left\{ \frac{\Gamma(p_i + r)\Gamma(p_i + q_i)}{\Gamma(p_i)\Gamma(p_i + q_i + r)} \right\}, \qquad r = 1, 2, \ldots, \tag{25.95a}$$

and Y has a standard beta (p, q) distribution if and only if (25.94b) is satisfied. This condition is satisfied if

$$p_i = p + \sum_{j=1}^{i-1} q_j \qquad \text{with } \sum_{j=1}^{0} q_j = 0.$$

Then

$$\begin{aligned} \mu'_r(Y) = {} & \left\{ \frac{\Gamma(p + r)\Gamma(p + q_1)}{\Gamma(p)\Gamma(p + q_1 + r)} \right\} \cdot \left\{ \frac{\Gamma(p + q_1 + r)\Gamma(p + q_1 + q_2)}{\Gamma(p + q_1)\Gamma(p + q_1 + q_2 + r)} \right\} \cdots \\ & \times \left\{ \frac{\Gamma(p + q_1 + \cdots + q_{k-1} + r)\Gamma(p + q_1 + \cdots + q_k)}{\Gamma(p + q_1 + \cdots + q_{k-1})\Gamma(p + q_1 + \cdots + q_k + r)} \right\} \\ = {} & \frac{\Gamma(p + r)\Gamma(p + q_1 + \cdots + q_k)}{\Gamma(p)\Gamma(p + q_1 + \cdots + q_k + r)} \end{aligned} \tag{25.95b}$$

so that Y has a standard beta $(p, \sum_{i=1}^k q_i)$ distribution. [See, e.g., Fan (1991) for a different analysis leading to this result.]

Using the Mellin transform [Chapter 1, Eq. (1.166)], Steece (1976) showed that the pdf of $Y = X_1X_2$ is

$$p_Y(y) = \frac{\Gamma(p_1+q_1)\Gamma(p_2+q_2)}{\Gamma(p_1)\Gamma(p_2)\Gamma(q_1+q_2)} y^{p_1-1}(1-y)^{q_1+q_2-1}$$

$$\times {}_2F_1(q_2, p_1-p_2+q_1; q_1+q_2; 1-y)$$

$$= \frac{\Gamma(p_1+q_1)\Gamma(p_2+q_2)}{\Gamma(p_1)\Gamma(p_2)\Gamma(q_2)\Gamma(p_1-p_2+q_1)}$$

$$\times \sum_{n=0}^{\infty} \frac{\Gamma(q_2+n)\Gamma(p_1-p_2+q_1+n)}{n!\Gamma(q_1+q_2+n)}$$

$$\times y^{p_1-1}(1-y)^{q_1+q_2+n-1}, \tag{25.96a}$$

where ${}_2F_1(\cdot)$ is the (Gauss) hypergeometric function (defined in Chapter 1, Section A6). The cdf of Y is

$$F_Y(y) = \frac{\Gamma(p_1+q_1)\Gamma(p_2+q_2)}{\Gamma(p_1)\Gamma(p_2)\Gamma(p_1-p_2+q_1)} \sum_{n=0}^{\infty} \frac{\Gamma(q_2+n)\Gamma(p_1-p_2+q_1+n)}{n!\Gamma(p_1+q_1+q_2+n)}$$

$$\times I_y(p_1, q_1+q_2+n), \tag{25.96b}$$

where $I_y(\cdot)$ denotes the incomplete beta function ratio [defined in Chapter 1, Eq. (1.91)]. [Note that different formulas are obtained by interchanging (p_1, q_1) with (p_2, q_2). See also Fan (1991).]

Exact formulas for the distributions of products and quotients of independent standard beta (p_i, q_i) random variables $X_1, \dots, X_k$ can be derived straightforwardly (though somewhat tediously) provided at least one parameter of each variable is an integer. Dennis (1994) has recently derived closed form expressions (involving infinite series) for the cumulative distribution function and probability density function of the product of k independent beta random variables when p_i and q_i $(i = 1, 2, \dots, k)$ are real-valued parameters. Through these expressions, he has also illustrated the computation of the distribution of the product of three independent beta variables.

The joint pdf of X_1 and X_2 is

$$p_{X_1,X_2}(x_1, x_2) = \{B(p_1, q_1)B(p_2, q_2)\}^{-1}$$

$$\times x_1^{p_1-1}x_2^{p_2-1}(1-x_1)^{q_1-1}(1-x_2)^{q_2-1},$$

$$0 < x_1, x_2 < 1. \tag{25.97}$$

If q_1 and q_2 are integers, (25.97) can be rewritten as

$$p_{X_1, X_2}(x_1, x_2) = \{B(p_1, q_1)B(p_2, q_2)\}^{-1} \times \sum_{t_1=0}^{q_1-1} \sum_{t_2=0}^{q_2-1} (-1)^{t_1+t_2} \binom{q_1-1}{t_1} \binom{q_2-1}{t_2} x_1^{p_1+t_1-1} x_2^{p_2+t_2-1}.$$

To evaluate the cdfs of $V = X_1X_2$ and $W = X_2/X_1$, we need to integrate functions of form $x_1^{c_1}x_2^{c_2}$ over regions $x_1x_2 \le v$, $x_2/x_1 \le w$, respectively. We have

$$J_v(c_1, c_2) = \iint_{x_1x_2 \le v} x_1^{c_1}x_2^{c_2}\, dx_2\, dx_1 = \begin{cases} \dfrac{(c_1+1)v^{c_2+1} - (c_2+1)v^{c_1+1}}{(c_1+1)(c_2+1)(c_1-c_2)}, & c_1 \ne c_2, \\ \dfrac{v^{c+1}}{(c+1)^2}\{1 - (c+1)\log v\}, & c_1 = c_2 = c, \end{cases} \tag{25.98a}$$

and

$$J_w^*(c_1, c_2) = \iint_{x_2/x_1 \le w} x_1^{c_1}x_2^{c_2}\, dx_2\, dx_1 = \begin{cases} \dfrac{w^{c_2+1}}{(c_2+1)(c_1+c_2+1)}, & w < 1, \\ \dfrac{1}{(c_1+1)(c_2+1)} - \dfrac{w^{-c_1-1}}{(c_1+1)(c_1+c_2+1)}, & w > 1. \end{cases} \tag{25.98b}$$

Then

$$F_V(v) = \frac{1}{B(p_1, q_1)B(p_2, q_2)} \sum_{t_1=0}^{q_1-1} \sum_{t_2=0}^{q_2-1} (-1)^{t_1+t_2} \binom{q_1-1}{t_1}\binom{q_2-1}{t_2} \times J_v(p_1+t_1-1, p_2+t_2-1) \tag{25.99a}$$

and

$$F_W(w) = \frac{1}{B(p_1, q_1)B(p_2, q_2)} \sum_{t_1=0}^{q_1-1} \sum_{t_2=0}^{q_2-1} (-1)^{t_1+t_2} \binom{q_1-1}{t_1}\binom{q_2-1}{t_2} \times J_w^*(p_1+t_1-1, p_2+t_2-1). \tag{25.99b}$$

Note that

$$\frac{dJ_t(c_1, c_2)}{dv} = \begin{cases} \dfrac{v^{c_2} - v^{c_1}}{c_1 - c_2} & \text{if } c_1 \neq c_2, \\ -\dfrac{v}{c+1}\log v & \text{if } c_1 = c_2 = c \end{cases} \qquad (25.100a)$$

and that

$$\frac{dJ_w^*(c_1, c_2)}{dw} = \begin{cases} \dfrac{w^{c_2}}{c_1 + c_2 + 2} & \text{if } w < 1, \\ \dfrac{w^{-c_1-2}}{c_1 + c_2 + 2} & \text{if } w > 1. \end{cases} \qquad (25.100b)$$

For the case where p_i and q_i are both integers, for all $i = 1, \dots, k$, the book by Springer *(1978)* presents the following formula for the pdf of $Y = \prod_{i=1}^{k} X_i$:

$$p_Y(y) = \sum_{i=1}^{m} \sum_{j=0}^{g_i - 1} \frac{K_{ij}}{(g_i - 1 - j)!j!} y^{d_i - 1}(-\log y)^{g_i - j - 1}, \quad (25.101)$$

where d_i is the number of different integers occurring with multiplicity g_i in the collection of integers

$$p_h - 1, p_h, p_h + 1, \dots, p_h + q_h - 2 \qquad \text{for } h = 1, 2, \dots, k,$$

and m is the number of different multiplicities. (Note that there are q_h members of the hth set, so $\sum_{i=1}^{m} g_i d_i = \sum_{h=1}^{k} q_h$.) The constants K_{ij} are determined by the recurrence relation

$$K_{ij} = \sum_{s=0}^{j-1} \sum_{t=1}^{m} (-1)^{s+1} \frac{\binom{j-1}{s} s! g_t}{(d_t - d_i)^{s+1}} K_{i,j-1-s}, \qquad j > 0, \quad (25.102)$$

with

$$K_{i0} = \sum_{\substack{t=1 \\ t \neq i}}^{m} (d_t - d_i)^{-g_t}.$$

For example, if we have $k = 3$ and $p_1 = 9$, $q_1 = 3$; $p_2 = 8$, $q_2 = 3$; $p_2 = 4$, $q_3 = 2$, the collection of integers is

$$8, 9, 10; 7, 8, 9; 3, 4.$$

The integers have the frequencies

$$g_1 = 1 \quad \text{for } 3, 4, 7, \text{ and } 10,$$

$$g_2 = 2 \quad \text{for } 8 \text{ and } 9.$$

Hence $m = 2$, $d_1 = 4$, $d_2 = 2$,

$$K_{10} = (2-4)^{-2} = \tfrac{1}{4}, \quad K_{20} = (4-2)^{-1} = \tfrac{1}{2},$$

$$K_{21} = -\frac{\binom{0}{0} 0!1}{(4-2)} K_{20} = -\frac{1}{4}.$$

The pdf is

$$p_Y(y) = \frac{3960}{7} y^3 - 1980 y^4 + 99{,}000 y^7 + (374{,}220 + 356{,}400 \log y) y^8$$

$$-(443{,}520 - 237{,}600 \log y) y^9 - \frac{198{,}000}{7} y^{10}. \qquad (25.103a)$$

It is shown in Figure 25.5.

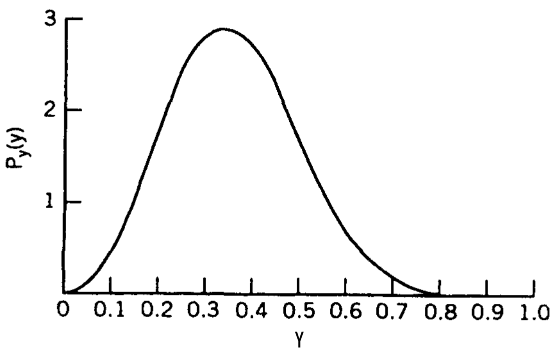


Figure 25.5 Probability density function of product of three beta variables

Pederzoli (1985), using a factorial expansion of ratios of gamma functions, obtained

$$p_Y(y) = \left\{\prod_{i=1}^{k} B(p_i, q_i)\right\}^{-1} \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \prod_{j=1}^{k} \frac{(1-q_i)^{[r_i]}}{r_i!}$$
$$\times \left\{\sum_{h=1}^{k} a_h y^{p_h+r_h-1}\right\}, \qquad 0<y<1, \qquad (25.103b)$$

where

$$a_h = \prod_{j=1}^{k} (p_i - p_j + q_i - q_h)^{-1}$$

and $a^{[m]} = a(a+1)\cdots(a+m-1)$ (provided that $p_j + s_i \neq p_i + s_j$ for all $i \neq j$, where the s_i's are nonnegative integers). The p_i's and q_i's need not be integers but must be positive.

If all of the q_i's are positive integers, (25.103b) becomes

$$p_Y(y) = \left\{\prod_{i=1}^{k} B(p_i, q_i)\right\}^{-1} \sum_{r_1=0}^{q_1-1} \cdots \sum_{r_k=0}^{q_k-1} (-1)^{r_1+\cdots+r_k}$$
$$\times \left\{\sum_{h=1}^{k} a_h y^{p_h+r_h-1}\right\}. \qquad (25.103c)$$

Tang and Gupta (1984) obtained

$$p_Y(y) = \prod_{j=1}^{k} \left\{\frac{\Gamma(p_j+q_j)}{\Gamma(q_j)}\right\} y^{p_k-1}(1-y)^{f(k)-1} \sum_{r=0}^{\infty} a_{r,k}(1-y)^r, \qquad (25.103d)$$

where $f(k) = \sum_{j=1}^{k} q_j$ and the $a_{r,k}$'s satisfy the recursive relations

$$a_{r,k} = \frac{\Gamma(f(k-1)+r)}{\Gamma(f(k)+r)} \sum_{s=0}^{r} (p_k + q_k - p_{k-1})^{[s]} a_{r-s,k-1},$$

with $a_{0,1} = 1/\Gamma(q_1)$. For $k = 2$, we obtain Steece's formula (25.96a).

Fan (1991) has provided an approximation to the distribution of Y in which Y has an approximate standard beta (p, q) distribution with

$$p = S(T - S^2)^{-1},$$
$$q = (1-S)(S-T)(T-S^2)^{-1},$$

where

$$S = \prod_{i=1} \left(\frac{p_i}{p_i + q_i} \right) \quad \text{and} \quad T = \prod_{i=1}^{k} \left\{ \frac{p_i(p_i + 1)}{(p_i + q_i)(p_i + q_i + 1)} \right\}.$$

This approach ensures that the approximation has the correct expected value and variance. He compared the first ten (!) moments of the approximate with those of the exact distribution over a considerable range, with quite good results. For example, with $k = 3$, $(p_1, p_2, p_3) = (778, 43, 23)$ and $(q_1, q_2, q_3) = (567, 57, 12)$, the approximate eighth moment is 0.10554×10^{-5}, while the exact value is 0.103925×10^{-5}.

Pham-Gia and Turkkan *(1993)* have derived the distribution of the difference $D = X_1 - X_2$. The pdf of D is

$$p_D(d) = \begin{cases} B(p_2, q_1) d^{q_1+q_2-1}(1-d)^{p_2+q_1-1} \dfrac{F_1(q_1, p_1+p_2+q_1+q_2-2, 1-p_1; p_2+q_1; 1-d, 1-d^2)}{A} & \text{for } 0 \le d \le 1, \\ B(p_1, q_2)(-d)^{q_1+q_2-1}(1+d)^{p_1+q_2-1} \dfrac{F_1(q_2, 1-p_2, p_1+p_2+q_1+q_2- \; p_1+q_2; 1-d^2, 1+}{A} & \text{for } -1 \le d \le 0, \end{cases} \tag{25.103e}$$

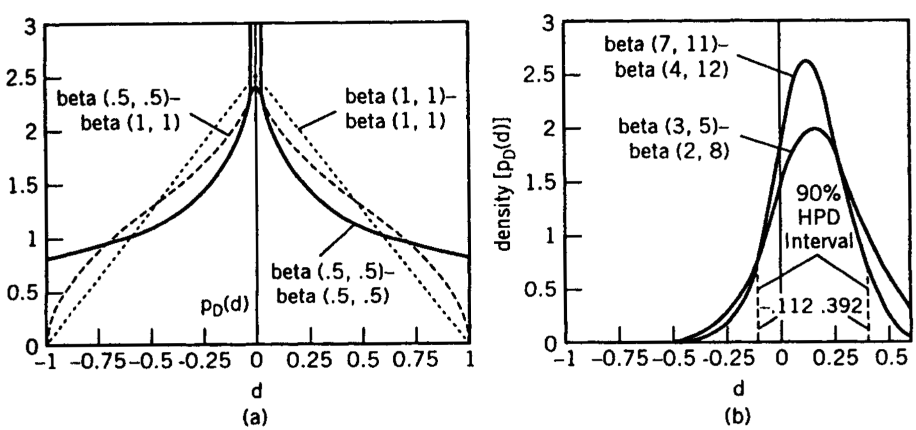


Figure 25.6 Probability density functions of differences between independent beta variables

where $A = B(p_1, q_1)B(p_2, q_2)$ and

$$F_1(a, b_1, b_2; c; x_1, x_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{a^{[i+j]}}{c^{[i+j]}} \frac{b_1^{[i]} b_2^{[j]} x_1^i x_2^j}{i! j!}$$

is the first Appell hypergeometric function in two variables; here, $a^{[b]} = a(a+1)\cdots(a+b-1)$. ($F_1$ is convergent for $|x_1| < 1$ and $|x_2| < 1$.) The distributions can have a great variety of shapes depending on the values of (p_i, q_i), $i = 1, 2$. Some cases are presented in Figures 25.6*a*, b [Pham-Gia and Turkkan (1993)].

Recently, Pham and Turkkan (1994) derived the distribution of sum of two independent beta random variables. Let X_1 and X_2 be independently distributed as beta with parameters (p_1, q_1) and (p_2, q_2), respectively. Then, they have shown that the density function of $S = X_1 + X_2$ is as follows:

$$p_S(s) = \begin{cases} B^*(p_1, q_1; p_2, q_2) s^{p_1+p_2-1}(1-s)^{q_1-1} F_1\left(p_2, 1-q_1, 1-q_2; p_1+p_2; \frac{s}{s-1}, s\right), \\ \qquad 0 \le s < 1, \\ B(p_1+q_2-1, p_2+q_1-1)/\{B(p_1, q_1)B(p_2, q_2)\}, \\ \qquad s = 1, \\ B^*(q_2, p_2; q_1, p_1)(s-1)^{p_2-1}(2-s)^{q_1+q_2-1} F_1\left(q_1, 1-p_1, 1-p_2; q_1+q_2; 2-s, \frac{2-s}{1-s}\right), \\ \qquad 1 < s \le 2, \end{cases} \tag{25.104}$$

where $B^*(p_1, q_1; p_2, q_2) = \Gamma(p_1+q_1)\Gamma(p_2+q_2)/\{\Gamma(q_1)\Gamma(q_2)\Gamma(p_1+p_2)\}$. Compare this density with that of the difference D presented in Eq. (25.103e).

REFERENCES

AbouRizk, S. M., and Halpin, D. W. (1992). Statistical properties of construction duration data, *Journal of Construction Engineering and Management*, **118**, 525–544.

AbouRizk, S. M., Halpin, D. W., and Wilson, J. R. (1991). Visual iterative fitting of beta distribution, *Journal of Construction Engineering and Management*, **117**, 589–605.

AbouRizk, S. M., Halpin, D. W., and Wilson, J. R. (1993). Fitting beta distributions based on sample data, *Unpublished manuscript*.

Abramowitz, M., and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, National Bureau of Standards, Applied Mathematics Series, Washington, DC: GPO.

Agrawal, A. R., and Yang, T.-T. (1991). Numerical simulations of low-BTU coal gas combustion in a gas turbine combustor, *International Gas Turbine and Aeroengine Congress*, Orlando, FL. (Publ. *American Society of Mechanical Engineers*, New York, NY).

Ahrens, J. H., and Dieter, U. (1974). Computer methods for sampling from Gamma, beta, Poisson and binomial distributions, ***Computing (Vienna)***, **12,** 223–246.

Alfero, D., and Dinges, H. (1984). A normal approximation for beta and gamma tail probabilities, ***Zeitschrift fur*** Wahrseheinlichkeitstheorie ***und ihre Verwandte Gebiete***, **65,** 399–420.

Amos, D. E. (1963). Additional percentage points for the incomplete beta distribution, ***Biometrika,*** **50,** 449–457.

Andersen, J. B., Lauritzen, S. L., and Thommesen, C. (1990). Distributions of phase derivatives in mobile communications, ***Proceedings of IEEE***, **137,** 197–201.

Armero, C., and Bayarri, M. J. (1994). Prior assessments for prediction in queues, ***The Statistician,*** **43,** 139–153.

Arnold, B. C., and Groeneveld, R. A. (1980). Some properties of the arcsine distribution, ***Journal of the American Statistical Association,*** **75,** 173–175.

Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices, ***Journal of Mathematical Analysis and Applications,*** **20,** 262–268.

Aroian, L. A. (1941). Continued fractions for the incomplete beta function, ***Annals of Mathematical Statistics,*** **12,** 218–223. (Correction: **30,** 1265).

Aroian, L. A. (1950). On the levels of significance of the incomplete beta function and the F-distributions, ***Biometrika,*** **37,** 219–223.

Atkinson, A. C., and Pearce, M. C. (1976). The computer generation of beta, gamma and normal random variables, ***Journal of the Royal Statistical Society, Series A,*** **139,** 431–461.

Bailey, R. W. (1992). Distributional identities of beta and chi-squared variables: A geometrical interpretation, ***The American Statistician,*** **46,** 117–120. (Correction: **47,** 86).

Bancroft, T. A. (1945). Note on an identity in the incomplete beta function, ***Annals of Mathematical Statistics,*** **16,** 98–99.

Bancroft, T. A. (1949). Some recurrence formulae in the incomplete beta function ratio, ***Annals of Mathematical Statistics,*** **20,** 451–455.

Bánkövi, G. (1964). A note on the generation of beta distributed and gamma distributed random variables, ***Publications of the Mathematical Institute, Hungarian Academy of Sciences, Series A,*** **9,** 555–562.

Bargmann, R. E., and Ghosh, S. P. (1963). Statistical distribution programs for a computer language, ***IBM Watson Research Center Report RC-1094.***

Barnard, G. A. (1967). [Discussion of Horsnell (1967)]. ***Journal of the Royal Statistical Society, Series A,*** **120,** 192–193.

Barrett, A. M., Normand, M. D., and Peleg, M. (1991). "Log beta" vs. the log-normal distribution for particle populations with a wide finite size range, ***Powder Technology, 66,*** 195–199.

Beckman, R., and Tietjen G. (1978). Maximum likelihood estimation for the beta distribution, ***Journal of Statistical Computation and Simulation,*** **7,** 253–258.

Békéssy, A. (1964). Remarks on beta distributed random numbers, ***Publications of the Mathematical Institute, Hungarian Academy of Sciences, Series A,*** **9,** 565–571.

Benedetti, C. (1956). Sulla rappresentabilità di una distribuzione binomiale mediante una distribuzione B e vice versa, ***Metron*, 18,** 121–131. (In Italian)

Boardman, T. J., and Kopitzke, R. W. (1975). Probability and table values for statistical distributions, ***Proceedings of the Statistical Computing Section, American Statistical Association*,** 81–86.

Bol'shev, L. N. (1964). Some applications of Pearson transformations, ***Reciew of the International Statistical Institute*, 32,** 14–15.

Boss, J. (1992a, b). Mixing time of grain materials. I. Theoretical considerations, ***Chemical Engineering Science*, 47,** 4027–4032; II. Experimental part, ***Ibid.*, 47,** 4033–4038.

Bosten, N. E., and Battiste, E. L. (1974). Remark on Algorithm 179 (S14): Incomplete beta ratio, ***Communications of the Association for Computing Machinery*, 17,** 156–157.

Bouver, H., and Bargmann, R. E. (1975). Computer algorithm and modules for the evaluation of statistical distribution functions, ***Technical Report* 110,** Statistics Department, University of Georgia, Athens, GA.

Bowman, K., and Shenton, L. (1979a). Approximate percentage points for Pearson distributions, ***Biometrika*,** 66, 147–151.

Bowman, K., and Shenton, L. (1979b). Further approximate Pearson percentage points and Cornish-Fisher, ***Communications in Statistics*, B8,** 231–244.

Brkich, D. M. (1987). B-test of homogeneity, ***Microelectronics and Reliability*, 27,** 639–641.

Bunz, H., Schoeck, W., Koyro, M., Gentry, J., Plunkett, S., Runyan, M., Pearson, C., and Wang, C. (1987). Particle deposition in expanding and oscillating gas bubbles, ***Journal of Aerosol Science*, 18,** 677–670.

Cadwell, J. H. (1952). An approximation to the symmetrical incomplete beta function, ***Biometrika*, 39,** 204–207.

Carnahan, J. V. (1989). Maximum likelihood estimation for the 4-parameter beta distribution, ***Communications in Statistics—Simulation and Computation*, 18,** 513–516.

Chaloner, K. M., and Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation, ***The Statistician*, 32,** 174–180.

Champernowne, D. G. (1953). The economics of sequential sampling procedures for defectives, ***Applied Statistics*,** 2, 118–130.

Chang, Y., Han, R.-L., Pearson, C. L., Runyan, M. R., Ranade, M. B., and Gentry, J. W. (1988). Applications of the log beta distribution to the evolution of aerosol growth, ***Journal of Aerosol Science*, 19,** 879–882.

Chen, J. C., and Novick, M. R. (1984). Bayesian analysis for binomial models with generalized beta prior distributions, ***Journal of Educational Statistics* 9,** 163–175.

Chen, R., Goodman, R., and Zame, A. (1984). On the limiting distribution of two random sequences, ***Journal of Multivariate Analysis*, 14,** 221–230.

Chen, R., Lin, E., and Zame, A. (1981). Another arc sine law, ***Sankhyā*, *Series A*, 43,** 371–373.

Chen, R. C. H. (1978). Generating beta variates with nonintegral shape parameters, ***Communications of the Association for Computing Machinery*, 21,** 317–322.

Clark, R. E. (1953). Percentage points of the incomplete beta function, *Journal of the American Statistical Association,* **48, 831–843.**

Cramér, H. (1946). *Mathematical Methods of Statistics,* Princeton, NJ: Princeton University Press.

Delignon, Y., Garello, R., and Hillion, A. (1990). A statistical characterization of sea-state SAR images, *Proceedings of Ocean's '90 Conference,* IEEE Service Center, Piscataway, NJ, pp. 398–401.

Dennis, S. Y., III. (1994). On the distribution of products of independent beta variates, *Communications in Statistics-Theory and Methods, 23,* 1895–1913.

Devroye, L. (1986). *Non-uniform Random Variate Generation,* New York: Springer-Verlag.

Dishon, M., and Weiss, G. H. (1980). Small sample comparison of estimation methods for the beta distribution, *Journal of Statistical Computation and Simulation, 11,* 1–11.

Dodge, H. F., and Romig, H. G. (1959). *Sampling Inspection Tables, Single and Double Sampling,* 2nd Edition, New York: Wiley. (Originally published in *Bell System Technical Journal*, 20 [1941].)

Dutka, J. (1981). The incomplete beta function—A historical profile, *Archive for History of Exact Sciences,* 24, 11–29.

Elderton, W. P., and Johnson, N. L. (1969). *Systems of Frequency Curves*, Cambridge: Cambridge University Press.

Ennis, D. M., and Johnson, N. L. (1993). Noncentral and central chi-square, F and beta distribution functions as special cases of the distribution of an indefinite quadratic form, *Communications in Statistics—Theory and Methods, 22,* 897–905.

Fan, D.-Y. (1991). The distribution of the product of independent beta variables, *Communications in Statistics—Theory and Methods,* 20, 4043–4052.

Fang, K. T., and Wang, Y. (1989). A sequential algorithm for optimization and its applications to regression analysis, *Manuscript,* Academia Sinica, Beijing, China.

Fang, K. T., and Yuan, K.-H. (1990). A unified approach to maximum likelihood estimation, *Manuscript,* Academia Sinica, Beijing, China.

Farnum, N. R., and Stanton, L. W. (1987). Some results concerning the estimation of beta distribution parameters in PERT, *Journal of the Operational Research Society, 38,* 287–290.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications,* vol. 2, New York: Wiley.

Fielitz, B. D., and Myers, B. L. (1975). Estimation of parameters in the beta distribution, *Decision Sciences,* 6, 1–13.

Fielitz, B. D., and Myers, B. L. (1976). Estimation of parameters in the beta distribution: Reply, *Decision Sciences* 7, 163–164.

Forsythe, G. E. (1972). von Neumann's comparison method for random sampling from the normal and other distributions, *Mathematics of Computation, 26,* 817–826.

Fosam, E. B., and Sapatinas, T. (1994). On characterizing the Pareto and power-function distributions within the context of multiplicative damage and generating models, *Presented at the Third World Congress of the Bernoulli Society,* Chapel Hill. North Carolina.

Fox, B. L. **(1963).** Generation of random samples from the beta and F distributions, ***Technometrics, 5*, 269–270.**

Frankl, P., and Maehara, H. **(1990).** Some geometric applications of the beta distribution, ***Annals of the Institute of Statistical Mathematics,* 42, 463–474.**

Ganter, W. A. **(1990).** Comments with reply, on Reliability of modified designs: A Bayes analysis of an accelerated test of electronic assemblies, ***IEEE Transactions on Reliability,* 39, 520–522.**

Gleissner, W. **(1984).** Some recursion formulae for the beta distribution, ***Statistische Hefte*, 25, 235–236.**

Gnanadesikan, R., Pinkham, R. S., and Hughes, L. P. **(1967).** Maximum likelihood estimation of the parameters of the beta distribution from smallest order statistics, ***Technometrics,* 9, 607–620.**

Goertzel, G. **(1958).** An algorithm for the evaluation of finite trigonometric series, ***American Mathematical Monthly,* 65, 34–35.**

Goldfarb, A. S., and Gentry, J. W. **(1979).** Applications of the Beta distribution to particle size distribution, ***Staub-Reinhaltung der Luft,* 49, 125–130.**

Golenko-Ginzburg, D. **(1988).** The distribution of activity time in PERT, ***Journal of the Operational Research Society,* 39, 767–771.**

Govindarajulu, Z., and Hubacker, N. W. **(1962).** ***Percentile points of order statistics in samples from beta, normal, chi (1 d.f.) populations***, Case Institute of Technology, Statistical Laboratory Publication No. **101,** Cleveland, OH.

Graham, V. A., and Hollands, K. G. T. **(1990).** Method to generate synthetic hourly solar radiation globally, ***Solar Energy,* 44, 333–341.**

Guenther, W. C. **(1967).** A best statistic with variance not equal to the Cramér-Rao lower bound, ***American Mathematical Monthly,* 74, 993–994.**

Han, R.-J., Chang, Y., Pao, J.-R., Ranade, M. B., and Gentry, J. W. **(1989).** Applications of probit analysis with the log Beta distribution, ***Staub-Reinhaltung der Luft,* 49, 125–130.**

Harrop-Williams, K. **(1989).** Random nature of soil porosity and related properties, ***Journal of Engineering Mechanics, 115,* 1129–1133.**

Hart, L. **(1990).** Reliability of modified designs: A Bayes analysis of an accelerated test of electronic assemblies, ***IEEE Transactions on Reliability,* 39, 140–144.**

Harter, H. L. **(1964).** ***New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-square and Beta Distributions,*** Aerospace Research Laboratories, Wright-Patterson Air Force Base, OH.

Harter, H. L. **(1978).** Adaptive robust estimation of location and scale parameters of symmetric populations, ***Tech. Rep.*** AFFDL-TR-78-128, Wright-Patterson Air Force Base, OH.

Hartley, H. O., and Fitch, E. R. **(1951).** A chart for the incomplete beta-function and the cumulative binomial distribution, ***Biometrika,* 38, 423–426.**

Haynes, L. W., and Yau, K. W. **(1990).** Single-channel measurement from the cyclic GMP-activated conductance of catfish retinal cones, ***Journal of Physiology (London),* 429, 1451–1481.**

Heselden, G. P. M. **(1955).** Some inequalities satisfied by incomplete beta functions, ***Skandinavisk Aktuarietidskrift*,** 38, **192–200.**

Hillier, F. S., and Lieberman, G. J. **(1980).** ***Introduction to Operations Research,*** 3rd ed., San Francisco: Holden-Day.

Hodges, J. L., Jr. **(1955).** On the noncentral beta-distribution, ***Annals of Mathematical Statistics,*** **26, 648–653.**

Horsnell, G. **(1957).** Economical acceptance sampling schemes, ***Journal of the Royal Statistical Society, Series A*****, 120, 148–191.**

IMSL Institute of Mathematical and Statistical Libraries **(1977, 1985).** Packages, published by IMSL, Houston, TX.

Jain, G. S., and Consul, P. C. **(1971).** **A** generalized negative binomial distribution, ***SIAM Journal of Applied Mathematics,*** **21, 501–513.**

Jambunathan, M. V. **(1954).** Some properties of beta and gamma distributions, ***Annals of Mathematical Statistics,*** **25,** 401–405.

Janardan, K. G., and Padmanabhan, G. **(1986).** Double bounded beta distribution for hydrologic variables, ***Proc.*** 17th ***Annual Pittsburgh Conference (Part 3*****), 17, 1107–1111.**

Johnk, M. D. **(1964).** Erzeugung von beta-verteilten und gamma-verteilten Zufall-szahlen, ***Metrika,*** **8, 5–15.** (In German)

Johnson, N. L. **(1949).** Systems of frequency curves generated by methods of translation, ***Biometrika,*** **36, 149–176.**

Johnson, N. L. **(1960).** An approximation to the multinomial distribution: Some properties and applications, ***Biometrika,*** **47, 93–102.**

Johnson, N. L., and Kotz, S. **(1990).** Use of moments in deriving distributions and some characterizations, ***Mathematical Scientist,*** **15, 42–52.**

Johnson, N. L., and Kotz, S. **(1994).** Further results on limit distributions arising from iterated random subdivisions of an interval, ***Statistics & Probability Letters*** (to appear).

Kalinin, V. M. **(1968).** Limit properties of probability distributions, ***Akademia Nauk SSSR,*** Steklov Institute, **104,** 88–134.

Karavias, E., and Myers, A. L. (1992). Molecular thermodynamics of absorption from gas mixtures. Composition of absorbed phase from gravimetric data, ***Chemical Engineering Science,*** **47, 1441–1451.**

Keeping, E. S. **(1962).** ***Introduction to Statistical Inference,*** New York: Van Nostrand.

Kennedy, D. P. **(1988).** A note on stochastic search methods for global optimization, ***Advances in Applied Probability,*** **20, 476–478.**

Kennedy, W. J., and Gentle, J. E. **(1980).** ***Statistical Computing,*** **New York: Dekker.**

Koppes, L. J., and Grover, N. B. **(1992).** Relationship between size of parent at cell division and relative size of progeny in ***Escherichia Coli*****,** ***Archives of Microbiology*****, 157, 402–405.**

Koshal, R. S. **(1933).** Application of the method of maximum likelihood to the improvement of curves fitted by the method of moments, ***Journal of the Royal Statistical Society, Series A*****, 96, 303–313.**

Koshal, R. S. (1935). Application of the method of maximum likelihood to the derivation of efficient statistics for fitting frequency curves, ***Journal of the Royal Statistical Society, Series A*****, 98, 265–272.**

Kotlarski, I. (1962). On groups of *n* independent random variables whose product follows the beta distribution, ***Colloquium Mathematicum***, 9, 325–332.

Kottes, J. F., and Lau, H.-S. (1978). On estimating parameters for beta distributions, ***Decision Sciences***, ***9***, 526–531.

Królikowska, T. (1966). O pewnej własności granicznej rozkładu beta, ***Zeszyty Naukowe Politechniki Lódzkiej***, ***77***, 15–20. (In Polish)

Kulldorff, G. (1957). On the conditions for asymptotic efficiency of maximum likelihood estimates, ***Skandinacisk Aktuarietidskift***, **40**, 129–144.

Laha, R. G. (1964). On a problem connected with beta and gamma distributions, ***Transactions of the American Mathematical Society***, **113**, 287–298.

Lamberson, L. R., and DeSouza, D. I. (1987). Bayesian Weibull estimation, ***Transactions, 41st Annual Quality Congress***, American Society for Quality Control, Milwaukee, WI, pp. 497–506.

Lau, H.-S., and Lau, A. H.-L. (1991). Effective procedures for estimating beta distribution's parameters and their confidence intervals, ***Journal of Statistical Computation and Simulation***, ***38***, 139–*150*.

Lauritzen, S. L., Thommesen, C., and Andersen, J. B. (1990). A stochastic model in mobile communication, ***Stochastic Processes and Their Applications***, **36**, 165–172.

Lee, C. M. S. (1989). On the computation of incomplete beta integral, ***Proc. ASA Statistical Computing Section***, pp. 89–93.

Lee, C. M. S. (1992a). Efficiency comparison of methods computing incomplete beta integral, ***Proceedings of ASA Statistical Computing Section***, pp. 108–122.

Lee, C. M. S. (1992b). On computation of central and non-central beta probabilities, ***Journal of Statistical Computation and Simulation***, **43**, 1–10.

Libby, D. L., and Novick, M. R. (1982). Multivariate generalized beta distributions with applications to utility assessment, ***Journal of Educational Statistics***, ***7***, 271–294.

Lingappaiah, G. S. (1988). Generalized length biased distributions, ***Aplikace Matematiky***, **33**, 354–361.

Ludwig, O. (1963). Algorithm 179. The incomplete beta ratio, ***Communications of the Association for Computing Machinery***, **6**, 314.

Maffett, A. L., and Wackerman, C. C. (1991). The modified beta density function as a model for synthetic aperture radar clutter statistics, ***IEEE Transactions on Geoscience and Remote Sensing***, ***29***, 277–283.

Magnussen, S. (1992). A distribution model for heritability, ***Genome***, 35, 931–938.

Majumder, K. L., and Bhattacharjee, G. P. (1973a). Algorithm AS63: The incomplete Beta integral, ***Applied Statistics***, 22, 409–411.

Majurnder, K. L., and Bhattacharjee, G. P. (1973b). Algorithm AS179: The Beta incomplete Beta integral, ***Proceedings of ASA Statistical Computing Section***.

Massonie, J. P. (1965). Utilization d'une statistique linéaire correcte pour estimer la vitesse d'évasion des étoiles dans le plan galactique, ***Publications de l'Institut de Statistique de l'Université de Paris***, **14**, 1–62. (In French)

McHenry, C. E. (1976). A note on expressing Wilks' Λ as a product of independent beta variables. ***Communications in Statistics***, **A5**, 1047–1053.

McNally, R. J. (1990). Maximum likelihood estimation of the parameters of the prior distributions of three variables that strongly influence reproductive performance in cows, *Biometrics*, **46**, 501–514.

Mendenhall, W., and Lehman, E. H. (1960). An approximation to the negative moment of the positive binomial useful in life testing, *Technometrics*, *2*, 227–242.

Meyer, D., Groebe, K., and Thews, G. (1990). Theoretical analysis of factors influencing recovery of ventilation distributions from inert gas washout data, *Advances in Medical and Experimental Biology*, *277*, 615–624.

Milyutin, E. R., and Yaromenko, Yu. I. (1991). Statistical characteristics of atmospheric transparency index over tilted routes, *Meteorologiya i Gidrologiya*, **12**, 72–76.

Moitra, S. D. (1990). Skewness and the beta distribution, *Journal of the Operational Research Society*, **41**, 953–961.

Molina, E. C. (1932). An expansion for Laplacian integrals in terms of incomplete gamma functions, and some appplications, *Bell System Technical Journal*, **11**, 563–575.

Morgan, M. G., and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitatice Risk and Policy Analysis*, Cambridge: Cambridge University Press.

Mudholkar, G. S., and Chaubey, Y. P. (1976). Some refinements of the Wise-approximation for the beta and F distributions, *Utilitas Mathematica*, **10**, 199–207.

Mühlbach, G. von (1972). Rekursionsformeln für die zentralen Momente der Pólya- und der Beta-Verteilung, *Metrika*, **19**, 173–177. (In German)

Nair, K. R. (1948). The studentized form of the extreme mean square test in the analysis of variance, *Biometrika*, *35*, 16–31.

Norton, R. M. (1975). On properties of the arc sine law, *Sankhyā*, *Series B*, *37*, 306–308.

Norton, R. M. (1978). Moment properties and the arc sine law, *Sankhyā*, *Series A*, **40**, 192–198.

O'Connor, C. A., Hook, A., and O'Connor, T. A. (1985). Beta distributions as limiting distributions, *Proceedings of ASA Statistical Computing Section*, pp. 292–293.

Onukogu, I. B., and Gupta, D. (1983). The distribution of the product of two independent beta variables, *Biometrical Journal*, **25**, 621–624.

Osborn, D., and Madey, R. (1968). The incomplete beta function and its ratio to the complete beta function, *Mathematics of Computation*, **22**, 159–162.

Patel, I. D., and Khatri, C. G. (1978). A Lagrangian beta distribution, *South African Statistical Journal*, **12**, 57–64.

Pearson, E. S., and Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, 1, Cambridge: Cambridge University Press (2d ed., 1958; 3d ed., 1966.)

Pearson, K. (1895). Mathematical contribution to the theory of evolution. II. Skew variation in homogeneous material, *Philosophical Transactions of the Royal Society of London, A*, **186**, 343–414.

Pearson, K. (ed.) (1968). *Tables of the Incomplete Beta Function*, 2d ed., Cambridge: Cambridge University Press. (1st ed., 1934).

Pearson, K., and Pearson, M. V. (1935). On the numerical evaluation of high order incomplete Eulerian integrals, ***Biometrika*, 27, 409–412.**

Pederzoli, G. **(1985).** Some properties of beta functions and the distribution for the product of independent beta random variables, ***Trabajos de Estadística y de Investigación Operatica*, 36, 122–128.**

Peizer, D. B., and Pratt, J. W. **(1968).** A normal approximation for binomial, F, beta, and other common, related tail probabilities, I, ***Journal of the American Statistical Association*, 63, 1416–1456.**

Pekh, P. A., and Marchenko, A. S. **(1992).** Computational probability properties of the Johnk algorithm, In ***Approximate Solutions of Some Problems in Mathematical Physics***, Lutsk Industrial Institute, Lutsk, Russia, pp. **15–17.** (In Russian)

Peleg, M., and Normand, M. D. (1986). Simulation of size reduction and enlargement processes by a modified version of the beta distribution function, ***American Institute of Chemical Engineers Journal*, 32, 1928–1930.**

Pham, T. G., and Turkkan, N. **(1994).** Reliability of a standby system with beta-distributed component lives, ***IEEE Transactions on Reliability*, R-43, 71–75.**

Pham-Gia, T. **(1994).** Value of the beta prior information, ***Communications in Statistics-Theory and Methods*, 23, 2175–2195.**

Pham-Gia, T., and Duong, Q. P. (1989). The generalized beta and F distributions in statistical modelling, ***Mathematical and Computer Modelling*, 13, 1613–1625.**

Pham-Gia, T., and Turkkan, N. **(1992).** Determination of the beta distribution from its Lorenz curve, ***Mathematical and Computer Modelling*, 16, 73–84.**

Pham-Gia, T., and Turkkan, N. **(1993).** Bayesian analysis of the difference of two proportions, ***Communications in Statistics—Theory and Methods*, 22, 1755–1771.**

Pham-Gia, T., Turkkan, N., and Duong, Q. P. **(1992).** Using the mean deviation in the elicitation of the prior distribution, ***Statistics & Probability Letters*, 13, 373–381.**

Popplewell, L. M., and Peleg, M. **(1992).** Theoretical kinetic model for particulates disintegration process that result in bimodal size distributions, ***Powder Technology*, *70*, 21–30.**

Posten, H. O. (1982). Computer algorithms for the classical distribution functions, ***Proceedins of 1st International Conference on Teaching Statistics, Unicersity of Sheffield, England***, pp. **313–330.**

Posten, H. O. **(1986).** Algorithms for the beta distribution, ***Proceedings of Computational Statistics, Rome 1986***, (eds., F. de Antonio, N. Lauro, and A. Rizzi), Heidelberg and Vienna: Physica-Verlag, pp. **309–319.**

Pratt, J. W. **(1968).** A normal approximation for binomial, F, beta and other common, related tail probabilities, II, ***Journal of the American Statistical Association*, 63,** 1457–1483.

Rahman, S., Khallat, M. A., and Salameh, Z. M. (1988). Characterization of insolation data for use in photovoltaic system analysis models, ***Energy (Oxford)*, 13, 63–72.**

Ratnaparkhi, M. V., and Mosimann, J. E. (1990). On the normality of transformed beta and unit-gamma random variables, ***Communications in Statistics—Theory and Methods*. 19. 3833–3854.**

Ressel, W. (1991). Traffic flow and capacity of work sites on freeways, ***Proc. Industrial Symposium on Highway Capacity and Level of Serrice,*** Karlsruhe, Germany. Rotterdam, Netherlands: Balkema, pp. 321–328.

Riggs, L. S. (1989). Numerical approach for generating beta random variables, ***Journal of the Computing Dicision of American Society of Chemical Engineers,*** **3,** 183–191.

Romesburg, H. C. (1976). Estimation of parameters in the beta distribution, ***Decision Sciences,*** 7, 162.

Roy, M. K., Roy, A. K., and Ali, M. Masoom (1993). Binomial mixtures of some standard distributions, ***Journal of Information & Optimization Sciences,*** **14,** 57–71.

Runyan, M. R., Pearson, C. L., Wang, C. C., Han, R., Plunkett, S., Kaplan, C., and Gentry, J. W. (1988) The use of the log Beta distribution to characterize atmospheric aerosols, In ***Lecture Notes in Physics,*** 309, Berlin: Springer-Verlag, pp. 65–68.

Scheffé, H. (1944). Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for a binomial ***p, Biometrika,*** **33,** 181.

Schmeiser, B. W., and Shalaby, M. A. (1980). Acceptance/rejection methods for beta variate generation, ***Journal of the American Statistical Association, 75,*** 673–678.

Seber, G. A. F. (1963). The non-central chi-squared and beta distributions, ***Biometrika,*** 50, 542–544.

Shantaram, R. M. (1981). On the stochastic equation $X + Y = XY$, In ***Statistical Distributions in Scientific Work,*** vol. *4* (eds. C. Taillie, G. P. Patil, and S. Kotz), Dordrecht: Reidel, pp. 419–431.

Shaw, M. (1991). Use of Bayes' theorem and the beta distribution for reliability estimation purposes, ***Reliability Engineering and System Safety*** **31,** 145–153.

Silver. E. A. (1963). Markovian decision processes with uncertain transition probabilities or rewards, ***Operations Research Center, MIT, Technical Report No.*** 1.

Sopel'nik, Yu. V., and Lerchenko, L. S. (1991). Detection characteristics of fluctuating radar signals with the beta distribution of power probabilities, ***Izvestiya*** WZ: ***Radioelektronika,*** **32,** 61–63. (In Russian)

Soper, H. E. (1921). The numerical evaluation of the incomplete B-function or the integral $\int_0^x x^{p-1}(1-x)^{q-1}\,dx$ for ranges of x between 0 and 1, ***Tracts for Computers,*** Vol. 7, Cambridge: Cambridge University Press.

Springer, M. D. (1978), ***Algebra of Random Variables,*** New York: Wiley.

Steece, B. M. (1976). On the exact distribution for the product of two independent beta-distributed variables, ***Metron,*** **34,** 187–190.

Stuart, A. (1962). Gamma-distributed products of independent random variables, ***Biometrika,*** **49,** 564–565.

Sukvittayawong, S., and Inasaki, I. (1991). Identification of chip form in turning process, ***Japanese Society of Mechanical Engineers, Series 3: Vibration, Control Engineering, Engineering for Industry,*** **34,** 553–560.

Swain, J., Venkataraman, S., and Wilson, J. (1988). Least squares estimation of distribution functions in Johnson's translation system, ***Journal of Statistical Computation and Simulation,*** **29, 271–297.**

Tang, J., and Gupta, A. K. (1984). On the distribution of the product of independent beta random variates, ***Statistics & Probability Letters,*** **2, 165–168.**

Temme, N. M. (1987). Incomplete Laplace integrals: Uniform asymptotic expansion with application to the incomplete beta function, ***SIAM Journal on Mathematical Analysis, 18,*** **1638–1663.**

Temme, N. M. (1992). Asymptotic inversion of the incomplete beta function, ***Journal of Computational and Applied Mathematics,*** **41, 145–147.**

Thompson, C. M. **(1941).** Tables of percentage points of the incomplete beta-function, ***Biometrika,*** **32, 151–181.**

Thompson, P. **(1990).** Combination of expert opinion approach to probabilistic information retrieval. Part 2: Mathematical treatment of CEO Model **3,** ***Information Processing and Management,*** **26, 383–394.**

Thomson, D. H. (1947). Approximate formulae for the percentage points of the incomplete beta function and the χ^2 distribution, ***Biometrika*****, 34, 368–372.**

Treacy, P., Ochs, J. B., Ozsoy, T. M., and Wang, N. (1991). Automated tolerance analysis for mechanical assemblies modeled with geometric features and relational data structure, ***Computer Aided Design,*** **23, 444–452**.

Tretter, M. J., and Walster, G. W. **(1979).** Continued fractions for the incomplete beta function: Additions and corrections, ***Annals of Statistics, 7,*** 462–465.

Van Assche, W. **(1987).** A random variable uniformly distributed between two random variables, ***Sankhyā*****,** ***Series A*****, 49, 207–211.**

Van Dingenan, R., Raes, F., and Vanmarcke, H. (1987). Application of the log Beta distribution for aerosol size distribution, ***Journal of Aerosol Science,*** **18, 663–666.**

Vogler, L. E. **(1964).** Percentage points of the beta distribution, ***Technical Note,*** No. **215,** National Bureau of Standards, Boulder, CO.

Volodin, I. N. **(1970).** Beta distribution for small values of parameters, ***Theory of Probability and Its Applications,*** **15, 563–566.**

Volodin, N. **(1994).** Some generalizations of the beta distribution, ***Private Communication.***

Wang, Z., and Dornfeld, D. A. **(1992).** In-process tool wear monitoring using neural networks, ***Proceedings of*** 1992 ***Japan–USA Symposium on Flexible Automation, Part I***, pp. **263–277.**

Whitby, O. (1971). Estimation of parameters in the generalized beta distribution, ***Technical Report,*** No. **29,** Dept. of Statistics, Stanford University, CA.

Wigner, F. (1958). On the distribution of the roots of certain symmetric matrices, ***Annals of Mathematics, 67,*** **325–327.**

Wiley, J. A., Herschokoru, S. J., and Padiau, N. S. **(1989).** Heterogeneity in the probability of HIV transmission per sexual contact: The case of male-to-female transmission in penile-vaginal intercourse, ***Statistics in Medicine,*** **8, 93–102.**

Wise, M. E. **(1950).** The incomplete beta function as a contour integral and a quickly converging series for its inverse, ***Biometrika,*** **37, 208–218.**

Wise, M. E. (1960). On normalizing the incomplete beta-function for fitting to dosage-response curves, ***Biometrika,*** 47, 173–175.

Wishart, J. (1927). On the approximate quadrature of certain skew curves, with an account of the researches of Thomas Bayes, ***Biometrika,*** 19, 1–38.

Woods, J. D., and Posten, H. O. (1968). Fourier series and Chebyshev polynomials in statistical distribution theory, ***Research Report*** No. 37, Department of Statistics, University of Connecticut, Storrs.

Woods, J. D., and Posten, H. O. (1977). The use of Fourier series in the evaluation of probability distribution functions, ***Communications in Statistics,*** B6, 201–219.

Yamazaki, H. (1990). Breakage models and intermittency, ***Journal of Fluid Mechanics,*** 219, 181–193.

Yang, S., Li, Z., and Li, G. (1988). Statistical inferences of form error distribution, ***Precision Engineering,*** 10, 97–99.

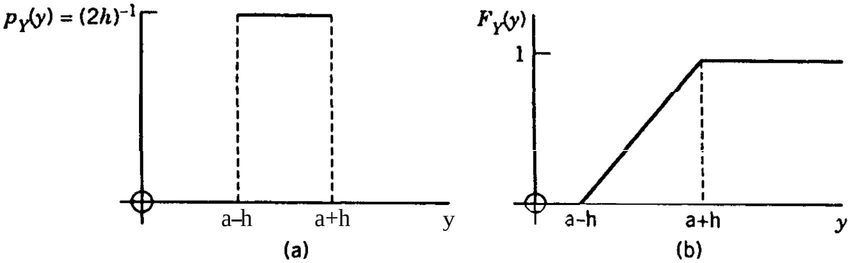


Figure 26.1 **(a) Uniform density function; (b) Uniform distribution function.**

The linear transformation transforming the uniform $(0, \theta)$ back to uniform $(a - h, a + h)$ is $Y = a - h + (2h/\theta)Z$. One may similarly define the uniform distribution on the interval $(a, a + \theta)$ with probability density function

$$p_X(x) = \begin{cases} \dfrac{1}{\theta}, & a \le x \le a + \theta, \\ 0, & \text{otherwise.} \end{cases}$$

The ***standardized*** uniform distribution (having mean 0 and standard deviation 1) has the pdf

$$p_U(u) = \frac{1}{2\sqrt{3}}, \qquad -\sqrt{3} \le u \le \sqrt{3}. \tag{26.3}$$

We will refer to the uniform distribution over (0, 1), with the pdf

$$p_R(r) = 1, \qquad 0 \le r \le 1, \tag{26.4}$$

as the ***standard*** uniform distribution. If R is uniformly distributed between 0 and 1, then

$$U = \sqrt{3}(2R - 1)$$

has the standardized uniform distribution.

2 GENESIS

The uniform distribution (26.1), with $a = 0$ and $h = \frac{1}{2} \times 10^{-k}$, is often used to represent the distribution of roundoff errors in values tabulated to the nearest k decimal places. Of course, if the rounding were applied to figures expressed in a binary scale, we would have $h = 2^{-(k+1)}$.

Nagaev and Mukhin (1966) have investigated conditions under which a rectangular distribution of roundoff errors is to be expected. In particular they have shown that if $X_1, X_2, \ldots$ are independent random variables with characteristic functions $E[e^{itX_j}] = \phi_j(t)$, then for any positive integer a, a necessary and sufficient condition for

$$\lim_{n\to\infty} \Pr\left\{ \sum_{j=1}^{n} X_j - a\left[\sum_{j=1}^{n} \frac{X_j}{a}\right] \le x \right\} = \frac{x}{a}, \qquad 0 \le x < a,$$

is that

$$\prod_{j=1}^{n} \phi_j\left(\frac{2\pi k}{a}\right) = 0 \qquad \text{for } k = \pm 1, \pm 2, \ldots.$$

In the expression above $[\sum_{j=1}^{n} X_j/a]$ means "integral part of $\sum_{j=1}^{n} X_j/a$," so $\sum_{j=1}^{n} X_j - a[\sum_{j=1}^{n} X_j/a]$ is the roundoff error of $\sum_{j=1}^{n} X_j$ in units of a. The condition $\prod_{j=1}^{n} \phi_j(2\pi k/a) = 0$ is certainly satisfied if all X's have the same distribution and if $|\phi_j(2\pi k/a)| < \eta_k < 1$ for infinitely many j and some $\eta_k < 1$. Holewijn (1969) showed that if

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \phi_j(2\pi k) = 0, \qquad k = 1, 2, \ldots,$$

then the fractional part sequence is uniformly distributed "almost certainly" —that is, nearly all the variables $\{X_j - [X_j]\}$ have unit uniform distributions.

A rectangular distribution also arises as a result of the probability integral transformation [see Quesenberry (1986)]. If X is a continuous variable and $\Pr[X \le x] = F(x)$, then $F(X)$ is distributed according to (26.1) with a = h = I or equivalently according to (26.4). This result [first employed by Fisher (1932)] has been applied in a number of ways [Durbin (1961), Pearson (1938), Stephens (1966)] in various techniques for combining results of statistical tests (see Section 9).

3 HISTORICAL REMARKS

The uniform distribution is so natural a conception that it has probably been in use far more than can be inferred from printed records. Among such records we may mention, in particular, descriptions of the use of the distribution by Bayes (1763) and Laplace (1812).

Some particular historical interest attaches to the distribution of the sum of independent random variables, each having the same rectangular distribution. Seal (1950) gives an extensive bibliography on this subject (see also Section 26.9).

4 GENERATING FUNCTIONS, MOMENTS, AND ORDER STATISTICS

The expected value of a random variable Y with probability density function (26.1) is a. The distribution is symmetrical, and all odd central moments are zero. If r is even, the rth central moment of Y is

$$\mu_r(Y) = (r+1)^{-1}h^r. \tag{26.5}$$

It follows that $\text{var}(Y) = (1/3)h^2$ and that $\sqrt{\beta_1} = \alpha_3 = 0$ and $\beta_2 = \alpha_4 = 1.8$. For the standard uniform distribution (26.4) the expected value is 1/2 and the variance is 1/12 (corresponding to $h = 1/2$).

Formula (26.5) gives the values of the rth absolute central moment for all positive r. In particular, the mean deviation is $h/2$. Hence for this distribution

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \frac{\sqrt{3}}{2} = 0.866. \tag{26.6}$$

The characteristic function is $E[e^{itY}] = e^{ita}\{\sin(th)/th\}$. The moment-generating function is $E[e^{tY}] = e^{ta}\{\sinh(th)/th\} = \frac{1}{2}(th)^{-1}\{e^{t(a+h)} - e^{t(a-h)}\}$. The cumulants are

$$\begin{cases} \kappa_1 = a,\ \kappa_r = 0 & (r > 1 \text{ and odd}), \\ \kappa_r = \dfrac{2^{r-1}h^r B_r}{r} & (r \text{ even}) \end{cases} \tag{26.7}$$

(B_r is the rth Bernoulli number; see Chapter 1, Section A9).

The information-generating function [$(u-1)$th frequency moment] is

$$T(u) = (2h)^{1-u}. \tag{26.8}$$

The entropy is $-T'(1) = \log(2h)$. The characteristic function of the uniform $(0,1)$ variable can be written

$$\phi_X(t) = \frac{e^{it}-1}{it} = \sum_{j=0}^{\infty} \frac{(it)^j}{(j+1)!}. \tag{26.9}$$

The moment-generating function of this variable is $(e^t - 1)/t$. More generally, the moment-generating function of uniform $(0, \theta)$ random variable is

$$\frac{e^{\theta t}-1}{\theta t} = e^{\theta t/2}\frac{\sinh(\theta t/2)}{\theta t/2}.$$

[Note that Haight (1961) gave an erroneous expression.]

For the uniform distribution on the interval $(a, a + \theta)$, the Lorenz curve is given by

$$L(u) = \{au + (\theta u^2/2)\}/\{a + (\theta/2)\},$$

and the Gini index is $\theta\{3(\theta + 2a)\}$.

The pdf's of order statistics $X_1' \le X_2' \le \cdots \le X_n'$ of random samples of size n from the uniform distribution (26.1) can be computed from the general formula [Chapter 12, Eq. (12.14)], which gives:

$$p_{X_1',\dots,X_n'}(x_1,\cdots,x_n) = n!(2h)^{-n}, \qquad a - h \le x_1 \le \cdots \le x_n \le a + h, \tag{26.10a}$$

or, for the standard uniform distribution $(a = h = \frac{1}{2})$

$$p_{X_1',\dots,X_n'}(x_1,\cdots,x_n) = n!, \qquad 0 \le x_1 \le \cdots \le x_n \le 1. \tag{26.10b}$$

The distribution of X_j' corresponding to (26.10b) is

$$p_{X_j'}(x) = \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1-x)^{n-j}, \qquad 0 \le x \le 1. \tag{26.11a}$$

This is a standard beta $(j, n - j + 1)$ distribution (cf. Equation (25.2) in Chapter 25).

The joint distribution of X_i' and X_j' is

$$p_{X_i', X_j'}(x_i, x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}x_i^{i-1}(x_j - x_i)^{j-i-1}(1-x_j)^{n-j},$$

$$0 \le x_i \le x_j \le 1. \tag{26.11b}$$

We have

$$E\left[X_i'^r(1-X_j')^s\right] = \frac{i^{[r]}(n-j+1)^{[s]}}{(n+1)^{[r+s]}}$$

$$\text{with} \quad a^{[b]} = a(a+1)\dots(a+b-1), \tag{26.11c}$$

whence

$$E[X_i'] = (n+1)^{-1}i,$$

$$\operatorname{var}(X_i') = (n+1)^{-2}(n+2)^{-1}i(n+1-i),$$

$$\operatorname{cov}(X_i', X_j') = (n+1)^{-2}(n+2)^{-1}i(n+1-j).$$

From the joint distribution of the least and greatest order statistics (X_1' and X_n') the distribution of range ($W = X_n' - X_1'$) can be shown to have the pdf

$$p_W(w) = n(n-1)w^{n-2}(1-w), \qquad 0 < w < 1. \tag{26.12}$$

So W has a beta $(n-1, 2)$ distribution.

The distribution of the ratio $W_{ij} = X_i'/X_j'$ $(i < j)$ corresponding to (26.11b) is just that of the ith order statistic in a sample of size $j-1$ from a standard uniform distribution; this is given by (26.11a) with n replaced by $j-1$ and j replaced by i, namely a standard beta $(i, j-i)$ distribution. The distribution of the product $Y_{ij} = X_i'X_j'$ $(i < j)$ based on a random sample of size n from the uniform (0, 1) distribution, can be derived from (26.11b). The pdf is

$$p_Y(y) = \begin{cases} \dfrac{1}{2}\dfrac{n!}{(i-1)!(n-j)!(j-i)!}y^{(j/2)-1}(1-y)^{j-i}\displaystyle\sum_{k=0}^{n-j}(-1)^k\binom{n-j}{k}y^{k/2} & \\ \quad \times {}_2F_1(j-i, \tfrac{1}{2}(j+k)-i+1; j-i+1; 1-y), & 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases} \tag{26.13a}$$

For $i = 1$ and $j = n$ we have

$$p_Y(y) = \begin{cases} ny^{(n/2)-1}(1-y)^{n-1}{}_2F_1(n-1, \tfrac{1}{2}n-1; n; 1-y), & 0 < y < 1, \\ 0, & \text{elsewhere}, \end{cases} \tag{26.13b}$$

where the function ${}_2F_1$ is as defined in Chapter 1, Section **A6.**

5 CHARACTERIZATIONS

Characterizations of the uniform distribution often provide useful tools for constructing goodness-of-fit statistics, simulation of complex statistical procedures, and testing the quality of pseudorandom number generators. Many characterizations of the uniform distribution can be traced to the corresponding characterizations of the exponential distribution (Chapter 19) since the simple monotone transformation $X = e^{-Y}$ of a standard exponential random variable Y yields a uniform (0, 1) random variable.

For example, Hamdan (1972) has shown that X is uniformly distributed over $(0, 1)$ if and only if

$$E[-\log(1 - X)|X > y] = -\log(1 - y) + 1 \qquad \text{for } y \in [0, 1). \quad (26.14a)$$

Pusz (1988) provides alternative characterizations of the type

$$E[h(X)|X \geq x] = g(x), \qquad (26.14b)$$

where $h(\cdot)$ and $g(\cdot)$ are certain known functions.

A random variable X has the uniform distribution on $(0, 1)$ if and only if for any $a \in (0, 1)$

$$E[X^{-\alpha}|X < y] = \frac{1}{1 - \alpha} y^{-\alpha} \qquad \text{for } y \in (0, 1). \qquad (26.14c)$$

[Compare characterizations of the exponential distributions (Chapter 19, Section 8) by the properties of conditional expectations. Galambos and Kotz (1978) provide additional details.] Moreover there are several characterizations of the uniform distributions on abstract spaces which would be more appropriately designated for spherical uniform distributions [see, e.g., Brown, Cartwright, and Eagleson (1986)]. Herer (1993) obtained the following characterization: A real random variable X, with finite support, has a (continuous or discrete) uniform distribution if and only if, for any $a < b$ in its support, $E[X|a \leq X \leq b] = \frac{1}{2}(a + b)$ when $Pr[a \leq X \leq b] > 0$. The same result follows from the paper by Das Gupta, Goswami, and Rao (1993) to be discussed below. In fact, Ouyang (1993) (see below), Herer (1993), and Das Gupta, Goswami, and Rao (1993) all provide, independently, essentially the same characterization. Popular characterizations are based on correlations of order statistics.

Terrell (1983) and Abdelhamid (1985), among others, provide the following characterization: Let $X_1' < X_2'$ be the order statistics of a sample of size 2 from a continuous distribution with finite variance. Then the correlation between X_1' and X_2' is less or equal to $\frac{1}{2}$, with equality if and only if F is a uniform distribution. Papathanasiou (1990) similarly established that, in a sample of size 2, $\operatorname{cov}(X_1', X_2') \leq \frac{1}{3}\operatorname{var}(X)$, with equality holding if and only if the population distribution is uniform; also see Ma (1992). Balakrishnan and Balasubramanian (1993) showed that this characterization is equivalent to the one based on the inequality $E[X_2' - E(X)] \leq \{\frac{1}{3}\operatorname{var}(X)\}^{1/2}$ due to Hartley and David (1954) and Gumbel (1954). Székely and Móri (1985) extended Terrell's result by showing that

$$\operatorname{corr}(X_r', X_s') \leq \left(\frac{r(n + 1 - s)}{s(n + 1 - r)} \right)^{1/2}, \qquad 1 \leq r < s \leq n, \quad (26.15)$$

with equality if and only if the population distribution is uniform. The proofs provided by Terrell, Abdelhamid, and Székely and Mori are all different, but all use the Cauchy-Schwarz inequality. These characterizations show that for the uniform distribution the order statistics are the most highly correlated.

The structure of the expected spacings between consecutive order statistics was utilized for a characterization of the uniform distribution by Saleh (1976) [implicit in Cox and Lewis (1966)]. This states that under appropriate conditions on $G(u) = \inf_x\{x: F(x) \geq u\}$ and finiteness of the mean (ξ), the property $E[V_i] = \sigma/(n+1)$, $i = 1, 2, \dots, n$, where $V_i = X_i' - X_{i-1}'$ $(X_0' = \xi - \sigma/2;\ X_{n+1}' = \xi + \sigma/2)$ characterizes the uniform distribution on $[\xi - \frac{1}{2}\sigma, \xi + \frac{1}{2}\sigma]$. Applications of this characterization are plentiful in problems related to queueing theory.

Huang, Arnold, and Ghosh (1979) show that for a superadditive continuous cdf $F(\cdot)$ [satisfying $F(x+y) \geq F(x) + F(y)$ for all x, y and $x+y$ in the support of F], identical distribution of V_1 and V_k for some $k = 2, \dots, n$ implies that F is uniform. If superadditivity is replaced by bounded support and absolute continuous and monotonic density then this modified condition also characterizes uniform distributions [Ahsanullah (1989)]. [Compare with corresponding characterizations of the exponential distribution (Chapter 19, Section 8) under somewhat more restrictive conditions.] Note that for a Bernoulli random variable with probability of success $n/(n+1)$, $V_1 \stackrel{d}{=} V_2$ and so some smoothness condition on F is essential.

Shimizu and Huang (1983) show that, in the class of absolutely continuous distributions, uniform $(0, \theta)$ distributions are characterized by the property

$$X_j' - X_i' \stackrel{d}{=} X_{j-i}'. \tag{26.16}$$

Lin (1986) showed that for a sample of size 2 from any distribution with finite second moment,

$$\{E[X_2']\}^2 \leq (\tfrac{4}{3})E[X^2], \tag{26.17}$$

with equality if and only if F is degenerate at $x = 0$ or uniform on $[0, c)$ with $c = \{3E[X^2]\}^{1/2} > 0$. [The proof involves the explicit expression of $E[X_2']$ in terms of $F^{-1}(\cdot)$:

$$E[X_2'] = \int_0^1 F^{-1}(t)2t\,dt$$

and utilization of the Cauchy-Schwarz inequality.]

Lukacs (1979) characterizes uniform distribution over $[-1, 1]$ by the property that the expected values of

$$S = \frac{1}{n(n-1)} \sum_{j \neq k}^{n} \left(2X_j^3 X_k - 4X_j X_k - 6X_j^2 X_k^2\right) + \frac{4}{n} \sum_{j=1}^{n} X_j^2 \quad (26.18a)$$

and

$$T = \frac{1}{n(n-1)} \sum_{j \neq k}^{n} \left(X_j^4 X_k - 2X_j^2 X_k^3\right) + \frac{1}{n} \sum_{j=1}^{n} \left(2X_j^3 - X_j\right) \quad (26.18b)$$

do not depend on $X_1 + \cdots + X_n$. The proof involves solution of complicated differential equations for the characteristic functions.

Das Gupta, Goswami, and Rao (1993) have discovered that the conditional expectation property,

$$\Pr\left[E[X_1 | X_1', X_n'] = \tfrac{1}{2}(X_1' + X_n')\right] = 1 \quad (26.19)$$

for some $n \geq 3$, implies that the underlying distribution is either uniform over an interval or discrete uniform (see Chapter 6, Section 10.1) supported on a set of equispaced points. The main idea is that the above cited property, involving conditional expectation, determines the structure of the support of the underlying distribution.

We now note some characterizations based on inequalities among the moments, and also the so-called Chernoff-type inequalities (which are variants of the Chebyshev Type I inequalities) due to Sumitra and Kumar (1990). If X has an absolutely continuous distribution, with support $[-1, 1]$, and a symmetric density with its only mode at $x = 0$, then

$$E[X^2] \leq \sup\left\{\frac{E\left[[g(X)]^2\right]}{E\left[[g'(X)]^2\right]}\right\} \leq (8\pi^{-2})E[|X|] \quad (26.20)$$

where the supremum is taken over all even *convex* functions on $[-1, 1]$ with $g(0) = 0$. The upper bound is achieved if and only if X has a uniform distribution on $[-1, 1]$. As is the case for many characterizations based on inequalities, the final step in the proof of (26.20) is provided by the Cauchy-Schwarz inequality.

A characterization useful for testing goodness-of-fit was introduced by Seshadri and Shuster (1971) in an unpublished manuscript. It asserts that under appropriate regularity conditions, a necessary and sufficient condition that i.i.d. random variables Y_1, Y_2 possess a uniform $[0, \theta]$ distribution for some $\theta > 0$ is that $T = \min(Y_1, Y_2)/\max(Y_1, Y_2)$ is uniformly distributed on $(0, 1)$ [see, e.g., Kotz (1974) for more details].

Some remarkable characterizations are related to the following property, noted by Feller (1966): If X_1 and X_2 are mutually independent random variables assuming values in $(0, 1)$ and X_1 has a uniform $(0, 1)$ distribution, then the fractional part of $(X_1 + X_2)$—namely $Z = X_1 + X_2 - [X_1 + X_2]$, where $[a]$ is an integer part of a—also has a uniform $(0, 1)$ distribution if and only if X_2 has a uniform $(0, 1)$ distribution. [Stapleton (1963) discusses this problem in a more abstract setting.] Arnold and Meeden (1976) point out that these assumptions imply that X_1 and Z are mutually independent.

Goldman (1968) showed that if X_1 and X_2 are *i.i.d.* and Z $[= (X_1 + X_2) \bmod 1]$ and X_1 have the same distribution, then X_1 has a uniform distribution on $(0, 1)$ or a discrete rectangular distribution on a set of values $\{0, m^{-1}, 2m^{-1}, \dots, (m - 1)m^{-1}\}$ for some m. An alternative version of this result, useful in operations research, is that if X_1 and X_2 are independent random variables on $(0, 1)$ with the same cdf F, the distribution of $[(X_1 + X_2) \bmod 1]$ is F if and only if X_1 (and X_2) are uniform on $(0, 1)$ (discrete or continuous, depending on F). Arnold and Meeden (1976) obtained a similar result. Driscoll (1978) generalized the result to variables X_1 and X_2, which are mutually independent, each having a positive pdf over a finite interval $[a, b]$, and

$$Z = \begin{cases} X_1 + X_2 - a & \text{for } 2a \le X_1 + X_2 \le a + b, \\ X_1 + X_2 - b & \text{for } a + b < X_1 + X_2 \le 2b \end{cases} \tag{26.21}$$

[$a = 0$, $b = 1$ gives $Z = (X_1 + X_2) \bmod 1$]. This result has been a source for examples of sets of three identically distributed variables that are pairwise but not jointly independent.

For a continuous random variable X defined on the interval $[a, b]$, Ouyang (1993) shows that $E[X|X > c] = (b + c)/2$ for $a < c < b$ iff X is uniform. Similarly, for a sample of size n, the condition $E[X'_{k+1} - X'_k | X'_k = c] = (b - c)/(n - k + 1)$ for any $1 \le k < n$ and $a < c < b$, has also been shown by Ouyang (1993) to be a characterization of the uniform distribution. There is considerable overlap among papers discussing this type of characterizations. Somewhat different results that are useful in random number generation were obtained by Deng and George (1992). These results are summarized below.

Let U and V be independent random variables distributed over $(0, 1)$ with continuous pdfs. Then statements (26.22a)–(26.22d) are equivalent:

$$U \sim U(0,1), \tag{26.22a}$$

$$W_1 = \min\left(\frac{U}{V}, \frac{1 - U}{1 - V}\right) \sim U(0,1), \quad \text{and independent of } V, \tag{26.22b}$$

$$W_2 = |U - V|\left(\frac{\Delta}{V} + \frac{1 - \Delta}{1 - V}\right) \sim U(0,1), \quad \text{and independent of } V, \tag{26.22c}$$

where $A = I_{(V>U)}$, the indicator function of $V > U$, and

$$W_3 = (U + V) \bmod 1 \sim U(0,1), \quad \text{and independent of } V. \tag{26.22d}$$

These results provide a partial answer to the important problem of determining the family of functions g for which the uniformity of U and V implies (and is implied by) uniformity of $g(U, V)$ if U and V are independent random variables having continuous pdfs with support $(0, 1)$. (This is relevant to construction of methods for improving pseudorandom number generators to make them give results closer to standard uniform distributions.)

6 ESTIMATION OF PARAMETERS

If observations in a random sample are represented by independent random variables $Y_1, Y_2, \cdots, Y_n$ each with distribution (26.1), the likelihood function is equal to $(2h)^{-n}$ for $a - h \le \min(Y_1, \cdots, Y_n) \le \max(Y_1, \cdots, Y_n) \le a + h$. This likelihood is maximized by making h as small as possible. In other words, the maximum likelihood estimator of h is

$$\hat{h} = \tfrac{1}{2}[\text{range}(Y_1, Y_2, \cdots, Y_n)]. \tag{26.23}$$

The maximum likelihood estimator of a is therefore

$$\hat{a} = \tfrac{1}{2}[\min(Y_1, \ldots, Y_n) + \max(Y_1, \ldots, Y_n)] = \text{midrange}(Y_1, \ldots, Y_n). \tag{26.24}$$

In fact the best linear unbiased estimators of h and a are

$$(n + 1 \qquad - 1 \qquad \text{and} \quad \hat{a}, \tag{26.25a}$$

respectively. The variances of these estimators are

$$2h^2(n-1)^{-1}(n+2)^{-1} \quad \text{and} \quad 2h^2(n+1)^{-1}(n+2)^{-1}, \tag{26.25b}$$

respectively.

The estimators $\hat{a}$ and $\hat{h}$ are uncorrelated but not independent. In fact their joint probability density function is

$$p_{\hat{a},\hat{h}}(a', h') = \left(\frac{2}{h}\right)^{n-1} n(n-1)h'^{n-2},$$

$$0 \le h' \le h, 0 \le a' - a - h \le a' - a + h' \le 2h. \tag{26.26}$$

The distribution of $\hat{h}$ alone is

$$p_{\hat{h}}(h') = 2\left(\frac{2}{h}\right)^{n-1} n(n-1)h'^{n-2}(h-h'), \qquad 0 \le h' < h. \tag{26.27}$$

The distribution of $\hat{a}$ alone is

$$p_{\hat{a}}(a') = \left(f \right)^{n-1} n\left(\frac{1}{2} - \left|a' - a - \frac{1}{2}\right|\right)^{n-2}, \qquad a - h \le a' \le a + h. \tag{26.28}$$

Cumulative probabilities $\Pr[\hat{h} < H]$, $\Pr[\hat{a} < A]$ are easily evaluated from these probability density functions. The arithmetic mean $\bar{Y}$ and the median $\tilde{Y}$ are also unbiased estimators of the parameter a.

It was noted by Carlton (1946) that

$$\frac{\operatorname{var}(\hat{a})}{\operatorname{var}(\bar{Y})} = \frac{6n}{(n+1)(n+2)}$$

and that

$$\frac{\operatorname{var}(\tilde{Y})}{\operatorname{var}(\bar{Y})} = \frac{3n}{(n+2)}.$$

If n is allowed to increase, $\operatorname{var}(\hat{a})/\operatorname{var}(\bar{Y})$ tends to zero and $\operatorname{var}(\tilde{Y})/\operatorname{var}(\bar{Y})$ tends to 3. Consequently the "efficiency" of the mean is zero, and the median is only one-third as efficient as the mean. (However, since $\hat{a}$ does not have a normal limiting distribution, the concept of efficiency cannot be strictly applied here.)

It will be noted that these estimators are functions of order statistics (in fact of the smallest and greatest values). The theory of order statistics for random samples from rectangular distributions (see Section 26.4) is remarkably simple. This has led to the proliferation of methods of estimation based on various combinations of order statistics. We will discuss some of these in the next section.

The maximum likelihood estimator of the population standard deviation in the case of a uniform distribution is the sample range divided by $2\sqrt{3}$ (equivalently, sample semirange divided by $\sqrt{3}$):

$$\hat{\sigma} = \frac{1}{2\sqrt{3}}(Y_n' - Y_1'),$$

which is also an adaptive robust estimator [Harter (1978)]. Sometimes the

Table 26.1 Comparison of $\hat{N}'$ and $\hat{N}'''$

n	Mean Square Error Ratio	Closeness
1	1.333	0.571
2	1.029	0.530
3	1.001	0.505
4	1.002	0.509
5	1.008	0.519
10	1.036	0.542
20	1.061	0.556
∞	1.094	0.571

Source: Johnson (1950).

lower limit $(a - h)$ is known, and it is desired to estimate 2h. Such a case was achieved as a continuous approximation to a problem proposed by Schrodinger [referred to by Geary (1944)], who called for estimation of a number N (positive integer) given n independent integers, each equally likely to be $1, 2, 3, \dots, N$. Geary's (1944) model regards N as a parameter that can take any positive value and assumes the observed values Y_i to be uniformly distributed continuous variables:

$$p_{Y_i}(y) = N^{-1}, \qquad 0 \le y \le N, \tag{26.29}$$

which is (26.1) with $a = \frac{1}{2}N$, $h = \frac{1}{2}N$.

Johnson (1950) discussed four estimators of N, each depending only on $\max(Y_1, Y_2, \cdots, Y_n) = Y_n'$:

1. The maximum likelihood estimator Y_n'.
2. The minimum mean square error estimator $\hat{N}' = (n+2)Y_n'/(n+1)$.
3. The unbiased estimator $\hat{N}'' = (n+1)Y_n'/n$.
4. The closest estimator $\hat{N}''' = 2^{1/n}Y_n'$.

Table 26.1 shows how Johnson's method of comparison affects the assessment of "relative merit" of estimators $\hat{N}'$ and $\hat{N}'''$. The second column shows the ratio of the mean square error of $\hat{N}'''$ to the mean square error of $\hat{N}'$; the "closeness criterion" is expressed as

$$\Pr\left[|\hat{N}''' - N| < |\hat{N}' - N|\right].$$

Clearly, $\hat{N}'''$ is always a closer estimator of N than $\hat{N}'$, but $\hat{N}'$ has the smaller mean square error.

Gibbons (1974) investigated three estimators of the parameter 8 in a population uniformly distributed over $(0, \theta)$:

1. The maximum likelihood estimator $\hat{\theta}$ which is the largest sample observation X_n' with moments

$$E[\hat{\theta}] = \frac{n\theta}{(n+1)},$$

$$\text{var}(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2(n+2)}. \tag{26.30a}$$

2. The unbiased estimator $8 = (n+1)X_n'/n$ with moments

$$E[\tilde{\theta}] = 8,$$

$$\text{var}(\tilde{\theta}) = \frac{\theta^2}{n(n+2)}. \tag{26.30b}$$

3. The symmetric estimator $\theta^* = X_1' + X_n'$ with moments

$$E[\theta^*] = \theta,$$

$$\text{var}(\theta^*) = \frac{2\theta^2}{(n+1)(n+2)}. \tag{26.30c}$$

She found that the probabilities of relative errors of the estimators $\hat{\theta}$ and θ^*, given by

$$\Pr\left[\left|\frac{\hat{\theta}-\theta}{\theta}\right| < \varepsilon\right] \quad \text{and} \quad \Pr\left[\left|\frac{\theta^*-\theta}{\theta}\right| < \varepsilon\right],$$

respectively, are the same $[= 1 - (1-\varepsilon)^n]$ for $0 < \varepsilon < 1$. Gibbons and Litwin (1974) studied simultaneous estimation of the parameters a and β in the uniform density (26.1)'—namely of $(a - h)$ and $(a + h)$ in (26.1).

The simultaneous maximum likelihood estimators $\hat{\alpha} = Y_1'$ and $\hat{\beta} = Y_n'$ are jointly sufficient, consistent, and complete for a and β, but they are only asymptotically unbiased estimators. The marginal distributions of these estimators are quite skewed, with modes at the end points of the interval (a, β), respectively. Gibbons and Litwin suggest using the linear unbiased uniformly minimum variance estimators given by

$$\hat{\alpha}' = \frac{nY_1' - Y_n'}{n - 1} \tag{26.31a}$$

and

$$\hat{\beta}' = \frac{nY_n' - Y_1'}{n - 1}, \tag{26.31b}$$

which are consistent and locally symmetric, that is, $p_{\hat{\alpha}'}(\alpha + \delta) = p_{\hat{\alpha}'}(\alpha - \delta)$ and $p_{\hat{\beta}'}(\beta + \delta) = p_{\hat{\beta}'}(\beta - \delta)$ provided that $\delta/(\beta - a) \to 0$. They concluded that the expected value of the maximum absolute error [where the maximum absolute error of a pair of estimators $\hat{\theta}_1, \hat{\theta}_2$ is defined as $e(\hat{\theta}_1, \hat{\theta}_2) = \max(|(\hat{\theta}_1 - \alpha)/(\beta - \alpha)|, |(\hat{\theta}_2 - \beta)/(\beta - \alpha)|)$] for estimators $\hat{\alpha}'$, $\hat{\beta}'$ is considerably smaller for all n than the corresponding expected value for estimators Y_1, Y_n, due to unbiasedness and local symmetry. The limiting values as $n \to \infty$ are $(0.5 + 4e^{-2})/(n + 1) = 1.04/(n + 1)$ and $1.5/(n + 1)$, respectively.

A numerical example of simultaneous confidence regions for (α, β) with confidence coefficient of at least 0.95, based on the (Y_1', Y_n') and $(\hat{\alpha}', \hat{\beta}')$ sets of estimators, respectively, is presented in Figure 26.2, corresponding to $n = 20$, $Y_1' = 10$, and $Y_2' = 50$. Note that the region includes the "impossible" sample values $Y_1' < a$ and $Y_n' > \beta$.

Rukhin, Kuo, and Dey (1990) studied minimax estimation of the scale parameter h of the uniform distribution (26.1) on the interval where both h and a are unknown parameters. A version of the complete sufficient statistic

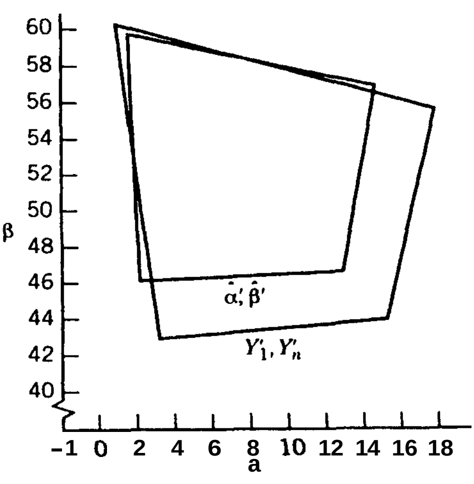


Figure 26.2 Confidence regions for (a, β) based on $(\hat{\alpha}', \hat{\beta}')$ and on (Y_1', Y_n').

for this distribution is (Y, S) with

$$Y = \tfrac{1}{2}\left(\max_i Y_i + \min_i Y_i\right) = \tfrac{1}{2}(Y_n' + Y_1'),$$

$$S = \max_i Y_i - \min_i Y_i = Y_n' - Y_1', \tag{26.32}$$

and an estimator (under a quadratic loss function) of the form $\delta(Y, S) = c_0 S[1 - g(Y/S)]$, where g is a symmetric nonnegative function (under a quadratic loss function). This is a minimax estimator provided $\theta \le g(z) \le D \le 1/(2^{n+2} - 1)$, where D is a constant and subject to some additional, rather cumbersome, conditions on $g(\cdot)$.

Uniform distributions thus form a location-scale family admitting a two-dimensional sufficient statistic for their two parameters. However, the best equivariant estimator (of the form above) can be shown to be inadmissible. Rukhin, Kuo, and Dey (1990) provide an explicit expression for a $g(z)$ that yields a minimax estimator:

$$g(z) = \begin{cases} D & \text{for } -0.5 \le z \le 0.5, \\ 0 & \text{elsewhere.} \end{cases} \tag{26.33}$$

Fan (1991) notes that the statistics

$$T_k = \frac{Y_{k+1}' + Y_{k+2}' + \cdots + Y_{n-k}'}{n - 2k} \tag{26.34a}$$

and

$$H_k = \frac{Y_1' + Y_2' + \cdots + Y_k' + Y_{n-k+1}' + Y_{n-k+2}' + \cdots + Y_n'}{2k},$$

$$k = 1, \ldots, \left[\frac{n}{2}\right], \tag{26.34b}$$

are each unbiased estimators of the population mean $(\frac{1}{2}\theta)$ for the uniform distribution over the interval $(0, \theta)$, and, for even n $(= 2s)$,

$$\mathrm{var}(H_1) < \mathrm{var}(H_2) < \cdots < \mathrm{var}(H_s) < \mathrm{var}(\bar{X}) < \mathrm{var}(T_1) < \mathrm{var}(T_2)$$
$$< \cdots < \mathrm{var}(T_{s-1}).$$

Móri (1983) [improving on an approach developed by Vincze (1979)] showed that for any unbiased estimator $\tilde{a}$ of a based on random samples of size n from distribution (26.1),

$$\liminf_{c\to\infty} \frac{1}{2c}\int_{-c}^{c} \text{var}(\tilde{a}|a)\,da \geq \frac{2h^2}{(n+1)(n+2)}.$$

In this way a lower bound is attained for "average variance" of a', using the estimator

$$a^* = \tfrac{1}{2}(Y_1' + Y_n') \qquad [\text{cf. } (26.24)]. \tag{26.35a}$$

Nikulin (1991) observed that a* and

$$h^* = \tfrac{1}{2}(n-1)^{-1}(n+1)(Y_n' - Y_1') \tag{26.35b}$$

are minimum variance unbiased estimators of a and h, respectively. We have

$$\text{var}(a^*) = 2(n+1)^{-1}(n+2)^{-1}h^2, \tag{26.36a}$$

$$\text{var}(h^*) = 2(n-1)^{-1}(n+2)^{-1}h^2, \tag{26.36b}$$

and

$$\text{cov}(a^*, h^*) = 0. \tag{26.36c}$$

Joó and Szabó (1992) present a discussion of the accuracy of a* for general symmetrical distributions. If a is known, we use the statistic

$$Z = \min(|Y_1' - a|, |Y_n' - a|), \tag{26.37}$$

which has the pdf

$$p_Z(z) = nh^{-n}z^{n-1}, \qquad 0 < z.$$

The minimum variance unbiased estimator of h is

$$h^{**} = (1 + n^{-1})Z. \tag{26.38}$$

Note that

$$\text{var}(h^{**}) = n^{-1}(n+2)^{-1}h^2, \tag{26.39}$$

which is less than half of var(h^*).

Eltessi and Pal (1992) investigate estimation of the minimum and the maximum of the scale parameters θ_1, θ_2 of two uniform $(0, \theta_j)$ $(j = 1, 2)$ distributions. Let $X_{j1}, \ldots, X_{jn}$ $(j = 1, 2)$ represent values from independent random samples of size n from uniform $(0, \theta_j)$, and define $\theta_{(L)} = \min(\theta_1, \theta_2)$, $\theta_{(U)} = \max(\theta_1, \theta_2)$. The statistic $X'_{jn} = \max(X_{j1}, \cdots, X_{jn})$ has the pdf

$$p_{X'_{jn}}(x) = \left(\frac{n}{\theta_j}\right)\left(\frac{x}{\theta_j}\right)^{n-1}, \qquad 0 \le x \le \theta_j,\ j = 1, 2. \tag{26.40}$$

Let $Z_{(L)} = \min(X'_{1n}, X'_{2n})$ and $Z_{(U)} = \max(X'_{1n}, X'_{2n})$. There are no unbiased estimators of $\theta_{(L)}$ and $\theta_{(U)}$ that are functions of only $Z_{(L)}$ and $Z_{(U)}$ [unless $\theta_1 = \theta_2$ $(= \theta_{(L)} = \theta_{(U)})$]. The estimators

$$\tilde{\theta}_{(\phi)} = \frac{2(n+1)(2n+1)}{n\{4n + 2g(\phi) - 1\}} Z_{(\phi)}, \qquad (\phi = (L, U)), \tag{26.41}$$

with $g(L) = 1$, $g(U) = 2$, are "minimax"-biased estimators of $\theta_{(L)}, \theta_{(U)}$, respectively, that minimize the maximum values of the proportional bias:

$$\left| E\left[\frac{\tilde{\theta}_{(\phi)}}{\theta_{(\phi)}} - 1\right]\right| \quad \text{over} \quad 0 \le \frac{\theta_{(L)}}{\theta_{(U)}} \le 1, \text{ where as above } \phi = (L, U).$$

Estimators minimizing the maximum risk are

$$\frac{2(n+2)}{2n+1} Z_{(\phi)}, \qquad (\phi = (L, U)), \tag{26.42}$$

which appear to perform better than the other estimators for $n \le 5$ and $\theta_1/\theta_2 = 0.1(0.1)0.9$.

7 ESTIMATION USING ORDER STATISTICS—CENSORED SAMPLES

The variances and covariances of ordered variables $Y'_1 \le Y'_2 \le \cdots \le Y'_n$ corresponding to a random sample of size n from (26.1) are given in (26.11c). These formulas make it possible to obtain best linear unbiased estimators of a and h. Such estimators were discussed by Lloyd (1952), Sarhan (1955), and Sarhan and Greenberg (1959). Some of their results are summarized below.

If the smallest r_1 and largest r_2 values (out of a random sample of total size n) are omitted, the best linear unbiased estimator of a is

$$\hat{a}^* = \tfrac{1}{2}[(n - 2r_2 - 1)\ (\text{Least observed value}) + (n - 2r_1 - 1)\ (\text{Greatest observed value})] \times (n - r_1 - r_2 - 1)^{-1}. \tag{26.43a}$$

The best linear unbiased estimator of h is

$$\hat{h}^* = \tfrac{1}{2}(n+1)\,[(\text{Greatest observed value}) - (\text{Least observed value})] \times (n - r_1 - r_2 - 1)^{-1}. \tag{26.43b}$$

The variances of these estimators are

$$\operatorname{var}(\hat{a}^*) = h^2[(r_1+1)(n-2r_2-1)+(r_2+1)(n-2r_1-1)] \times (n+1)^{-1}(n+2)^{-1}(n-r_1-r_2-1)^{-1}, \tag{26.44a}$$

$$\operatorname{var}(\hat{h}^*) = h^2(r_1+r_2+2)(n+2)^{-1}(n-r_1-r_2-1)^{-1}. \tag{26.44b}$$

The correlation between $\hat{a}^*$ and $\hat{h}^*$ is

$$(r_2 - r_1)\left[\frac{n+1}{(r_1+r_2+2)\{(r_1+1)(n-2r_2-1)+(r_2+1)(n-2r_1-1)\}}\right]^{1/2}. \tag{26.45}$$

It will be noted that we again use only the least and greatest of all the observed values (just as when all the sample values are available).

By putting $r_1 = 0$, the special case of censoring only the largest r_2 values is obtained. In this case

$$\hat{a}^* = \tfrac{1}{2}[(n-2r_2-1)(\text{Least observed value}) + (n-1)(\text{Greatest observed value})] \times (n-r_2-1)^{-1} \tag{26.46a}$$

and

$$\hat{h}^* = \tfrac{1}{2}(n+1)[(\text{Greatest observed value}) - (\text{Least observed value})] \times (n-r_2-1)^{-1}. \tag{26.46b}$$

The corresponding estimator of the lower limit of the range of variation is

$$\hat{a}^* - \hat{h}^* = \tfrac{1}{2}[2(n-r_2)(\text{Least obsreved value}) - 2(\text{Greatest observed value})](n-r_2-1)^{-1}.$$

Putting $r_2 = 0$, the special case of censoring only the r_1 smallest values is obtained.

8 TABLES OF RANDOM NUMBERS

A set of numbers 0 to 9 chosen independently of each other, with each number equally likely to be any one of the ten digits 0–9, is known as a table of random numbers. Although the distribution corresponding to the individual recorded number is a discrete rectangular distribution (Chapter 6, Section 10.1), good approximations to samples from (continuous) rectangular distributions are obtainable by combining several integers together and applying an appropriate linear transformation. For example, taking groups of four numbers, adjoining a 5 at the right-hand end, and dividing by 100,000 gives a good approximation to random samples from a standard uniform distribution (over the interval 0 to 1). The best-known published sets of random numbers are (in chronological order):

Tippett (1927)
Kendall and Babington Smith (1938, 1940)
Rand Corporation (1955)
Clark (1966).

At present, tables of random numbers have been largely supplanted by random number generators [see, e.g., Devroye (1986) and Section 11].

By combining uniformly distributed random variables in various ways, a number of other distributions can be built up. For example, if Y has a standard rectangular distribution (over the interval 0 to 1) then $-2\log Y$ is distributed as χ^2 with two degrees of freedom (see the beginning of Section 26.9). Additional examples are given in Marsaglia (1961).

Dharmadhikari [cited in Troutt (1991)] notes that given the Laplace density (Chapter 24)

$$p_X(x) = \tfrac{1}{2}\exp(-|x|), \qquad -\infty < x < \infty,$$

the ordinate $p_X(X)$, considered as a random variable (the so-called vertical density function) is uniform. The same is true for the standard exponential density $p_X(x) = e^{-x}$ $(0 < x)$. Similarly for the uncorrelated bivariate normal density [Troutt (1991)]

$$p_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(x_1^2 + x_2^2)\right\},$$

the ordinate $p_{X_1,X_2}(X_1,X_2)$, considered as a random variable, is uniform $[0, 1/(2\pi)]$ providing an intuitive interpretation of the Box-Muller method for generating normal variables (Chapter 13, Section 9). (At the time of writing, extension to a general multivariate setting remains an unproven conjecture.)

9 RELATED DISTRIBUTIONS

As we mentioned in Chapter 25, the rectangular distribution is a special form of beta distribution. If X is rectangularly distributed over (0, 1), according to (26.4), then $Z = -\log X$ has the exponential distribution (Chapter 19)

$$p_Z(z) = e^{-z}, \qquad z > 0, \tag{26.47}$$

[and conversely, if Z^* has a standard exponential distribution, $X = \exp(-Z^*)$ has a standard uniform distribution]. That is, Z is distributed as χ^2 with two degrees of freedom (Chapter 18). This relationship is used in the construction of certain methods for the combination of tests. Fisher (1932) proposed that values of independent random variables $Z_1, Z_2, \cdots, Z_k$ obtained from k independent tests be combined by checking the value of $\sum_{i=1}^{k} Z_i$ against a χ^2 distribution with 2k degrees of freedom. [See also Quesenberry (1986).] The distribution of $\prod_{i=1}^{k} X_i$ is easily derived from the fact that its logarithm is distributed as $-\frac{1}{2}\chi^2_{2k}$.

The distribution of $S_n = \sum_{i=1}^{n} X_i$, where $X_1, X_2, \cdots, X_n$ are independent $U(0,1)$ random variables, can be derived successively using standard convolution formulas. The result is the Irwin-Hall distribution [Hall (1932); Irwin (1932)]:

$$p_{S_n}(s) = \begin{cases} \dfrac{1}{(n-1)!} \displaystyle\sum_{j=0}^{k} (-1)^j \binom{n}{j} (s-j)^{n-1}, & k \le s \le k+1,\ 0 \le k \le n-1, \\ 0, & \text{elsewhere.} \end{cases} \tag{26.48}$$

For the more general case of $U(0, a)$ the pdf of S_n is

$$p_{S_n}(s) = \begin{cases} \dfrac{1}{(n-1)!a^n} \displaystyle\sum_{j=0}^{k} (-1)^j \binom{n}{j} (s-aj)^{n-1}, & k \le sa^{-1} < k+1,\ 0 \le k \le n-1, \\ 0, & \text{elsewhere.} \end{cases} \tag{26.49}$$

The rectangular mean distribution is the distribution of the sum (S_n) as in (26.48), divided by the sample size n, namely the arithmetic mean T. It has the pdf

$$p_T(t) = \frac{n^n}{(n-1)!} \sum_{j=0}^{[tn]} (-1)^j \binom{n}{j} \left(t - \frac{j}{n}\right)^{n-1}, \qquad 0 \le t \le 1. \quad (26.50)$$

This is also known as ***Bates's distribution*** [Bates (1955)]. This distribution is sometimes confused with the Irwin-Hall distribution. The remarkable history of investigation of these distributions [as described by Seal (1950)] has been mentioned in Section **26.3.**

When $n = 2$, the distribution of $\overline{X}_2 = \frac{1}{2}S_2$ is a symmetrical ***triangular distribution***:

$$p_{\overline{X}_2}(x) = \begin{cases} \dfrac{x - a + h}{h^2}, & a - h \le x \le a, \\ \dfrac{a + h - x}{h^2}, & a \le x \le a + h, \end{cases} \qquad (26.51)$$

or equivalently

$$p_{\overline{X}_2}(x) = \frac{h - |x - a|}{h^2}, \qquad a - h \le x \le a + h. \qquad (26.51)'$$

The standard triangular distribution [see Ayyangar (1941)] is represented (possibly after linear transformation) by a probability density function of form

$$p_X(x) = \begin{cases} \dfrac{2x}{H}, & 0 \le x \le H, \\ \dfrac{2(1-x)}{1-H}, & H \le x \le 1. \end{cases} \qquad (26.52)$$

The graph of $p_X(x)$ sketched in Figure **26.3** indicates why the name ***triangular*** is given to these distributions. If $H = \frac{1}{2}$, the distribution is symmetrical.

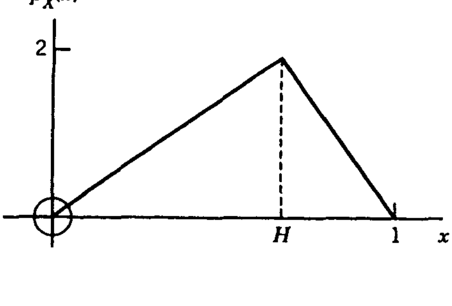


Figure 26.3 Triangular density function.

Symmetrical triangular distributions have been called ***tine distributions*** [Schmidt (1934)].

The rth moment about H is

$$ {}_H\mu_r(X) = E\left[(X-H)^r\right] = \frac{2\left[(-1)^r H^{r+1} + (1-H)^{r+1}\right]}{(r+1)(r+2)}. \quad (26.53)$$

The expected value is

$$H + {}_H\mu_1(X) = \tfrac{1}{3}(1+H), \quad (26.54)$$

and the variance is

$$ {}_H\mu_2(X) - \left[{}_H\mu_1(X)\right]^2 = \tfrac{1}{18}(1-H+H^2). \quad (26.55)$$

The median is at

$$\begin{cases} 1 - \sqrt{\tfrac{1}{2}\max(H, 1-H)}, & H \le \tfrac{1}{2}, \\ \sqrt{\tfrac{1}{2}H} & H \ge \tfrac{1}{2}. \end{cases}$$

The mean deviation is

$$\begin{cases} 1 \quad -H)^{-1}(2-H)^3 & \text{if } H < \tfrac{1}{2}, \\ \tfrac{2}{81}H^{-1}(1+H)^3 & \text{if } H > \tfrac{1}{2}. \end{cases} \quad (26.56)$$

The ratio (mean deviation)/(standard deviation) has the following values:

H	0.5	0.6	0.7	0.8	0.9
Ratio	0.816	0.820	0.827	0.833	0.837

When each X_j can have different values of a_j and h_j, the distribution of the sum S_n is much more complicated. Tach (1958) gives five decimal place tables of the cumulative distribution function of S_n for n = 2, 3, 4, with a, = 0 for all j, and various h, (subject to $\sum_{j=1}^{n} h_j = 1$).

Barrow and Smith (1979) provided a succinct formula for the cdf of a ***linear*** function of independent uniform (0, 1) random variables. A detailed investigation of the probability distribution of roundoff errors (essentially a linear combination of independent uniform $(-\frac{1}{2}, \frac{1}{2})$ random variables; see Section 26.2) was provided by Mitra and Banerjee (1971). Their derivation was based on the formula for the volume of the intersection of a "half-space" with a hypercube in R^n, involving concepts related to the theory of splines.

Barrow and Smith's formula is based on the relation:

$$\Pr\left[\sum_{i=1}^{n} \alpha_i X_i \le w\right] = \frac{1}{\alpha_1}\int_0^1 \cdots \int_0^1$$
$$\times\left[\left(w - \sum_{i=2}^{n} \alpha_i x_i\right)_+^n - \left(w - \alpha_1 - \sum_{i=2}^{n} \alpha_i x_i\right)_+^n\right] dx_2 \cdots dx_n \qquad (26.57a)$$

where

$$(x)_+^k = \begin{cases} x^k & \text{if } x > 0, \\ 0 & \text{if } x < 0, \end{cases}$$

and w is a real number, leading to

$$\Pr\left[\sum_{i=1}^{n} \alpha_i X_i \le w\right] = \frac{\sum_{\mathbf{v} \in C}(\operatorname{sgn} \mathbf{v}) \cdot (w - \boldsymbol{\alpha} \cdot \mathbf{v})_+^n}{n!(\prod_{i=1}^{n} \alpha_i)}, \qquad (26.57b)$$

where C is the "cube" $\{\mathbf{x} \in R^n; 0 \le x_i \le 1 \text{ for } i = 1, \dots, n\}$; the summation is over the 2^n vertices of C and $\operatorname{sgn} \mathbf{v} \equiv (-1)^m$, where $m = \sum_{i=1}^{n} v_i$.

Mitra and Banerjee's (1971) formula [earlier results are due to Lowan and Laderman (1939)] for the cdf of

$$\rho_n = \sum_{s=0}^{n} (-1)^s \omega_s R_s,$$

where

$$p_{R_s}(x) = \begin{cases} 1 & \text{for } -\frac{1}{2} \le x \le \frac{1}{2}, \\ 0 & \text{elsewhere}, \end{cases} \qquad (26.58)$$

[note that the transformed variables $T_s = (-1)^s R_s + \frac{1}{2}$ are $U(0,1)$] is

$$\Pr[\rho_n \le x] = \int_D \cdots \int dt_0 \, dt_1 \cdots dt_n \qquad (26.59a)$$

with the region D defined by $0 \le t_i \le 1$ and

$$\omega_0 t_0 + \omega_1 t_1 + \cdots + \omega_n t_n \le x + 2^{n-1}.$$

The integral is similar to the Dirichlet integral, for which the domain of integration is $(0 \le t_s\}$ $(s = 0, 1, \ldots, n)$ and $t_0 + t_1 + \cdots + t_n \le 1$. The authors utilize the Dirichlet integrals and obtain

$$\Pr[\rho_n \le x] = \frac{\Sigma_s(-1)^{s_0+s_1+\cdots+s_n}\left(x + \frac{1}{2}\Sigma_{i=0}^n \omega_i - \Omega\right)^{n+1}}{(n+1)!\,\omega_0\omega_1\cdots\omega_n}$$

$$\text{for } -\frac{1}{2}\sum_{i=1}^n \omega_i \le x \le \frac{1}{2}\sum_{i=1}^n \omega_i, \qquad (26.59b)$$

where

$$\Omega = \sum_{i=1}^n s_i\omega_i \le x + \frac{1}{2}\sum_{i=0}^n \omega_i,$$

and the summation is over $s_i = 0, 1$ for all i.

The even moments are

$$E[\rho_n^{2\nu}] = 2^{-2\nu}\sum \frac{(2\nu)!\prod_{i=0}^n \omega_i^{h_i}}{\prod_{i=0}^n (h_i+1)!}, \qquad (26.60)$$

where summation is over all even integral values of $h_0, h_1, \ldots, h_n$ such that $\Sigma_{i=0}^n h_i = 2\nu$; all odd order moments vanish. The variance is $\Sigma_{i=0}^n \omega_i^2/12$, and the kurtosis is

$$\beta_2 - 3 = \frac{\kappa_4}{\kappa_2^2} = -\frac{6}{5}\left(\sum_{i=1}^n \omega_i^4\right)\left(\sum_{i=1}^n \omega_i^2\right)^{-2},$$

which is always negative, indicating that the distribution is always *platykurtic*. The density is more flattened than the corresponding normal distribution near the median. For small values of x the asymptotic pdf is

$$p_{\rho_n}(x) \approx \left(6\pi^{-1}\sum_{i=0}^n \omega_i^2\right)^{-1/2} \exp\left\{-\left(6\sum_{i=0}^n \omega_i^2\right)^{-1} x^2\right\}. \qquad (26.61)$$

[Mitra and **Banerjee's** formula is of course equivalent to that of Barrow and Smith **(1979)**, derived at the request of H. O. **Hartley**.]

The joint distribution of the differences of successive order statistics of standard uniform variables is a Dirichlet distribution (Chapter 40). If $Y_1' \le Y_2' \le \cdots \le Y_n'$ are defined as at the beginning of Section 7, and $V_i = Y_i' - Y_{i-1}'$ [$i = 1, \ldots, (n+1)$], $Y_0' = a - h$; $Y_{n+1}' = a + h$, then

$$p_{V_1, \ldots, V_n}(v_1, \cdots, v_n) = n!(2h)^{-n}, \qquad 0 \le v_i, \ \sum_{i=1}^{n} v_i \le 2h. \quad (26.62)$$

Putting $h = \frac{1}{2}$ (corresponding to a range of 1) we have

$$p_{V_1, \ldots, V_n}(v_1, \cdots, v_n) = n!, \qquad 0 \le v_i, \ \sum_{i=1}^{n} v_i \le 1 \ [\text{cf. (26.10a) and (26.10b)}]. \quad (26.63)$$

More insight into the nature of this joint distribution is gained by noting that the same distribution (26.63) would be obtained if

$$V_i = \frac{W_i}{\sum_{j=1}^{n+1} W_j}, \qquad i = 1, 2, \ldots, (n+1),$$

where $W_1, W_2, \cdots, W_{n+1}$ are mutually independent random variables each distributed as χ^2 with two degrees of freedom. From this point of view it is clear that the ratio $(Y_{n-s+1}' - Y_s')/(Y_{n-r+1}' - Y_r')$, with $s > r$, has a standard beta distribution with parameters $(n - 2s + 1), 2(s - r)$ (Chapter 25, Section 2). Criteria of this kind were suggested by David and Johnson (1956) as tests for kurtosis when the probability integral transformation can be used. As another example demonstrating that the range $(Y_n' - Y_1')$ is equal to $\sum_{i=2}^{n} V_i$, it is clear that range has a beta distribution with parameters $(n-1), 2$, as can be confirmed from (26.22); see also (26.27).

The distribution of the ratio (U) of ranges calculated from independent random samples of sizes n', n'' from distributions (26.1) with the same value of h has been studied by Rider (1951, 1963). The probability density function of this ratio (sample size n' in numerator) is

$$p_U(u) = \begin{cases} C\left[(n' + n'')u^{n'-2} - (n' + n'' - 2)u^{n'-1}\right], & 0 \le u \le 1, \\ C\left[(n' + n'')u^{-n''} - (n' + n'' - 2)u^{-n''-1}\right], & 1 \le u, \end{cases} \quad (26.64)$$

with $C = n'(n'-1)n''(n''-1)/[(n'+n'')(n'+n''-1)(n'+n''-2)]$. Tables of upper 10%, 5%, and 1% points are given in Rider (1951); more extensive tables are given in Rider (1963).

The distribution of the ratio (V) of maximum values when both random samples come from $U(0, 2h)$ **distribution(s)** was studied by Murty (1955). He found that

$$p_V(v) = \begin{cases} \dfrac{n'n''}{n' + n''} v^{n'-1} & \text{for } 0 \le v \le 1, \\ \dfrac{n'n''}{n' + n''} v^{-n''-1} & \text{for } v \ge 1. \end{cases} \tag{26.65}$$

Murty gives tables of the upper 5% point of the distribution of $\max(V, V^{-1})$ for $n', n'' = 2(1)20$, (n' = sample size with ***greater*** maximum value).

These last two criteria might be used in testing for identity of two rectangular distributions with respect to changes in range of variation (in the second case, with known initial point). A criterion introduced by Hyrenius (1953) uses "cross-ranges." If we denote by L', L'' the smallest observations in the two samples, and by U', U'' the largest (with the samples chosen so that $L' \le L''$), then the cross-ranges are $U'' - L'$, $U' - L''$. For $V = (U'' - L')/(U' - L'')$, Hyrenius obtained the probability density function

$$p_V(v) = \begin{cases} \dfrac{(n'-1)n''}{n' + n'' - 1} {}^{n''-1} & \text{for } 0 \le v \le 1, \\ \dfrac{(n'-1)n''}{n' + n'' - 1} {}^{-n'} & \text{for } v \ge 1. \end{cases} \tag{26.66}$$

He also considered $T = (L'' - L')/(U' - L') = V +$ (ratio of ranges of the two samples).

9.1 Mixtures of two uniform distributions

Mixtures of uniform distributions have some importance in data analysis. They provide a tool for constructing histograms from a data set without attempting to estimate the underlying distribution. Gupta and Miyawaki (1978) studied estimation problems in the mixture of two uniform distributions with a pdf of the form

$$p_Y(y) = pf_1(y, \beta) + (1 - p)f_2(y, \beta), \tag{26.67}$$

where

$$f_1(y, \beta) = \begin{cases} \dfrac{1}{\beta} & \text{for } 0 < y < \beta, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_2(y, \beta) = \begin{cases} \dfrac{1}{1-\beta} & \text{for } \beta < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The mixture is identifiable unless $p = \beta$. The kth crude moment of distribution **(26.67)** is

$$\mu'_k = p\frac{\beta^k}{k+1} + q\frac{1-\beta^k}{1-\beta}\frac{1}{k+1}, \qquad q = 1-p. \tag{26.68}$$

Given observed values of $Y_1, \ldots, Y_n$-**independent** random variables distributed with the pdf **(26.67)** with β known, one can use the moment estimator **(ME)** of p,

$$\tilde{p} = 1 + \beta - 2\bar{Y}, \tag{26.69}$$

where $\bar{Y} = n^{-1}\sum_{i=1}^n Y_i$, which is unbiased, with variance

$$\sigma^2 = \frac{4}{n}\left[\frac{\beta^2}{12} + \frac{q}{6}(8 - 6q - \beta)\right], \tag{26.70}$$

while the maximum likelihood estimator is

$$\hat{p} = \frac{\text{Number of } Y\text{'s} \le \beta}{n} \tag{26.71}$$

which is **UMVUE** and consistent.

If p is known, a consistent and unbiased moment estimator of β is

$$\tilde{\beta} = 2\bar{Y} - q. \tag{26.72}$$

The maximum likelihood estimator **(MLE)** of β is in the interval $[Y'_r, Y'_{r+1}]$, where $Y'_r < p < Y'_{r+1}$. In fact

$$\hat{\beta} = \begin{cases} Y'_r & \text{for } \left(\dfrac{Y'_{r+1}}{Y'_r}\right)^r > \left(\dfrac{1 - Y'_r}{1 - Y'_{r+1}}\right)^{n-r}, \\ Y'_{r+1} & \text{for } \left(\dfrac{Y'_{r+1}}{Y'_r}\right)^r < \left(\dfrac{1 - Y'_r}{1 - Y'_{r+1}}\right)^{n-r}, \end{cases} \tag{26.73}$$

where r is determined by the inequalities $Y_r' \leq p \leq Y_{r+1}'$. No comparison of the properties of $\tilde{\beta}$ and $\hat{\beta}$ is available in the literature, as far as we know.

In the general case where both p and β are unknown, there are moment estimators of the form

$$\begin{pmatrix} \tilde{q} \\ \tilde{\beta} \end{pmatrix} = \begin{bmatrix} \dfrac{4M_1^2 - 3M_2}{2M_1 - 1} \\ \dfrac{3M_2 - 2M_1}{2M_1 - 1} \end{bmatrix}, \tag{26.74}$$

where $M_r = n^{-1}\Sigma_{i=1}^{n} Y_i'^r$. These estimators are asymptotically unbiased and consistent. Gupta and Miyawaki (1978) provide expressions for the asymptotic variance-covariance matrix of the joint distribution of $\begin{pmatrix} \tilde{q} \\ \tilde{\beta} \end{pmatrix}$, which is asymptotically normal.

To calculate the maximum likelihood estimators $\hat{p}, \hat{\beta}$, the following iterative procedure has been suggested. Calculate the moment estimators $\tilde{p}$ [$= (1 - \tilde{q})$] and $\tilde{\beta}$ as indicated above. Next calculate $\hat{\beta}$ from (26.73). Then turn to $\hat{p}$ = (number of Y's $< \hat{\beta})/n$ and go back to (26.731, repeating until the values of $\hat{p}$ and $\hat{\beta}$ stabilize. Again, there are no comparative investigations of the MLE and ME for the general case in the literature, so far as we know.

Roy, Roy, and Ali (1993) have introduced the binomial mixture of uniform $(0, a)$ distributions with density given by

$$p_X(x|n, p, a) = \sum_{r=0}^{n} \binom{n}{r} p^r(1-p)^{n-r}(r+1)x^r/a^{r+1}, \qquad 0 < x < a.$$

The kth moment about zero of X is given by

$$E[X^k] = a^k \sum_{r=0}^{n} \binom{n}{r} p^r(1-p)^{n-r}(r+1)/(r+k+1).$$

9.2 Other Related Distributions

Among other distributions derived (at least in part) from the rectangular distribution there may be mentioned five:

1. The Student's t (Chapter 28) when the variables $X_1, \cdots, X_n$ from which the t is calculated are independent and have a common rectangular distribution. Rider (1929) showed that for samples of size $n = 2$, the probability density function is of the form $\frac{1}{2}(1 + |t|)^{-2}$. Perlo (1933) derived the distribution for samples of size $n = 3$, and Siddiqui (1964)

obtained a general formal result for any size of sample and noted some inequalities for the cumulative distribution function.

2. Various distributions arising from the construction of tests of adherence to a specified distribution, using the probability integral transformation. **Pearson (1938)** pointed out that if Y has distribution **(26.1)** with $a = h$, so do $2h - Y$, $2|Y - h|$, and so on. Durbin (1961) showed that if the differences $V_1, V_2, \cdots, V_{n+1}$ ($V_j = Y_j' - Y_{j-1}'$, with $Y_0' = 0$, $Y_{n+1}' = 1$) are arranged in ascending order of magnitude $V_1' \le V_2' \le \dots \le V_{n+1}'$, the joint probability density function of $V_1', \dots, V_n'$ is (for $h = \frac{1}{2}$):

$$p_{V_1', \cdots, V_n'}(y_1, y_2, \cdots, y_n) = (n+1)!n!, \qquad 0 \le y_1 \le \cdots \le y_n, \ \sum_{i=1}^{n} y_i \le 1. \tag{26.75}$$

Also, if

$$G_j = (n + 2 - j)(V_j' - V_{j-1}')$$

[Sukhatme's (1937) transformation], then

$$p_{G_1, \cdots, G_n}(g_1, g_2, \cdots, g_n) = n!, \qquad g_j \ge 0, \ \sum_{j=1}^{n} g_j \le 1. \tag{26.76}$$

Hence the quantities

$$W_r' = \sum_{j=1}^{r} G_j, \qquad r = 1, \dots, n, \tag{26.77}$$

have the same joint distribution as the original ordered variables Y_r', so any function of the W'**'s** has the same distribution as the corresponding function of the Y_r'**'s**. Durbin gave a considerable number of references to other work on "random division of an interval," namely on uniform distributions. More recent works on random division of an interval are due to Chen, Lin, and Zame (1981), Chen, Goodman, and Zame (1984), Van Assche (1987), and Johnson and Kotz (1990); see also Chapter **25** for further discussions.

3. The distribution of the ratio of a rectangular variable Y to an independent normal variable Z was described by Broadbent **(1954).** Calculations are facilitated by noting that, for example, with $a = 0$ in (26.1),

$$\Pr\left[\frac{Y}{Z} \le K(>0) | Z > 0\right] = \Pr[Y \le KZ | Z > 0] = \frac{\int_0^\infty p_Z(z) F_Y(Kz)\, dz}{\int_0^\infty p_Z(z)\, dz}.$$

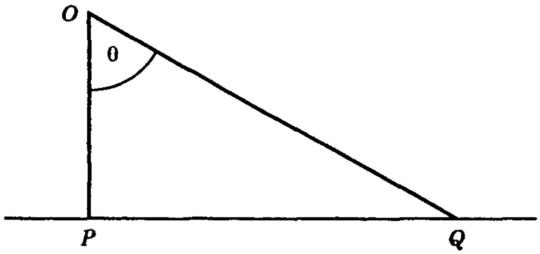


Figure 26.4 Connection between a uniform angular distribution around a semicircle and a Cauchy distribution

4. There is a connection between a uniform angular distribution around a semicircle and a Cauchy distribution (Chapter 16) on a line. If, as shown in Figure **26.4,** θ has the uniform probability density

$$p_\theta(t) = \pi^{-1}, \qquad -\frac{\pi}{2} \le t \le \frac{\pi}{2}, \tag{26.79}$$

then the probability density function of $X = PQ$ (OP is perpendicular to PQ) is

$$p_X(x) = \pi^{-1}\frac{d}{dx}\tan^{-1}\frac{x}{|OP|} = \frac{1}{\pi|OP|}\,\frac{1}{\left[1 + (x/|OP|)^2\right]}. \tag{26.80}$$

Hence X has a Cauchy distribution with scale parameter $|OP|$.

Cowan (1980) noted that if W and Z are independent random variables distributed as gamma $(a, 1)$ and beta $(a, a - a)$, respectively, then WZ has a gamma $(a, 1)$ distribution. Taking $a = 2$ and $a = 1$ (so that WZ has a standard exponential distribution, and Z has a standard uniform distribution) and remembering that

a. since W has a gamma $(2, 1)$ distribution, we have $W = W_1 + W_2$, where W_1 and W_2 are independent standard exponential variables, so that,
b. $X_i = \exp(-W_i)$ $(i = 1, 2)$ are independent, each being a standard uniform variable as is $\exp(-WZ) = \exp\{-(W_1 + W_2)Z\} = (X_1X_2)^Z$, we obtain the following remarkable result: If X_1, X_2, and Z are mutually independent standard uniform variables, then $(X_1X_2)^Z$ also has a standard uniform distribution.

Alternative proofs were given by Zhao-Guo and Hong-Zhi (1980) who use characteristic functions, Scott (1980) who uses moments, and Westcott (1980) who provides an interpretation using Poisson processes. Westcott also obtained the following generalization: If

$$Y_n = \left(\prod_{i=1}^{n} X_i\right)^z,$$

where $X_1, X_2, \cdots, X_n$ are mutually independent standard uniform random variables, then the pdf of $Y_n = -\log W_n$ is

$$p_{Y_n}(y) = \frac{1}{(n-1)!}\int_y^\infty t^{n-2}e^{-t}\,dt = (n-1)^{-1}\left\{1 - \frac{\Gamma_y(n-1)}{\Gamma(n-1)}\right\}, \tag{26.81}$$

where $\Gamma_y(n-1)/\Gamma(n-1)$ is the incomplete gamma function ratio (see Chapter *1*, Section A5).

5. Proctor (1987) introduced "generalized uniform" distributions, with cdfs of

$$F_X(x) = 1 - \{1 - k(x-a)^c\}^h, \qquad k, c, h > 0;\ a\ r\ x\ r\ a + k^{-1/c}. \tag{26.82}$$

These distributions are a counterpart to Burr type XII distributions (Chapter *12,* Section 4.5) which have cdfs of form

$$F_X(x) = 1 - (1 + x^c)^{-k}, \qquad 0 < x. \tag{26.83}$$

They cover parts of the (β_1, β_2) plane not covered by the latter.

10 APPLICATIONS

A frequent use of uniform distributions is in an explanatory capacity to clarify difficult theory in the classroom and in the literature. Chu *(1957)* and Leone (1961) utilized uniform distributions in connection with sample quasi-ranges. Anderson (1942) provided an early example of their use in stratified sampling. Levene (1952) among many others applied them to determine power functions of tests of randomness. Naus (1966) applied uniform distributions in a power comparison of tests of nonrandom clustering. There are also numerous applications in nonparametric tests, such as the Kolmogorov-Smirnov type.

Irwin-Hall and Bates distributions (see Section *26.9)* have found applications in accident proneness models [e.g., Haight (1965)]. Some applications in physics are presented in Feller (1966).

10.1 Corrections for Grouping

The uses of uniform distributions to represent the distribution of roundoff errors, and in connection with the probability integral transformation, have already been mentioned in Sections 26.2 and 26.9. These distributions also play a central role in the derivation of Sheppard's corrections (1907), which adjust the values of sample moments to allow for (on the average) effects of grouping.

Suppose that the probability density function of X is $p_X(x)$. If the observed value is not X but the nearest value ($\tilde{X}$) to X in the set of values $\{\alpha + jh\}$, where j can take any (positive, negative, or zero) integer value, then

$$\Pr[\tilde{X} = \alpha + jh] = \int_{\alpha+(j-\frac{1}{2})h}^{\alpha+(j+\frac{1}{2})h} p_X(x)\,dx. \tag{26.84}$$

We now seek to find an "average" relation between the cumulants of X and those of $\tilde{X}$. If we assume that $(\tilde{X} - X)$ has a uniform distribution over $(-\frac{1}{2}h, \frac{1}{2}h)$, then

$$\tilde{X} = X + Y,$$

where Y has a uniform distribution with a probability density function

$$p_Y(y) = h^{-1}, \qquad -\tfrac{1}{2}h \le y \le \tfrac{1}{2}h.$$

Also X and Y are mutually independent, so

$$\kappa_r(\tilde{X}) = \kappa_r(X) + \kappa_r(Y).$$

Using (26.7), we find that

$$\begin{aligned} \kappa_1(X) &= \kappa_1(\tilde{X}), \\ \kappa_2(X) &= \kappa_2(\tilde{X}) - \tfrac{1}{12}h^2, \\ \kappa_3(X) &= \kappa_3(\tilde{X}), \\ \kappa_4(X) &= \kappa_4(\tilde{X}) + \tfrac{1}{120}h^4. \end{aligned} \tag{26.85}$$

The last equation implies that $\mu_4(X) = \mu_4(\tilde{X}) - h^2\mu_2(\tilde{X})/2 + 7h^4/240$.

10.2 Life Testing

We give the distribution of a statistic based on the *r* smallest of *n* independent observations from a standard uniform distribution. In life-testing terminology, this statistic includes as special cases (1) the sum of the r earliest failure times, (2) the total observed life up to the rth failure, and (3) the sum of all n failure times.

The statistic is given by

$$T_{r,m}^{(n)} = t_1 + t_2 + \cdots + t_r + (m - r)t_r, \tag{26.86}$$

where t_i is the ith smallest of n independent observations and m is greater than $r - 1$ but is not necessarily an integer. For the case where $m = n$, this statistic can be interpreted as the total observed life in a life-testing experiment without replacement. The results presented below are due to Gupta and Sobel (1958).

The density function and the distribution function of $T_{r,m}^{(n)}$ are given by

$$p_{T_{r,m}^{(n)}}(t) = A_{r-1,m}^{(n,n)}(m - t) \tag{26.87a}$$

and

$$F_{T_{r,m}^{(n)}}(t) = 1 - \frac{1}{n+1} A_{r-1,m}^{(n,n+1)}(m - t), \tag{26.87b}$$

respectively, where $0 \le t \le m$ and

$$A_{r-1,m}^{(n,n)}(m - t) = \frac{n}{(r-1)!} \times \left\{ \frac{(m-t)^{n-1}}{m^{n-r+1}} - \binom{r-1}{1} \frac{(m-1-t)^{n-1}}{(m-1)^{n-r+1}} + \cdots \right\}.$$

From these expressions we can get as special cases the densities and cumulative distribution functions of case 1: $T_{r,r}^{(n)}$; case 2: $T_{r,n}^{(n)}$; and case 3: $T_{n,n}^{(n)}$.

The mean and the variance of $T_{r,m}^{(n)}$ are given by the formulas

$$E\left[T_{r,m}^{(n)}\right] = \frac{r(2m - r + 1)}{(2n + 1)}, \tag{26.88a}$$

$$\operatorname{var}\left[T_{r,m}^{(n)}\right] = \frac{r(n - r + 1)(2m - r + 1)^2}{4(n+1)^2(n+2)} + \frac{r(r+1)(r-1)}{12(n+1)(n+2)}. \tag{26.88b}$$

The distribution of the statistic $T_{r,m}^{(n)}$ is asymptotically normal for $r = \lambda n$, $m = \gamma n$ (γ and λ fixed with $0 < \lambda \le 1$ and $\lambda \le \gamma < \infty$), and $n \to \infty$. Other earlier applications for life testing are given in Epstein (1948).

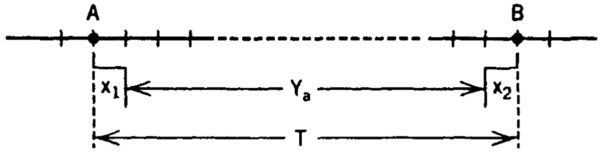


Figure 26.5 Representation of the distance from a vehicle *A* to the next vehicle B, ahead of it

10.3 Traffic Flow Applications

Allan (1966) applied the uniform distribution to form a model of the distribution of traffic along a straight road. The road is divided into intervals each of length h, and it is supposed that for each interval there are probabilities p of one vehicle being in the interval, q of no vehicles being there. (For the purpose of this model the possibility of two vehicles being in the same interval is neglected, and likewise are the lengths of the vehicles.) It is further supposed that given that a vehicle is in an interval, its position is uniformly distributed over the interval. From Figure 26.5 it can be seen that the distance from a vehicle A to the next vehicle B ahead of it is distributed as

$$L = hY + X_1 + X_2,$$

where Y represents the number of empty intervals between A and B and has the geometric distribution (Chapter 5, Section 2)

$$\Pr[Y = y] = q^y p, \qquad y = 0, 1, \dots,$$

and X_1, X_2 are independent random variables (also independent of Y) each having the distribution

$$p_X(x) = h^{-1}, \qquad 0 \le x \le h.$$

The distribution of $S = X_1 + X_2$ has density function

$$p_S(s) = h^{-2}(h - |s - h|), \qquad 0 \le s \le 2h \tag{26.89}$$

[cf. (26.51)']. The distribution of $T = hY + S$ has density function

$$p_T(t) = \begin{cases} ph^{-2}t, & 0 \le t \le h, \\ ph^{-2}\{q^{k-1}[(k+1)h - t] + q^k(t - kh)) \\ = pq^{k-1}h^{-2}\{(1 + kp)h - pt), & kh \le t \le (k+1)h, k \le 1 \end{cases} \tag{26.90}$$

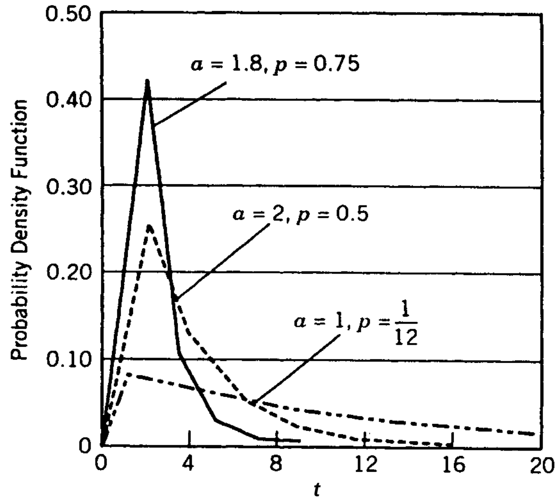


Figure 26.6 Allan's Binomial-Uniform (Headway) Distribution

Allan called this distribution a binomial-uniform distribution. It should not be confused with the binomial-beta distribution of Chapter 8 (Section 3.3). The graph of the density function (26.90) consists of a succession of straight lines. Figure 26.6, taken from Allan (1966), shows some examples.

Since Y, X_1, and X_2 are mutually independent, it is easy to find the moments of T. The rth cumulant is

$$\kappa_r(T) = h^r\kappa_r(Y) + 2\kappa_r(X_1),$$

whence we find

$$E[T] = hp^{-1}, \tag{26.91a}$$

$$\operatorname{var}(T) = (hp^{-1})^2(q + \tfrac{1}{6}p^2), \tag{26.91b}$$

$$\sqrt{\beta_1(T)} = q(1+q)(q + \tfrac{1}{6}p^2)^{-3/2}, \tag{26.91c}$$

$$\beta_2(T) = 3 + (6q^2 + qp^2 - \tfrac{1}{60}p^4)(q + \tfrac{1}{6}p^2)^{-2}. \tag{26.91d}$$

Allan (1966) also obtained the distribution of the sum of an independent identically distributed binomial-uniform variables and provided tables of the cumulative distribution function to four decimal places for p = 0.4(0.1)0.9, t/h = 0.5(0.5)25, n = 1(1)20. Applications to queueing theory—with waiting time uniformly distributed—were provided by Haight (1958, 1960).

10.4 Applications in Statistical Tests and Models

We have already noted the probability integral transformation (at the end of Section 26.2) and its use in formulation of procedures for combining results of significance tests [see (26.77) and beginning of Section 26.91. The use of uniform distributions in constructing models of distributions of roundoff error has been described in Section 26.2.

11 RANDOM NUMBER GENERATION

Routines for generating uniform random numbers play an important role in many Monte Carlo or simulation studies. Uniform random generators facilitate in principle simulation of pseudorandom numbers from any continuous distribution, using the inverse cumulative distribution function transformation.

There are many random number generators available in the literature. A common one is the "multiplicative congruential method" which has a generator of the form

$$x_i = cx_{i-1} \bmod(2^{31} - 1). \qquad (26.92)$$

The generator needs to be initialized (x_0) by selecting a "seed" or an initial value. Each x, is then scaled into the unit interval (0, 1). If the multiplier c is a primitive root modulo $2^{31} - 1$ (which is a prime), then the generator in (26.92) will have maximal period of $2^{31} - 2$.

The lattice structure induced by the congruential generator above can be assessed by Marsaglia's (1972) lattice test or by Coveyou and MacPherson's (1967) spectral test. Fishman and Moore (1982) carried out an empirical study and observed that different values of the multiplier, all of which perform very well under both lattice and spectral tests, can yield quite different performances based on the criterion of similarity of samples generated to actual samples from uniform distribution.

The possible choices for c are 16807, 37204094, and 950706376. The first choice has been observed to result in the fastest execution time, while the last choice has been observed by Fishman and Moore to give the best performance. The method or routine described above is portable in the sense that given the same seed or initial value, it will produce the same sequence in all computer/compiler environments.

One may also use a "shuffled" version of this generator. The shuffled generators use a scheme due to Learmonth and Lewis (1973) under which a table is filled with the first 128 standard uniform pseudorandom numbers resulting from the multiplicative congruential generator in (26.92). Then, for each x, obtained from the generator, the low-order bits of x_i are used to select a random integer I from 1 to 128. The ith entry in the table is then delivered as the pseudorandom number, and x_i (after being scaled into the unit interval) is inserted into the Ith position in the table. Books by Kennedy and Gentle (1980) and Devroye (1986) provide elaborate discussions on other simulational methods for uniform as well as nonuniform populations.

Order statistics from a uniform population play a key role in many statistical problems. Computer simulation of uniform order statistics is essential, since it enables one to evaluate the performance of these statistical procedures through Monte Carlo simulations. A simple and direct way of simulating uniform order statistics is to generate a pseudorandom sample from the uniform distribution (by the routine described above) and then to sort the sample through an efficient algorithm like quick-sort. This direct method will naturally be time-consuming, slow, and expensive. A considerable amount of work has been done to improve this general method and to provide efficient algorithms for generating uniform order statistics.

Schucany (1972) suggested a method of simulating uniform order statistics by making use of the fact that X_i' (in a sample of size n) is distributed as beta $(i, n - i + 1)$ (see Section 26.4). For example, if the largest order statistic X_n' is required, it may be generated as $u_1^{1/n}$, where u_1 is a pseudorandom uniform $(0, 1)$ observation. Then the second largest order statistic X_{n-1}' may be generated as $u_1^{1/n} u_2^{1/(n-1)}$, where u_2 is another independent (of U_1) pseudorandom uniform $(0, 1)$ observation. Proceeding similarly, the order statistic X_{n-i+1}' may be generated as

$$X_{n-i+1}' = u_1^{1/n} u_2^{1/(n-1)} \cdots u_i^{1/(n-i+1)}. \qquad (26.93)$$

This method is called the descending method, and it avoids sorting altogether.

Lurie and Hartley (1972) presented a similar method that generates the uniform order statistics in an ascending order starting from the smallest order statistic. This method is called the ascending method. By conducting an empirical comparison, Lurie and Mason (1973) observed the descending method to be slightly faster than the ascending method.

Ramberg and Tadikamalla (1978) and Horn and Schlipf (1986) have presented algorithms for generating some central uniform order statistics. Lurie and Hartley (1972) provided another interesting algorithm for generating uniform order statistics. This algorithm is based on the result that if $Y_1, Y_2, \dots, Y_{n+1}$ are independent standard exponential random variables, then (with $Z = Y_1 + Y_2 + \cdots + Y_{n+1}$),

$$\frac{Y_1}{Z}, \frac{Y_2}{Z}, \dots, \frac{Y_n}{Z}$$

are distributed as $X_1', X_2' - X_1', \dots, X_n' - X_{n-1}'$. Therefore the uniform order statistic X_i' for a random sample of size n may be generated as

$$X_i' = \frac{Y_1 + Y_2 + \cdots + Y_i}{Y_1 + Y_2 + \cdots + Y_{n+1}}, \tag{26.94}$$

or equivalently as

$$X_i' = \frac{\sum_{j=1}^{i} \log U_j}{\sum_{j=1}^{n+1} \log U_j}, \tag{26.94$'$}$$

where $U_1, U_2, \ldots, U_{n+1}$ are pseudorandom uniform (0, 1) observations. This "exponential method" needs fewer steps to be performed, but one extra uniform observation is necessary than for the descending method.

Recently Balakrishnan and Sandhu (1995) proposed a simple and efficient simulational algorithm for generating progressive Type-II censored samples from the uniform (0, 1) distribution. Under this censoring scheme, n units are placed on a life test; after the first failure, R_1 surviving items are removed at random from further observation; after the second failure, R_2 surviving items are removed at random, and so on; finally, after the mth failure (last observed failure), R_m remaining items are withdrawn so that $n = m + (R_1 + R_2 + \cdots + R_m)$. Denoting $X_{(1)}, X_{(2)}, \cdots, X_{(m)}$ to be the progressive censored sample from the uniform (0, 1) distribution and

$$Y_i = \frac{1 - X_{(m-i+1)}}{1 - X_{(m-i)}}, \qquad i = 1, 2, \ldots, m-1, \; Y_m = 1 - X_{(1)}, \tag{26.95}$$

Balakrishnan and Sandhu (1995) have established that

$$U_i = Y_i^{i + R_m + R_{m-1} + \cdots + R_{m-i+1}}, \qquad i = 1, 2, \ldots, m, \tag{26.96}$$

are independent with a common uniform (0, 1) distribution. Their simulational algorithm is based on this distributional result. All of these algorithms for generating uniform order statistics can be used to generate order statistics from any other continuous population through the inverse cdf method (since it is order preserving). [See also Gerontidis and Smith (1982).]

REFERENCES

Abdelhamid, S. N. (1985). On a characterization of rectangular distributions, *Statistics & Probability Letters*, **3**, 235–238.

Ahsanullah, M. (1989). On characterizations of the uniform distribution based on functions of order statistics, *Aligarh Journal of Statistics*, **9**, 1–6.

Ali, M. M. (1975). Tail distribution of 'Student's' ratio for $t \geq n - 1$ in samples of size n from rectangular distribution, ***Journal of Statistical Research, 9,*** 11–24.

Ali, M. M. (1976). Tail distribution of 'Student's ratio for $t \geq (n - 2)/2$ in samples of size n from rectangular distribution, ***Journal of Statistical Research,*** **10,** 43–71.

Allan, R. R. (1966). Extension of the binomial model of traffic flow to the continuous case, ***Proceedings of the Third Conference of the Australian Road Research Board,*** **3,** 276–316.

Anderson, P. H. (1942). Distributions in stratified sampling, ***Annals of Mathematical Statistics,*** **13,** 42–52.

Arnold, B. C., and Meeden, G. (1976). A characterization of the uniform distribution based on summation modulo one with application to fractional backlogs, ***Australian Journal of Statistics,*** **18,** 173–175.

Ayyangar, A. A. K. (1941). The triangular distribution, ***Mathematics Student,*** 9, 85–87.

Balakrishnan, N., and Balasubramanian, K. (1993). Equivalence of Hartley-David-Gumbel and Papathanasiou bounds and some further remarks, ***Statistics & Probability Letters, 16,*** 39–41.

Balakrishnan, N., and Sandhu, R. A. (1995). A simple simulational algorithm for generating progressive Type-I1 censored samples, ***The American Statistician*** (to appear).

Barrow, D. L., and Smith, P. W. (1979). Spline notation applied to a volume problem, ***American Mathematical Monthly,*** **86,** 50–51.

Bates, G. E. (1955). Joint distributions of time intervals for the occurrence of successive accidents in a generalized Pólya scheme, ***Annals of Mathematical Statistics,*** **26,** 705–720.

Bayes, T. (1763). An essay towards solving a problem in the Doctrine of Chances, ***Philosophical Transactions of the Royal Society of London,*** **53,** 370–418.

Bhate, D. H. (1951). A note on the significance level of the distribution of the means of a rectangular population, ***Bulletin of the Calcutta Statistical Association,*** **3,** 172–173.

Broadbent, S. R. (1954). The quotient of a rectangular or triangular and a general variate, ***Biometrika,*** **41,** 330–337.

Brown, T. C., Cartwright, O. I., and Eagleson, G. K. (1986). Correlations and characterizations of the uniform distribution, ***Australian Journal of Statistics,*** **28,** 89–96.

Buss, O. L., Jr. (1972). ***A Supplement to Haight's "Index to the Distributions of Mathematical Statistics,"*** M.Sc. thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, OH, GSA/MA/72-2.

Carlton, A. G. (1946). Estimating the parameters of a rectangular distribution, ***Annals of Mathematical Statistics,*** **17,** 355–358.

Chen, R., Goodman, R., and Zame, A. (1984). On the limiting distribution of two random sequences, ***Journal of Multivariate Analysis,*** **14,** 221–230.

Chen, R., Lin, E., and Zame, A. (1981). Another arc sine law, ***Sankhyā, Series A,*** **43,** 371–383.

Chu, J. T. (1957). Some uses of quasi-ranges, ***Annals of Mathematical Statistics,*** **28,** 173–180.

Clark, C. E. (1966). *Random Numbers in Uniform and Normal Distribution,* San Francisco: Chandler.

Constantine, G. (ed.) (1980). Problem Corner, *DMS Newsletter,* 66, *2.* (CSIRO, Glen Osmond, South Australia).

Coveyou, R. R., and MacPherson, R. D. (1967). Fourier analysis of uniform random number generators, *Journal of the Association for Computing Machinery,* **14,** 100–119.

Cowan, R. (1980). A letter, In Constantine (1980).

Cox, D. R., and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events,* London: Methuen.

Das Gupta, S., Goswami, A., and Rao, B. V. (1993). On a characterization of uniform distributions, *Journal of Multicariate Analysis, 44,* 102–114.

David, F. N., and Johnson, N. L. (1956). Some tests of significance with ordered variables, *Journal of the Royal Statistical Society, Series B,* **18,** 1–20; Discussion, 20–31.

Deng, L.-Y., and George, E. 0 . (1992). Some characterizations of the uniform distribution with application to random number generation, *Manuscript,* Department of Mathematical Sciences, Memphis State University, Memphis, TN.

Devroye, L. (1986). *Non-uniform Random Variate Generation,* New York: Springer-Verlag.

Driscoll, M. F. (1978). On pairwise and mutual independence characterizations of rectangular distributions, *Journal of the American Statistical Association, 73,* 432–433.

Durbin, J. (1961). Some methods of constructing exact tests, *Biometrika,* **48,** 41–55.

Eisenhart, C., Deming, L. S., and Martin, C. S. (1963). *Tables describing small-sample properties of the mean, median, standard deviation, and other statistics in sampling from various distributions,* U.S. National Bureau of Standards, Technical Note 191.

Eltessi, A., and Pal, N. (1992). Estimation of the smallest and the largest of two uniform scale parameters, *Communications in Statistics—Theory and Methods,* 21, 2185–2201.

Epstein, B. (1948). Some applications of the Mellin transform in statistics, *Annals of Mathematical Statistics,* **19,** 370–379.

Fan, D.-Y. (1991). On a property of the order statistics of the uniform distribution, *Communications in Statistics—Theory and Methods, 20,* 1903–1909.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications,* Vol. 2, New York: Wiley.

Fisher, R. A. (1932). *Statbtical Methods for Research Workers,* 4th ed., Edinburgh: Oliver & Boyd.

Fishman, G. S., and Moore, L. R. (1982). A statistical evaluation of multiplicative congruential random number generators with modulus $2^{31} - 1$, *Journal of the American Statistical Association, 77,* 129–136.

Galambos, J., and Kotz, S. (1978). *Characterizations of Probability Distributions,* Lecture Notes in Mathematics, 675, New York: Springer-Verlag.

Geary, R. C. (1944). Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter, *Biometrika*, **33**, 123–128.

Gerontidis, I., and Smith, R. L. (1982). Monte Carlo generation of order statistics from general distributions, *Applied Statistics*, **31**, 238–243.

Gibbons, J. D. (1974). Estimation of the unknown upper limit of a uniform distribution, *Sankhyā, Series B*, **36**, 29–40.

Gibbons, J. D., and Litwin, S. (1974). Simultaneous estimation of the unknown upper and lower limits in a two-parameter uniform distribution, *Sankhyā, Series B*, **36**, 41–54.

Goldman, A. J. (1968). Fractional container loads and topological groups, *Operations Research*, **16**, 1218–1221.

Graybill, F. A., and Connell, T. L. (1964). Sample size required to estimate the parameter in the uniform density within d units of the true value, *Journal of the American Statistical Association*, **59**, 550–556.

Gumbel, E. J. (1954). The maxima of the mean largest value and of the range, *Annals of Mathematical Statistics*, **25**, 76–84.

Gupta, A. K., and Miyawaki, T. (1978). On a uniform mixture model, *Biometrical Journal*, **20**, 631–637.

Gupta, S. S., and Sobel, M. (1958). On the distribution of a statistic based on ordered uniform chance variables, *Annals of Mathematical Statistics*, **29**, 274–281.

Haight, F. A. (1958). Two queues in parallel, *Biometrika*, **45**, 401–410.

Haight, F. A. (1960). Queueing with balking, *Biometrika*, 47, 285–296.

Haight, F. A. (1961). Index to the distributions of mathematical statistics, *Journal of Research of the National Bureau of Standards*, **65B**, 23–60.

Haight, F. A. (1965). On the effect of removing persons with N or more accidents from an accident-prone population, *Biometrika*, **52**, 298–300.

Hall, P. (1932). The distribution of means for samples of size N drawn from a population in which the variate takes values between 0 and 1, *Biometrika*, **19**, 240–249.

Hamdan, M. A. (1972). On a characterization by conditional expectations, *Technometrics*, **14**, 497–499.

Harter, H. L. (1978). Adaptive robust estimation of location and scale parameters of symmetric populations, *Technical Report AFFDL-TR-78-128*, Wright-Patterson Air Force Base, OH.

Hartley, H. O., and David, H. A. (1954). Universal bounds for mean range and extreme observation, *Annals of Mathematical Statistics*, **25**, 85–99.

Herer, W. (1993). A characterization of uniformly distributed random variable, *Demonstratio Mathematica*, **26**, 207–212.

Holewijn, P. J. (1969). Note on Weyl's criterion and the uniform distribution of independent random variables, *Annals of Mathematical Statistics*, *40*, 1124–1125.

Horn, P. S., and Schlipf, J. S. (1986). Generating subsets of order statistics with applications to trimmed means and means of trimmings, *Journal of Statistical Computation and Simulation*, **24**, 83–97.

Huang, J. S., Arnold, B. C., and Ghosh, M. (1979). On characterization of the uniform distribution based on identically distributed spacings, ***Sankhyā*, *Series B*, *41*,** 109–115.

Hull, T. E., and Dobell, A. R. (1962). Random number generators, ***Journal of the Society of Industrial and Applied Mathematics*, *4*,** 230–254.

Hyrenius, H. (1953). On the use of ranges, cross-ranges and extremes in comparing small samples, ***Journal of the American Statistical Association*, *48*,** 534–545. (Correction: 48, 907.)

Irwin, J. D. (1932). On the frequency distribution of the means of samples from a population having any law of frequency, ***Biometrika*,** 19, 234–239.

Johnson, N. L. (1950). On the comparison of estimators, ***Biometrika*,** 37, 281–287.

Johnson, N. L., and Kotz, S. (1990). Randomly weighted averages: Some aspects and extensions, ***The American Statistician*, *44*,** 243–249.

Joó, I., and Szabó, S. (1992). On the estimate $(x_{\min} + x_{\max})/2$, ***Studia Scientiarum Mathematicarum Hungarica*, *27*,** 409–432.

Kendall, M. G., and Babington Smith, B. (1938). Randomness and random sampling numbers, ***Journal of the Royal Statistical Society, Series A*, *101*,** 147–166.

Kendall, M. G., and Babington Smith, B. (1940). ***Tables of Random Sampling Numbers*,** Tracts for Computers, vol. 24, Cambridge: Cambridge University Press.

Kennedy, W. J., and Gentle, J. E. (1980). ***Statistical Computing*,** New York: Dekker.

Kotz, S. (1974). Characterizations of statistical distributions: A supplement to recent surveys, ***International Statistical Review*, *42*,** 39–65.

Laplace, P. S. (1812). ***Théorie Analytique des Probabilités***, 1st ed., Paris.

Learmonth, G., and Lewis, P. A. W. (1973). Naval Postgraduate School random number generator package LLRANDOM, In ***Computer Science and Statistics: 7th Annual Symposium on the Interface*** (ed., *W. J.* Kennedy), pp. 163–171.

Leone, F. C. (1961). The use of sample quasi-ranges in setting confidence intervals for the population standard deviation, ***Journal of the American Statistical Association*,** 56, 260–272.

Levene, H. (1952). On the power function of tests of randomness based on runs up and down, ***Annals of Mathematical Statistics*, *23*,** 34–56.

Lin, G. D. (1986). Characterizations of uniform distributions and of exponential distributions, ***Technical Report*,** Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics, ***Biometrika*,** 39, 88–95.

Lowan, A. N., and Laderman, J. (1939). On the distribution of errors in N-th tabular differences, ***Annals of Mathematical Statistics*, *10*,** 360–364.

Lukacs, E. (1979). A characterization of the rectangular distribution, ***Stochastic Processes and Their Applications*, *9*,** 273–279.

Lurie, D., and Hartley, H. O. (1972). Machine generation of order statistics for Monte Carlo simulations, ***The American Statistician*, *26*,** 26–27.

Lurie, D., and Mason, R. L. (1973). Empirical investigation of several techniques for computer generation of order statistics, ***Communications in Statistics*, *2*,** 363–371.

Ma, C. (1992). Variance bound of function of order statistics, ***Statistics & Probability Letters,*** 13, 25–27.

Marsaglia, G. (1961). Expressing a random variable in terms of uniform random variables, ***Annals of Mathematical Statistics,*** 32, 894–898.

Marsaglia, G. (1972). The structure of linear congruential sequences. In ***Applications of Number Theory to Numerical Analysis*** (ed., S. K. Zaremba), pp. 249–286, San Diego, CA: Academic Press.

Massey, J. L. (1988). An introduction to contemporary cryptology, ***Proceedings of IEEE, 76,*** 533–549.

Mitra, S. K., and Banerjee, S. N. (1971). On the probability distribution of round-off errors in tabular differences, ***Australian Computer Journal,*** **3(2),** 60–68.

Móri, T. F. (1983). Note on the Cramér-Rao inequality in the nonregular case: The family of uniform distributions, ***Journal of Statistical Planning and Inference, 7,*** 353–358.

Morimoto, H., and Sibuya, M. (1967). Sufficient statistics and unbiased estimation of restricted selection parameters, ***Sankhyā, Series A,*** 29, 15–40.

Murty, V. N. (1955). The distribution of the quotient of maximum values in samples from a rectangular distribution, ***Journal of the American Statistical Association,*** 50, 1136–1141.

Nagaev, S. V., and Mukhin, A. B. (1966). On a case of convergence to a uniform distribution on an interval, In ***Limit Theorems and Statistical Inference*** (ed., S. H. Siraždinov), Tashkent: FAN, pp. 113–116. (In Russian)

Naus, I. (1966). A power comparison of two tests of non-random clustering, ***Technometrics,*** 8, 493–517.

Nikulin, M. S. (1991). Remark on estimation of parameters of the uniform distribution, In ***Statistical Estimation and Testing Hypothesis Methods,*** Perm', Russia: Perm' University, pp. 36–38. (In Russian)

Ouyang, L. Y. (1993). Characterizations of the uniform distribution by conditional expectation, ***International Journal of Information and Management Sciences, 4,*** 107–111.

Packer, L. R. (1950). The distribution of the sum of *n* rectangular variates, *I*, ***Journal of the Institute of Actuaries Students' Society, 10,*** 52–61.

Papathanasiou, V. (1990). Some characterization of distributions based on order statistics, ***Statistics & Probability Letters, 9,*** 145–147.

Pearson, E. S. (1938). Tests based on the probability integral transformation, ***Biometrika,*** 30, 134–148.

Perlo, V. (1933). On the distribution of Student's ratio for samples of three drawn from a rectangular distribution, ***Biometrika,*** 25, 203–204.

Proctor, J. W. (1987). Estimation of two generalized curves covering the Pearson system, ***Proceedings of ASA Computing Section,*** pp. 287–292.

Pusz, J. (1988). On a characterization of probability distributions by conditional expectations, ***Demonstratio Mathematica,*** 21, 247–253.

Quesenberry, C. P. (1986). Probability integral transformations, In ***Encyclopedia of Statistical Sciences,*** vol. 7 (eds., S. Kotz, N. L. Johnson, and C. B. Read), New York: Wiley, pp. 225–231.

Ramberg, J. S., and Tadikamalla, P. R. **(1978).** On the generation of subsets of order statistics, ***Journal of Statistical computation and simulation,* 6, 239–241.**

Rand Corporation **(1955). A *Million Random Digits and 100,000 Normal Deviates,*** Glencoe, IL: Free Press.

Rider, P. R. **(1929).** On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes, ***Biometrika,* 21, 124–143.**

Rider, P. R. (1951). The distribution of the quotient of ranges in samples from a rectangular population, ***Journal of the American Statistical Association, 46,* 502–507.**

Rider, P. R. **(1963). *Percentage Points of the Ratio of Ranges of Samples from a Rectangular Distribution,*** ARL **63-194,** Aerospace Research Laboratories, U.S. Air Force, Wright-Patterson Air Force Base, Dayton, OH.

Rizzi, A. **(1990).** Some theorems on the sum modulo ***n*** of two random variables, ***Metron, 48,* 149–160.**

Roach, S. A. **(1963).** The frequency distribution of the sample mean where each member of the sample is drawn from a different rectangular distribution, ***Biometrika,* 50, 508–513.**

Roy, M. K., Roy, A. K., and Ali, M. **Masoom (1993).** Binomial mixtures of some standard distributions, ***Journal of Information & Optimization Sciences,* 14, 57–71.**

Rukhin, A. L., Kuo, L., and Dey, D. K. **(1990).** A class of minimax estimators of the scale parameter of the uniform distribution, ***Statistics & Probability Letters, 9,* 317–321.**

Sakamoto, H. (1943). On the distribution of the product and the quotient of the independent and uniformly distributed random variables, ***Tohoku Mathematical Journal,* 49, 243–260.**

Saleh, A. K. Md. E. **(1976).** Characterization of distributions using expected spacings between consecutive order statistics, ***Journal of Statistical Research,* 10, 1–13.**

Sarhan, A. **E.** (1955). Estimation of the mean and standard deviation by order statistics, **III, *Annals of Mathematical Statistics,* 26, 576–592.**

Sarhan, A. E., and Greenberg, B. G. **(1959).** Estimation of location and scale parameters for the rectangular population from censored samples, ***Journal of the Royal Statistical Society, Series B,* 21, 356–363.**

Sarhan, A. E., and Greenberg, B. G. **(eds.) (1961). *Contributions to Order Statistics,*** New York: Wiley.

Schmidt, R. **(1934).** Statistical analysis of one-dimensional distributions, ***Annals of Mathematical Statistics,* 5,** 30–43.

Schucany, **W.** R. **(1972).** Order statistics in simulation, ***Journal of Statistical Computation and Simulation,* 1, 281–286.**

Scott, D. (1980). A letter, In Constantine (1980).

Scozzafava, S. (1991). Sum and difference modulo ***m*** between two random variables, ***Metron,* 49, 495–511.**

Scozzafava, S. **(1993).** Uniform distribution and sum modulo ***m*** of random variables, ***Statistics & Probability Letters, 18,* 313–314.**

Seal, H. L. **(1950).** Spot the prior reference, ***Journal of the Institute of Actuaries Students' Society,* 10, 255–256.**

Seshadri, V., and Shuster, J. J. **(1971).** A characterization of the uniform distribution and an application to goodness of fit testing, ***Manuscript,*** **McGill** University, Montreal, Canada.

Sheppard, W. F. **(1907).** Calculation of moments of a frequency distribution, ***Biometrika,*** **5, 450–459.**

Shimizu, R., and Huang, J. S. **(1983).** On a characteristic property of the uniform distribution, ***Annals of the Institute of Statistical Mathematics,*** **35, 91–94.**

Siddiqui, M. M. **(1964).** Distribution of Student's t in samples from a rectangular universe, ***Revue de l'Institut International de Statistique*****, 32, 242–250.**

Springer, M. D. **(1978).** ***The Algebra of Random Variables,*** New York: Wiley.

Stapleton, J. H. **(1963).** A characterization of the uniform distribution on a compact topological group, ***Annals of Mathematical Statistics*****, 34, 319–326.**

Stephens, M. A. **(1966).** Statistics connected with the uniform distribution: percentage points and application to testing for randomness of directions, ***Biometrika,*** **53, 235–238.**

Sukhatme, P. V. **(1937).** Tests of significance for samples from the χ^2 population with two degrees of freedom, ***Annals of Eugenics, London,*** **8, 52–56.**

Sumitra, P., and Kumar, B. S. **(1990).** Characterization of uniform distributions by inequalities of Chernoff-type, ***Sankhyii, Series A*****, 52, 376–382.**

Székely, G. J., and **Móri,** T. F. **(1985).** An extremal property of rectangular distributions, ***Statistics & Probability Letters,*** **3, 107–109.**

Tach, L. T. **(1958).** ***Tables of Cumulative Distribution Function of a Sum of Independent Random Variables,*** Convair Aeronautics Report No. ZU-7-119-TN, San **Diego, CA.**

Terrell, G. R. **(1983).** A characterization of rectangular distributions, ***Annals of Probability,*** *11,* **828–836.**

Tippett, L. H. C. **(1927).** Random sampling numbers, ***Tracts for Computers,*** vol. **15,** Cambridge: Cambridge University Press.

Troutt, M. D. **(1991).** A theorem on the density of the density ordinate and an alternative interpretation of the Box-Muller method, ***Statistics*****, 22, 463–466.**

Van Assche, W. **(1987).** A random variable uniformly distributed between two random variables, ***Sankhyā*****,** ***Series A****, 49,* **207–211.**

Vincze, I. **(1979).** On the **Cramér-Fréchet-Rao** inequality in the non-regular case, In ***Contributions to Statistics: The J. Hajek Memorial Volume,*** Prague: Academic, pp. **253–262.**

Westcott, M. **(1980).** A letter, In Constantine **(1980).**

Zhao-Guo, C., and Hong-Zhi, A. **(1980).** A letter, In Constantine **(1980).**

CHAPTER 27

F-Distributions

1 INTRODUCTION

If X_1, X_2 are independent random variables distributed as $\chi^2_{\nu_1}, \chi^2_{\nu_2}$, respectively, then the distribution of

$$\left(\frac{X_1}{\nu_1}\right)\left(\frac{X_2}{\nu_2}\right)^{-1}$$

is the F-distribution with ν_1, ν_2 degrees of freedom. As a special case note that if Y_1, Y_2 are independent random variables with a common Laplace distribution (see Chapter 24) centered at zero then $|Y_1/Y_2|$ has a $F_{2,2}$ distribution. We will use the symbol F_{ν_1,ν_2} generically to denote a random variable having this distribution. The phrase "F_{ν_1,ν_2}-distribution" can be used as an abbreviation for "F-distribution with ν_1, ν_2 degrees of freedom."

Note that the order ν_1, ν_2 is important. From the definition it is clear that the random variables F_{ν_1,ν_2} and $F^{-1}_{\nu_2,\nu_1}$ have identical distributions. In particular, using the suffix α to denote "lower $100\alpha\%$ point,"

$$F_{\nu_1,\nu_2,\alpha} = F^{-1}_{\nu_2,\nu_1,1-\alpha} \tag{27.1}$$

(since $\Pr[F_{\nu_1,\nu_2} \le K] = \Pr[F_{\nu_2,\nu_1} \ge K^{-1}]$).

The importance of the F-distribution in statistical theory derives mainly from its applicability to the distribution of ratios of independent 'estimators of variance'. If $\{X_{ti}\}$ ($t = 1,2$; $i = 1,2,\dots,n_t$; $n_t \ge 2$) denote independent random variables each normally distributed, and the expected value and standard deviation of X_{ti} are ξ_t, σ_t, respectively (not depending on i), then

$$S_t^2 = (n_t - 1)^{-1}\sum_{i=1}^{n_t}(X_{ti} - \bar{X}_{t.})^2 \qquad \text{with } \bar{X}_{t.} = n_t^{-1}\sum_{i=1}^{n_t} X_{ti}$$

is distributed as $\chi^2_{n_t-1}\sigma_t^2(n_t - 1)^{-1}$ ($t = 1,2$). The ratio (S_1^2/S_2^2) is therefore distributed as

$$\frac{\chi^2_{n_1-1}\sigma_1^2(n_1-1)^{-1}}{\chi^2_{n_2-1}\sigma_2^2(n_2-1)^{-1}},$$

that is, as

$$\left(\frac{\sigma_1}{\sigma_2}\right)^2 F_{n_1-1,n_2-1}.$$

The statistic $(S_1/S_2)^2$ is used in testing the hypothesis of equality between σ_1 and σ_2. The hypothesis is rejected if either

$$\left(\frac{S_1}{S_2}\right)^2 \le F_{n_1-1,n_2-1,\alpha_1} \quad \text{or} \quad \left(\frac{S_1}{S_2}\right)^2 \ge F_{n_1-1,n_2-1,1-\alpha_2}$$

with $\alpha_1 + \alpha_2 < 1$. The significance level of the test is $(\alpha_1 + \alpha_2)$, and the power (with respect to a specified value of the ratio σ_1/σ_2) is

$$1 - \Pr\left[F_{n_1-1,n_2-1,\alpha_1} < \left(\frac{S_1}{S_2}\right)^2 < F_{n_1-1,n_2-1,1-\alpha_2}\middle|\frac{\sigma_1}{\sigma_2}\right]$$

$$= 1 - \Pr\left[\left(\frac{\sigma_2}{\sigma_1}\right)^2 F_{n_1-1,n_2-1,\alpha_1} < F_{n_1-1,n_2-1} < \left(\frac{\sigma_2}{\sigma_1}\right)^2 F_{n_1-1,n_2-1,1-\alpha_2}\right]. \tag{27.2}$$

Since

$$\Pr\left[\left(\frac{\sigma_1}{\sigma_2}\right)^2 F_{n_1-1,n_2-1,\alpha_1} < \left(\frac{S_1}{S_2}\right)^2 < \left(\frac{\sigma_1}{\sigma_2}\right)^2 F_{n_1-1,n_2-1,1-\alpha_2}\middle|\frac{\sigma_1}{\sigma_2}\right]$$
$$= 1 - \alpha_1 - \alpha_2,$$

the values

$$\left(\frac{S_1}{S_2}\right)^2 (F_{n_1-1,n_2-1,1-\alpha_2})^{-1},$$

$$\left(\frac{S_1}{S_2}\right)^2 (F_{n_1-1,n_2-1,\alpha_1})^{-1},$$

enclose a confidence interval for $(\sigma_1/\sigma_2)^2$, with confidence coefficient $(1 - \alpha_1 - \alpha_2)$. For a given value α, say of $(\alpha_1 + \alpha_2)$, the length of this interval is minimized by choosing α_1, α_2 (subject to $\alpha_1 + \alpha_2 = \alpha$) to minimize

$$\left(F_{n_1-1,n_2-1,\alpha_1}\right)^{-1} - \left(F_{n_1-1,n_2-1,1-\alpha_2}\right)^{-1} = F_{n_2-1,n_1-1,1-\alpha_1} - F_{n_2-1,n_1-1,\alpha_2}.$$

This is achieved by making the ordinates of the probability density function of F_{n_2-1,n_1-1} equal at the values $F_{n_2-1,n_1-1,1-\alpha_1}$ and F_{n_2-1,n_1-1,α_2} if $n_2 \geq 3$; if $n_2 < 3$, then α_1 must be taken equal to α, and $\alpha_2 = 0$, so that

$$F_{n_2-1,n_1-1,\alpha_2} = 0.$$

It must be noted that minimization of the length of confidence interval for $(\sigma_1/\sigma_2)^2$ does not usually minimize the length of interval for $(\sigma_2/\sigma_1)^2$ (or for σ_1/σ_2 or σ_2/σ_1). It is more natural to minimize the length of interval for $\log(\sigma_1/\sigma_2)$—this will minimize the length for $\log(\sigma_2/\sigma_1)$, or indeed for $\log[(\sigma_1/\sigma_2)^r]$ for any $r \neq 0$. To do this, we need to choose α_1, α_2 (subject to $\alpha_1 + \alpha_2 = \alpha$) so that the probability density function of F_{n_1-1,n_2-1} has the same value at $f = F_{n_1-1,n_2-1,1-\alpha_2}$ and at $f = F_{n_1-1,n_2-1,\alpha_1}$. Tables giving appropriate values for $\alpha = 0.01, 0.05, 0.10, 0.25$, and $n_1, n_2 = 6(1)31$ (with $n_1 \geq n_2$) have been constructed by Tiao and Lochner (1966). Further applications for the F-distribution are described in Section 27.6.

Figure 27.1 [reproduced from Hald (1952)] contains graphs of the probability density function of F_{ν_1,ν_2} for $\nu_1 = 10$ and a number of values of ν_2. It can be seen that the graphs appear to approach a limiting form. This is, in fact the distribution of $\chi^2_{\nu_1}/\nu_1$ with $\nu_1 = 10$.

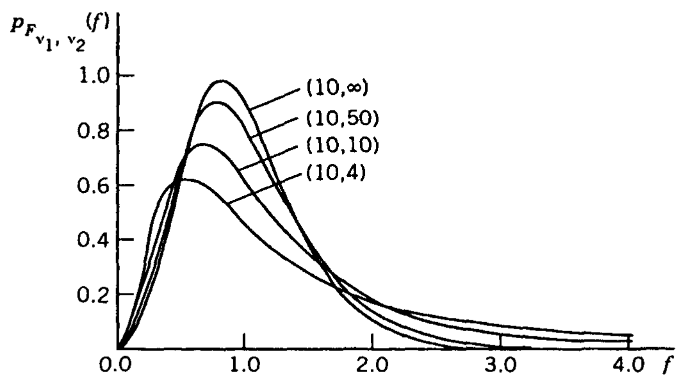


Figure 27.1 ***F*** **Density Functions**

For a brief recent description of various properties of the F-distribution and information regarding available tables and approximations, interested readers may refer to Stuart and Ord (1994, pp. 549–555).

2 PROPERTIES

If X_1, X_2 are distributed as described in Section 27.1, then the probability density function of $G_{\nu_1,\nu_2} = X_1/X_2$ is

$$p_{G_{\nu_1,\nu_2}}(g) = \frac{1}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \frac{g^{(\nu_1/2)-1}}{(1+g)^{(\nu_1+\nu_2)/2}}, \qquad 0 < g. \tag{27.3}$$

This is a Pearson Type VI distribution (Chapter 12, Section 4), also known as a ***beta distribution of the second kind*** (see Chapter 25, Section 7). We will omit the subscripts ν_1, ν_2 unless there might be confusion arising from the omission.

The probability density function of $F = \nu_2 G/\nu_1$ is

$$p_F(f) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \frac{f^{(\nu_1/2)-1}}{\left(1 + \nu_1\nu_2^{-1}f\right)^{(\nu_1+\nu_2)/2}}, \qquad 0 < f. \tag{27.4}$$

As f tends to infinity, $p_F(f)$ tends to zero; this is also the case as f tends to zero, provided that $\nu_1 > 2$. In this case there is a single mode, at

$$f = [\nu_2(\nu_1 - 2)][\nu_1(\nu_2 + 2)]^{-1}.$$

If $\nu_1 = 2$, there is a mode at $f = 0$; if $\nu_1 = 1$, $p_F(f) \to \infty$ as $f \to 0$.

The rth moment of F about zero is

$$\mu'_r = \left(\frac{\nu_2}{\nu_1}\right)^r E\left[\left(\chi^2_{\nu_1}\right)^r\right] E\left[\left(\chi^2_{\nu_2}\right)^{-r}\right]$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\nu_1(\nu_1+2)\dots(\nu_1 + 2\,\overline{r-1})}{(\nu_2-2)(\nu_2-4)\cdots(\nu_2-2r)}. \tag{27.5}$$

Note that if $r \geq \frac{1}{2}\nu_2$, μ'_r is infinite. In particular,

$$E[F] = \frac{\nu_2}{\nu_2 - 2}, \qquad \nu_2 > 2, \tag{27.6a}$$

$$\text{var}(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \qquad \nu_2 > 4, \tag{27.6b}$$

$$\text{Coefficient of variation } [\text{CV}(F)] = \left\{ \frac{2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)} \right\}^{1/2}, \tag{27.6b$'$}$$

$$\sqrt{\beta_1} = \sqrt{\frac{8(\nu_2 - 4)}{(\nu_1 + \nu_2 - 2)\nu_1}} \cdot \frac{(2\nu_1 + \nu_2 - 2)}{(\nu_2 - 6)}, \qquad \nu_2 > 6, \tag{27.6c}$$

$$\beta_2 = 3 + \frac{12\{(\nu_2 - 2)^2(\nu_2 - 4) + \nu_1(\nu_1 + \nu_2 - 2)(5\nu_2 - 22)\}}{\nu_1(\nu_2 - 6)(\nu_2 - 8)(\nu_1 + \nu_2 - 2)}$$

$$= \frac{3\{\nu_2 - 4 + \frac{1}{2}(\nu_2 - 6)\beta_1\}}{\nu_2 - 8}, \qquad \nu_2 > 8 \tag{27.6d}$$

[Wishart (1946)].

After pointing out the error in the form of the characteristic function given by Ifram (1970) (and also in the first edition of this volume), Awad (1980) presented the characteristic function of the distribution of G_{ν_1,ν_2} in (27.3) as

$$\frac{1}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \sum_{r=0}^{\infty} \sum_{j=0}^{\infty} \frac{(it)^r}{r!} \frac{1}{r + j + \frac{1}{2}\nu_1} \binom{r + j - \frac{1}{2}\nu_2}{j}; \tag{27.7}$$

see also Pestana (1977).

Phillips (1982) has presented the characteristic function of the F_{ν_1,ν_2} distribution in another form involving the confluent hypergeometric function of the second kind. To this end, noting that for the complex variable z and complex parameters a and c such that $\text{Re}(z) > 0$ and $\text{Re}(a) > 0$

$$\int_0^{\infty} e^{-zt} t^{a-1}(1 + t)^{c-a-1}\, dt = \Gamma(a)\Psi(a, c; z)$$

which defines the confluent hypergeometric function of the second kind, Phillips (1982) has shown that the characteristic function of F_{ν_1,ν_2} is given by

$$\frac{\Gamma[\frac{1}{2}(\nu_1+\nu_2)]}{\Gamma(\nu_2/2)}\Psi\left(\frac{\nu_1}{2},1-\frac{\nu_2}{2};-\frac{\nu_2}{\nu_1}it\right). \qquad (27.7)'$$

From (27.3) it can be seen that $G(1+G)^{-1}=\nu_1F(\nu_2+\nu_1F)^{-1}$ has a standard beta distribution (as defined in Chapter 25) with parameters $\frac{1}{2}\nu_1, \frac{1}{2}\nu_2$. It follows that

$$\Pr[F\le f]=I_{f_1}(\tfrac{1}{2}\nu_1,\tfrac{1}{2}\nu_2) \qquad (27.8)$$

where

$$f_1=\frac{\nu_1 f}{\nu_2+\nu_1 f}.$$

This identity can be used for computer evaluation, using the methodology described in Chapter 25, Section 6.2, for the beta distribution. Note also that the relation [Chapter 1, Eq. (3.37)] connecting binomial distributions and incomplete beta function ratios means that cdfs of F-distributions can be expressed in terms of sums of binomial probabilities, and vice versa. Exact series expansions of the distribution function of an F_{ν_1,ν_2} variable are given by

$$\Pr[F_{\nu_1,\nu_2}\le y]$$

$$=\begin{cases}\dfrac{2\theta(y)}{\pi} & \text{if } \nu_1=\nu_2=1,\\ A(\nu_2) & \text{if } \nu_1=1,\ \nu_2>1 \text{ odd},\\ \dfrac{2\theta(y)}{\pi}-c(\nu_1,1) & \text{if } \nu_2=1,\ \nu_1>1 \text{ odd},\\ A(\nu_2)-c(\nu_1,\nu_2) & \text{if } \nu_1>1,\ \nu_2>1 \text{ both odd},\\ 1-x^{\nu_2/2}\left\{1+\dfrac{\nu_2}{2}(1-x)+\dfrac{\nu_2(\nu_2+2)}{2\cdot 4}(1-x)^2+\cdots\right. & \\ \qquad\left.+\dfrac{\nu_2(\nu_2+2)\cdots(\nu_2+\nu_1-4)}{2\cdot 4\cdots(\nu_1-2)}(1-x)^{(\nu_1-2)/2}\right\} & \text{if } \nu_1 \text{ even},\\ (1-x)^{\nu_1/2}\left\{1+\dfrac{\nu_1}{2}x+\dfrac{\nu_1(\nu_1+2)}{2.4}x^2+\cdots\right. & \\ \qquad\left.+\dfrac{\nu_1(\nu_1+2)\cdots(\nu_2+\nu_1-4)}{2.4\cdots(\nu_2-2)}x^{(\nu_2-2)/2}\right\} & \text{if } \nu_2 \text{ even},\end{cases}$$

$$(27.9)$$

where

$$\theta(y) = \tan^{-1}\left(y\sqrt{\frac{\nu_1}{\nu_2}}\right),$$

$$x = \frac{\nu_2}{\nu_2 + \nu_1 y},$$

$$A(\nu_2) = \frac{2}{\pi}\Bigg[\theta + \sin\theta\cos\theta$$

$$\times\left\{1 + \frac{2}{3}\cos^2\theta + \cdots + \frac{2\,4\cdots(\nu_2 - 3)}{3\,5\cdots(\nu_2 - 2)}\cos^{\nu_2 - 3}\theta\right\}\Bigg], \qquad \nu_2 > 1,$$

$$c(\nu_1, \nu_2) = \frac{2}{\sqrt{\pi}}\frac{((\nu_2 - 1)/2)!}{((\nu_2 - 2)/2)!}\sin 6\cos^{\nu_2} 6$$

$$\times\Bigg(I + \frac{\nu_2 + 1}{3}\sin^2 6 + \ldots$$

$$+ \frac{(\nu_2 + 1)(\nu_2 + 3)\cdots(\nu_1 + \nu_2 - 4)}{3\,5\cdots(\nu_1 - 2)}\sin^{\nu_1 - 3}\theta\Bigg\}, \qquad \nu_2 > 1.$$

For fractional a, $a! \equiv \Gamma(a + 1)$. These formulas were presented in Abramowitz and Stegun (1964) (Chapter 26, written by M. Zelen and N. C. Severo) and were verified by Chen and Makowsky (1976) (who discovered a minor misprint and provided a FORTRAN subprogram). Lee (1988) used (27.9) (with corrections) to evaluate cdf values for F and reported some advantages over the MDFD subroutine in IMSL (1985).

Random variates following F-distributions are easily constructed from beta variates, which can be obtained by methods described in Chapter 25, Section 2. Grzegórski (1972) provided a procedure, "PF *Snedecor*," to evaluate the cdf of F_{ν_1,ν_2}. The procedure is much faster than an algorithm "Fisher," proposed for integral degrees of freedom ν_1 and ν_2 by Donner (1968), and in some cases faster than Morris's (1969) procedure "F test." ("*F* test" is twice as long as "PF Snedecor" and uses twice as many variables.)

Grzegórski uses the formulas

$$\Pr[F_{\nu_1,\nu_2} \le f] = \begin{cases} P(\nu_1, \nu_2, x) & \text{for } \nu_2 \text{ even},\\ 1 - P(\nu_2, \nu_1, 1 - x) & \text{for } \nu_1 \text{ even},\\ 0.5 + [R(\nu_1, \nu_2, x) & \\ \quad - \sqrt{2(1 - x)}\dfrac{[(\nu_1 - 1)/2]!}{[(\nu_2 - 11/21!}P(\nu_1, \nu_2, x)]/\pi & \text{for } \nu_1, \nu_2 \text{ odd}, \end{cases}$$

(27.10)

where $x = \nu_1 f/(\nu_1 f + \nu_2)$ and

$$P(\nu_1, \nu_2, x) = \frac{1}{2} x^{\nu_1/2} \sum_{k=0}^{(\nu_2/2)-1} \frac{((\nu_1/2) + k - 1)!}{((\nu_1/2) - k)!k!} \{2(1-x)\}^k,$$

$$R(\nu_1, \nu_2, x) = \sqrt{\frac{1}{2} x(1-x)} \sum_{k=0}^{(\nu_1/2)-1} \frac{k!}{\left(k + \frac{1}{2}\right)!} x^k + \tan^{-1}\text{id\$\&))}.$$

[If y is fractional, one would use $y! = \Gamma(y+1)$.] As ν_2 increases, the value of $\chi^2_{\nu_2}/\nu_2$ tends to 1 with probability 1. If ν_1 remains constant, the distribution of F_{ν_1,ν_2} tends to that of $\chi^2_{\nu_1}/\nu_1$ as ν_2 tends to infinity.

Let $U_n = \chi^2_\nu/(\chi^2_n/n)$, where χ^2_ν and χ^2_n are mutually independent chi-square variables with ν and n degrees of freedom, respectively. Note that U_n/ν has a F-distribution with ν, n degrees of freedom. Then, with $G(x)$ and $g(x)$ denoting the distribution and the density functions of χ^2_ν (see Chapter 18), Fujikoshi (1987) has shown that

$$\begin{aligned}\Pr[U_n \le x] = G(x) - g(x) &\times \left[\frac{1}{n}\left(\frac{x^2}{2} - \frac{x}{2}(\nu - 2)\right)\right.\\ &+ \frac{1}{n^2}\left\{\frac{x^4}{16} - \frac{x^3}{48}(9\nu - 2) + \frac{x^2}{48}(\nu - 2)(9\nu - 4)\right.\\ &\left.\left. - \frac{x}{16}(\nu - 2)(\nu - 4)(3\nu - 2)\right\} + \cdots\right].\end{aligned}$$

Further, by considering the transformed random variable

$$V_n = \left\{n + \frac{1}{2}(\nu - 2)\right\}\log\left(1 + \frac{U_n}{n}\right),$$

where $n > \max\{0, (2-\nu)/2\}$, Fujikoshi and Mukaihata (1993) have shown that

$$\Pr[V_n \le x] = G(x) + O(n^{-2})$$

for all real x. (These authors have also presented some approximations and bounds for the quantiles of the distribution of V_n.)

For purposes of approximation it is often convenient to study the **distribution** of $z_{\nu_1,\nu_2} = \frac{1}{2}\log F_{\nu_1,\nu_2}$* rather than that of F_{ν_1,ν_2} itself. [This variable was in fact used by Fisher (1924).] The distribution of $\log F_{\nu_1,\nu_2}$ is sometimes called the logarithmic F-distribution. Approximations themselves will be discussed in the next section, but here we give formulas for the moments of z_{ν_1,ν_2}. Dropping the suffices for convenience, we have the moment-generating function

$$E[e^{tz}] = E[F^{t/2}] = \left(\frac{\nu_2}{\nu_1}\right)^{t/2} \frac{\Gamma(\frac{1}{2}(\nu_1 + t))\Gamma(\frac{1}{2}(\nu_2 - t))}{\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}\nu_2)}. \quad (27.11)$$

The cumulants of z are

$$\kappa_1(z) = \frac{1}{2}\left[\log\left(\frac{\nu_2}{\nu_1}\right) + \psi\left(\frac{1}{2}\nu_1\right) - \psi\left(\frac{1}{2}\nu_2\right)\right], \quad (27.12a)$$

$$\kappa_r(z) = 2^{-r}\left[\psi^{(r-1)}\left(\frac{1}{2}\nu_1\right) + (-1)^r\psi^{(r-1)}\left(\frac{1}{2}\nu_2\right)\right], \quad r \geq 2, \quad (27.12b)$$

where $\psi(z) = (d/dz)\log\Gamma(z)$ is the **digamma** function. Note that all moments of z are finite. For $r \geq 2$ an alternative formula is

$$\kappa_r(z) = (r-1)!\sum_{j=0}^{\infty}\left[(-1)^r(\nu_1 + 2j)^{-r} + (\nu_2 + 2j)^{-r}\right]. \quad (27.12c)$$

There are also a number of special expressions for $\kappa_r(z)$ depending on the parity of ν_1 and ν_2. We now give these in summary form [see Aroian (1941); Wishart (1947)].

1. ν_1 and ν_2 both *even*:

$$E[z] = \frac{1}{2}\left[\log\left(\frac{\nu_2}{\nu_1}\right) - \sum_j{}^* j^{-1}\right], \quad (27.13a)$$

where $\sum_j^*$ denotes summation from $j = \frac{1}{2}\min(\nu_1, \nu_2)$ to $(\frac{1}{2}\max(\nu_1, \nu_2) - 1)$ inclusive and $\sum_j^* j^{-1} = 0$ if $\nu_1 = \nu_2$.

$$\operatorname{var}(z) = 0.822467 - \frac{1}{4}\left(\sum_{j=1}^{(\nu_1/2)-1} j^{-2} + \sum_{j=1}^{(\nu_2/2)-1} j^{-2}\right) \quad (27.13b)$$

*Although this is a random variable, it is usually denoted in the literature by a small z

and generally (for $r \geq 2$)

$$\kappa_r(z) = 2^{-r}(r-1)!\left[\left(S_r - \sum_{j=1}^{(\nu_1/2)-1} j^{-r}\right) + (-1)^r\left(S_r - \sum_{j=1}^{(\nu_2/2)-1} j^{-r}\right)\right] \tag{27.13c}$$

where

$$S_r = \sum_{j=1}^{\infty} j^{-r}.$$

2. *ν_1 and ν_2 both odd:*

$$\kappa_r(z) = (r-1)!\left[\left(T_r - \sum_{j=0}^{(\nu_2-3)/2} (2j+1)^{-r}\right) + (-1)^r\left(T_r - \sum_{j=0}^{(\nu_1-3)/2} (2j+1)^{-r}\right)\right] \tag{27.14}$$

for $r \geq 2$, where

$$T_r = S_r(1-2^{-r}) = \sum_{j=0}^{\infty} (2j+1)^{-r}.$$

In particular,

$$\text{var}(z) = 2.467401 - \frac{1}{4}\left(\sum_{j=0}^{(\nu_2-3)/2} \left(j+\tfrac{1}{2}\right)^{-2} + \sum_{j=0}^{(\nu_1-3)/2} \left(j+\tfrac{1}{2}\right)^{-2}\right). \tag{27.15a}$$

Also

$$E[z] = \frac{1}{2}\log\left(\frac{\nu_2}{\nu_1}\right) - \sum_j{}^{*} (2j+1)^{-r}, \tag{27.15b}$$

where the limits of summation in $\sum_j^*$ are *now* from $j = \frac{1}{2}\min(\nu_1, \nu_2) - 1$ to $j = \frac{1}{2}\max(\nu_1, \nu_2) - 3/2$.

3. ν_1 ecen, ν_2 *odd:*

$$\kappa_r(z) = (r-1)!\left[\left(T_r - \sum_{j=0}^{(\nu_2-3)/2}(2j+1)^{-r}\right) + \left(-\tfrac{1}{2}\right)^r\left(S_r - \sum_{j=1}^{(\nu_1/2)-1} j^{-r}\right)\right] \quad \text{for } r \geq 2. \quad (27.16)$$

4. ν_2 *even,* ν_1 *odd:*

$$\kappa_r(z) = (r-1)!\left[2^{-r}\left(S_r - \sum_{j=0}^{(\nu_2/2)-1} j^{-r}\right) + (-1)^r\left(T_r - \sum_{j=0}^{(\nu_1-3)/2}(2j+1)^{-r}\right)\right]$$

$$\text{for r} \geq 2. \quad (27.17)$$

It is helpful, in remembering these formulas, to think of the relation

$$\kappa_r(z) = 2^{-r}\left[\kappa_r\left(\log\chi^2_{\nu_1}\right) + (-1)^r\kappa_r\left(\log\chi^2_{\nu_2}\right)\right] \quad \text{for } r \geq 2. \quad (27.17)'$$

3 ORDER STATISTICS

Order statistics $G'_1 \leq \cdots \leq G'_n$ for random samples of n from the G_{ν_1,ν_2} distribution (27.3) were considered by Patil, Raghunandanan, and Lee (1985a, b). Tables of $E[G'_i]$, $E[G'^2_i]$, and $E[G'_iG'_j]$ were provided for n = 2(1)5, $\nu_1 = 2(1)4$, and $\nu_2 = 5(1)7$. Expressions for the pdf and cdf are quite complicated even in the relatively simple case of even degrees of freedom. The case of odd degrees of freedom presents substantial computational difficulties.

4 TABLES

In view of relation (27.8), tables of the incomplete beta function ratio can be used to evaluate the cumulative distribution function of F_{ν_1,ν_2}. In suitable cases tables of binomial (or negative binomial) probabilities can of course be used. This has been noted by a number of authors [e.g., Bizley (1950); Johnson (1959); Mantel (1966)]. Similarly from tables of percentile points of the beta distribution, corresponding points of F-distributions can be obtained with little computational effort. It is, however, convenient to have available tables giving percentile points of F-distributions (i.e., values of $F_{\nu_1,\nu_2,\alpha}$) directly for given values of ν_1, ν_2, and α. Such tables are commonly available. Here we give only the more notable sources, excluding reproductions (in part or whole) in textbooks. It is to be noted that it is customary to give only upper percentiles (i.e., $\alpha \geq 0.5$). Lower percentiles are easily obtained from the formula

$$F_{\nu_1,\nu_2,1-\alpha} = F^{-1}_{\nu_2,\nu_1,\alpha}$$

Greenwood and Hartley (1961) classified tables of $F_{\nu_1,\nu_2,1-\alpha}$ according to two broad categories:

1. With $\alpha = 0.005, 0.001, 0.025, 0.05, 0.1$, and 0.25.
2. With $\alpha = 0.001, 0.01, 0.05$, and 0.20.

Merrington and Thompson (1943) gave tables of the first type to five significant figures for $\nu_1 = 1(1)10, 12, 15, 20, 24, 30, 40, 60, 120, \infty$ and $\nu_2 = 1(1)30, 40, 60, 120, \infty$. (The higher values are chosen to facilitate harmonic interpolation with $120\nu^{-1}$ as the variable.) Fisher and Yates (1953) gave tables of the second type to two decimal places for $\nu_1 = 1(1)6, 8, 12, 24, \infty$ and $\nu_2 = 1(1)30, 40, 60, 120, \infty$. More extensive tables (though only to three significant figures) have been given by Hald (1952). He gave values of $F_{\nu_1,\nu_2,1-\alpha}$ for $\alpha = 0.0005, 0.0001, 0.1, 0.3$, and 0.5 with

$$\nu_1 = 1(1)10(5)20, 30, 50, 100, 200, 500, \infty$$

and

$$\nu_2 = 1(1)20(2)30(5)50, 60(20)100, 200, 500, \infty,$$

and also for

$$\alpha = 0.005, 0.01, 0.025, 0.05$$

with

$$\nu_1 = 1(1)20(2)30(5)50, 60(2)100, 200, 500, \infty$$

and

$$\nu_2 = 1(1)30(2)50(5)70(10)100(25)150, 200, 300, 500, 1000, \infty.$$

Many other tables of $F_{\nu_1,\nu_2,\alpha}$ are derived from one or more of these tables. However, the earliest tables (as has already been mentioned in Section 27.2) gave values not of $F_{\nu_1,\nu_2,\alpha}$ but of $z_{\nu_1,\nu_2,\alpha} = \frac{1}{2}\log F_{\nu_1,\nu_2,\alpha}$. For large values of ν_1 or ν_2, interpolation with respect to numbers of degrees of freedom is much easier for z than for F. This does not appear to be the original reason for introducing z [see Fisher (1924)], but it is now the main reason for using such tables [Fisher and Yates (1953)]. Of course for occasional interpolation it is possible to use tables of the F-distribution and to calculate the needed values of

$$z_{\nu_1,\nu_2,\alpha} = \tfrac{1}{2}\log F_{\nu_1,\nu_2,\alpha}.$$

Zinger (1964) has proposed the following method for interpolating in tables of percentile points of F_{ν_1,ν_2} in order to evaluate $\Pr[F_{\nu_1,\nu_2} > K]$. One seeks to find tabled values $F_{\nu_1,\nu_2,1-\alpha}$ and $F_{\nu_1,\nu_2,1-m\alpha}$ such that $F_{\nu_1,\nu_2,1-m\alpha} \leq K \leq F_{\nu_1,\nu_2,1-\alpha}$. Then

$$\Pr\left[F_{\nu_1,\nu_2} > K\right] \doteqdot 1 - \alpha m^k,$$

where k satisfies the equation

$$kF_{\nu_1,\nu_2,1-m\alpha} + (1-k)F_{\nu_1,\nu_2,1-\alpha} = K.$$

Laubscher (1965) has given examples showing the accuracy of harmonic interpolation (with arguments ν_1^{-1}, ν_2^{-1}) for either univariate or bivariate interpolation with respect to ν_1 and ν_2, where α is fixed. Bol'shev, Gladkov, and Shcheglova (1961) presented auxiliary tables for accurate computations of the beta and z-distribution functions.

Mardia and Zemroch (1978) compiled tables of F and related distributions and gave algorithms for their evaluation. Their tables include values of $F_{\nu_1,\nu_2,\alpha}$ to five significant figures for

$$\nu_1 = 0.1(0.1)1.0(0.2)2.0(0.5)5(1)16, 18, 20, 24, 30, 40, 60, 120, \infty,$$

$$\nu_2 = 0.1(0.1)3.0(0.2)7.0(0.5)11(1)40, 60, 120, \infty,$$

$$\alpha = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.025, 0.03(0.01)0.1, 0.2, 0.25, 0.3, 0.4, 0.5.$$

The inclusion of fractional values of ν_1 and ν_2 is useful when an F-distribution is used as an approximation.

5 APPROXIMATIONS AND NOMOGRAMS

Because of Eq. (27.8), that expresses the probability integral of the F-distribution as an incomplete beta function ratio, approximations to the latter can be applied to the former. Such approximations have already been described in Chapters 3 and 25. Here we will describe more particularly some approximations to the F-distribution. They also can be used as approximations to the incomplete beta function ratio. In fact this section may be regarded as an extension of the discussions of Chapter **3**, Section **6**, and Chapter **25**, Section 6.

The approximations cited below for are quite ingenious for the most part and historically valuable, and we have decided to include them to mirror the developments in this field of investigation realizing fully well that practical utility of a number of them becomes less prominent with the advances in computer technology.

We have already noted that $z = \frac{1}{2}\log F$ has a more nearly normal distribution than does F itself. Several approximations are based on either a normal approximation to the distribution of z or some modification thereof —such as that provided by using a Cornish-Fisher expansion (Chapter **12**, Section 5).

For large values of both ν_1 and ν_2 the distribution of z may be approximated by a normal distribution with expected value $\frac{1}{2}(\nu_2^{-1} - \nu_1^{-1})\ (= \delta)$ and variance $\frac{1}{2}(\nu_1^{-1} + \nu_2^{-1})\ (= \sigma^2)$. This leads to the simple approximate formula

$$z_{\nu_1,\nu_2,\alpha} \doteqdot \tfrac{1}{2}(\nu_2^{-1} - \nu_1^{-1}) + U_\alpha\sqrt{\tfrac{1}{2}(\nu_1^{-1} + \nu_2^{-1})} \qquad (= \delta + U_\alpha\sigma), \quad (27.18)$$

suggested by Fisher **(1924)**. Fisher also suggested that replacement of ν_1^{-1}, ν_2^{-1} by $(\nu_1 - 1)^{-1}, (\nu_2 - 1)^{-1}$ might improve accuracy.

More elaborate approximations can be obtained using expansions of Cornish-Fisher **(1937)** type. One such approximation [Aroian (1941); Fisher and Cornish (1960); Wishart (1957)], using ***approximate*** formulas for the cumulants, is

$$\begin{aligned} z_{\nu_1,\nu_2,\alpha} \doteqdot\ & U_\alpha\sigma + \frac{1}{3}\delta(U_\alpha^2 + 2) \\ & + \sigma\left\{\frac{\sigma^2}{12}(U_\alpha^3 + 3U_\alpha) + \frac{1}{36}\left(\frac{\delta}{\sigma}\right)^2(U_\alpha^3 + 11U_\alpha)\right\} \\ & + \frac{1}{30}\delta\sigma^2(U_\alpha^4 + 9U_\alpha^2 + 8) \\ & - \frac{1}{810}\frac{\delta^3}{\sigma^2}(3U_\alpha^4 + 7U_\alpha^2 - 16) + \cdots . \end{aligned} \qquad (27.19)$$

The first two terms of approximation **(27.19)** can be written

$$\delta + U_\alpha\sigma + \frac{1}{3}\delta(U_\alpha^2 - 1) = \delta\left[1 + \frac{1}{3}(U_\alpha^2 - 1)\right] + U_\alpha\sigma.$$

Fisher *(1924)* suggested the approximation

$$z_{\nu_1,\nu_2,\alpha} \doteqdot \delta\left[1 + \frac{1}{3}(U_\alpha^2 - 1)\right] + U_\alpha\sigma(1 - \sigma^2)^{-1/2}, \qquad (27.20)$$

which makes some allowance for later terms in *(27.19).* This approximation was improved by Cochran *(1940)* with the formula

$$z_{\nu_1,\nu_2,\alpha} \doteqdot \delta\left[1 + \frac{1}{3}(U_\alpha^2 - 1)\right] + U_\alpha\sigma\left[1 - \frac{1}{6}(U_\alpha^2 + 3)\sigma^2\right]^{-1/2}, \quad (27.21)$$

which differs from the sum of the first three terms of *(27.19)* by approximately $[(U_\alpha^3 + 11U_\alpha)/36](\delta^2/\sigma)$.

Carter *(1947)* obtained another formula using more accurate expressions for the cumulants of z derived by Wishart, together with certain modifications in the expansions. His formula for ν_1, ν_2 large is

$$z_{\nu_1,\nu_2,\alpha} \doteqdot \frac{U_\alpha\left\{(U_\alpha^2/6) - (1/2) + 2\nu_1'\nu_2'(\nu_1' + \nu_2')^{-1}\right\}^{1/2}}{2\nu_1'\nu_2'(\nu_1' + \nu_2')^{-1}}$$
$$- \frac{1}{6}\left(\frac{1}{\nu_1'} - \frac{1}{\nu_2'}\right)\left(U_\alpha^2 + 2 - \frac{2(\nu_1' + \nu_2')}{\nu_1'\nu_2'}\right), \qquad (27.22)$$

where

$$\nu_j' = \nu_j - 1, \qquad j = 1, 2.$$

We may also note the following approximate result of Aroian (1942): For ν_1 and ν_2 large, he showed

$$z_{\nu_1,\nu_2}\frac{\sqrt{2(\nu_1 + \nu_2 - 1)\nu_1\nu_2}}{\nu_1 + \nu_2}$$

to be approximately distributed as t with $(\nu_1 + \nu_2 - 1)$ degrees of freedom. Aroian also proposed the approximation

$$\Pr\left[z_{\nu_1,\nu_2}\frac{\sqrt{2(\nu_1 + \nu_2 - 1)}}{\nu_1 + \nu_2} > t_{\nu_1+\nu_2-1,1-\alpha}\right]$$
$$\doteqdot \alpha\sqrt{\frac{(\nu_1 + \nu_2 - 1)}{\nu_1 + \nu_2}}\exp\left[\frac{1}{6}(\nu_1^{-1} + \nu_2^{-1}) - \frac{5}{12}(\nu_1 + \nu_2 - 1)^{-1}\right],$$
$$(27.23)$$

where $t_{\nu_1+\nu_2-1,1-\alpha}$ is the *100 (1 – α)%* point of t distribution with $(\nu_1 + \nu_2$

− 1) degrees of freedom. The right-hand side is always less than α. The approximation is reported to be generally not as good as the Cornish-Fisher type formula (27.21).

These formulas were further modified by Aroian (1947), who gave a number of interesting numerical comparisons of the accuracy of various approximations. His conclusions were that a better approximation, than that provided by expansions of the type we have so far discussed [except (27.22)], is provided (at least for $\nu_1, \nu_2 > 20$), by an approximation suggested by Paulson (1942).

Paulson's formula is based on the Wilson-Hilferty approximation to the distribution of χ^2, the remarkable accuracy of which has been described in Chapter 18, Section 5. If the distributions of X_1 and X_2 (at the beginning of this chapter) are each approximated in this way, we see that the distribution of $F_{\nu_1,\nu_2}^{1/3}$ is approximated by the distribution of the ratio of two independent normal variables. In fact $F_{\nu_1,\nu_2}^{1/3}$ is approximately distributed as

$$\frac{1 - \frac{2}{9}\nu_1 + U_1\sqrt{\frac{2}{9}\nu_1}}{1 - \frac{2}{9}\nu_2 + U_2\sqrt{\frac{2}{9}\nu_2}}, \tag{27.24}$$

where U_1, U_2 are independent unit normal variables. Using a further approximate formula for the distribution of this ratio (Chapter 13, Section 6.3), we are led to the approximation of taking

$$W = \left[\left(1 - \frac{2}{9\nu_2}\right)F_{\nu_1,\nu_2}^{1/3} - \left(1 - \frac{2}{9\nu_1}\right)\right]\left(\frac{2}{9\nu_2}F_{\nu_1,\nu_2}^{2/3} + \frac{2}{9\nu_1}\right)^{-1/2} \tag{27.25}$$

to have a unit normal distribution [Paulson (1942)]. This approximation is also remarkably accurate for $\nu_2 \geq 10$. Smillie and Anstey (1964) utilized it in a computer routine.

For $\nu_2 \leq 10$ values of upper percentage points $F_{\nu_1,\nu_2,\alpha}$ calculated from (27.24) can be improved by using the formula

$$\text{(Improved value)} = m \times \text{(Calculated value)} + c$$

where m and c depend on ν_2 and the percentile but not on ν_1. Ashby (1968) gave values of m and c for $\alpha = 0.95, 0.99, 0.999$; $\nu_2 = 1(1)10$ with which accuracy to three significant figures is attained. For small values of $\nu_2 \leq 3$ Kelley (1948) recommended replacing W by

$$W' = W(1 + 0.08W^4\nu_2^{-3}).$$

One of the earliest computer programs for calculating $\Pr[F > f]$ using this corrected formula was published by Jaspen (1965). He gave comparisons of exact and calculated values for $\nu_1 = 2$, $\nu_2 = 2$. Further comparisons, for

$\nu_1, \nu_2 = 1, 2, 4, 10, 20, 60, 1000$ but only for $f = F_{\nu_1, \nu_2, 0.95}$, were made by Golden, Weiss, and Davis (1968). It appears from Jaspen (1965) that accuracy is better in the lower rather than the upper tail of the F-distribution.

A similar approximation can be obtained using Fisher's approximation to the χ^2 distribution ($\sqrt{2\chi_\nu^2} - \sqrt{2\nu - 1}$ approximately a unit normal variable) in place of the Wilson-Hilferty approximation.

The resulting approximation is to regard

$$\left[\sqrt{1 - \frac{1}{2\nu_2}}\, F_{\nu_1, \nu_2}^{1/2} - \sqrt{1 - \frac{1}{2\nu_1}}\right]\left[\frac{1}{2\nu_2} F_{\nu_1, \nu_2} + \frac{1}{2\nu_1}\right]^{-1/2} \tag{27.26}$$

as a unit normal variable. Since the Fisher transformation is generally less accurate than the Wilson-Hilferty transformation, one would expect (27.25) to be generally more accurate than (27.26). While this is so, the comparison is not as disadvantageous to (25.26) as might be expected.

If only one of ν_1 and ν_2 is large (e.g., ν_2), then the natural approximation to use is F_{ν_1, ν_2} distributed as $\chi_{\nu_1}^2/\nu_1$. (It is clearly always possible to arrange that $\nu_2 > \nu_1$ in this case.) Scheffé and Tukey (1944) have proposed a simple improvement on this which, in terms of the incomplete beta function ratio [Chapter 25, Eq. (27.3)], is that if

$$I_p(n - r + 1, r) = \alpha,$$

then

$$n \doteqdot \tfrac{1}{4}\chi_{2r, \alpha}^2 (1 + p)(1 - p)^{-1} + \tfrac{1}{2}(r - 1).$$

In terms of F-distributions this can be stated as

$$F_{\nu_1, \nu_2, \alpha} \doteqdot \left[\frac{\nu_1}{\chi_{\nu_1, \alpha}^2} + \frac{\nu_1}{\nu_2}\left(\frac{(\nu_1/2) - 1}{\chi_{\nu_1, \alpha}^2} - \frac{1}{2}\right)\right]^{-1}. \tag{27.27}$$

(At one stage the derivation of this entails reversing the degrees of freedom of F.) McIntyre and Ward (1968) constructed a computer program using this approximation.

Mudholkar and Chaubey (1976) utilize the relation between F- and beta distributions to adjust their Patnaik-type, Pearson-type, and Sankaran-type approximations to $I_x(p, q)$ (see Chapter 25, Section 6.1) for approximating percentage points of F-distributions:

$$\Pr\left[F_{\nu_1, \nu_2} \geq f\right] \doteqdot 1 - \Phi(u),$$

where

$$u = \frac{[(-\log x + b)/(\kappa_1 - b)]^{1/3} - [1 - \kappa_3^2/(36\kappa_2^3)]}{\kappa_3/(6\kappa_2^{3/2})} \qquad (27.28)$$

and $x = \{1 + (\nu_1/\nu_2)f\}^{-1}$. Here κ_1, κ_2, and κ_3 are the cumulants of $(-\log X)$, where X has a standard beta $(\frac{1}{2}\nu_2, \frac{1}{2}\nu_1)$ distribution:

$$\kappa_r = (-1)^r[\psi^{(r-1)}(\tfrac{1}{2}) - \psi^{(r-1)}(\tfrac{1}{2}(\nu_1 + \nu_2))],$$

where

$$\psi^{(r-1)}(x) = \left(\frac{d^r}{dx^r}\right)\log \Gamma(x)$$

[see Chapter 25, Eq. (25.13)‴]. If ν_1 is odd, an approximate "interpolated" expression,

$$\kappa_r^* \doteqdot (r-1)!\left\{\sum_{j=0}^{(\nu_1-3)/2}\left(\frac{1}{2}\nu_2 + j\right)^{-r} + \frac{1}{2}\left(\frac{\nu_1+\nu_2-1}{2}\right)^{-r}\right\}, \qquad (27.29)$$

is used for κ_r. If both ν_1 and ν_2 are even, the formula

$$\Pr[F_{\nu_1,\nu_2} \le f] = I_{\nu_1 f/(\nu_2+\nu_1 f)}(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2) \qquad (27.8)'$$

enables one to calculate values of the cdf of F_{ν_1,ν_2} as a sum of probabilities for a binomial distribution with parameters $\frac{1}{2}(\nu_1 + \nu_2) - 1, \nu_1 f/(\nu_2 + \nu_1 f)$. However, if ν_1 or ν_2 (or both) are odd, this is not possible. For such cases George and Singh (1987) suggested an approximate formula of form

$$\Pr[F_{\nu_1,\nu_2} \le f] \doteqdot I_{\alpha+\beta\log f}([\tfrac{1}{2}(\nu_1 + 1)], [\tfrac{1}{2}(\nu_2 + 1)]), \qquad (27.30a)$$

where $[a]$ denotes "integer part of a," and α and β are constants depending on ν_1 and ν_2. [Formula (27.30a) was based on approximating the distribution of log F_{ν_1,ν_2} by a generalized logistic distribution.]

Earlier Mantel (1966) had suggested that when ν_1 and/or ν_2 are odd, values of $I_p(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)$ can be interpolated from neighboring values for even ν_1 and ν_2—for example, if $\nu_1 = 3$ and $\nu_2 = 5$—the values $I_p(1,2)$, $I_p(1,3)$, $I_p(2,2)$, and $I_p(2,3)$ can be used to estimate $I_p(1.5, 2.5)$. [Note that in each case $p = \nu_1 f/(\nu_2 + \nu_1 f)$ $(= 3f/(5 + 3f)$ here).] George and Singh (1987) found that when both ν_1 and ν_2 are odd, (27.30a) gives superior results, especially in the tails of the distribution, but when only one of ν_1 and ν_2 is odd, Mantel's approximation is slightly better.

If $(-\log X)$ is approximated as $(a\chi_\nu^2 + b)$, then ν is determined by equating the first three moments of $(-\log X)$ and $(a\chi_\nu^2 + b)$, obtaining

$$\nu = 8\{\beta_1(-\log X)\}^{-1},$$

$$a = \left(\frac{\text{var}(-\log X)}{2\nu}\right)^{1/2},$$

$$b = E[-\log X] - a\nu. \tag{27.30b}$$

Davenport and Herring (1979) have suggested a so-called adjusted Cornish-Fisher approximation for $F_{\nu_1,\nu_2,\alpha}$. Johnson (1973) devised a number of empirical formulas for $F_{\nu_1,\nu_2,\alpha}$, for $\alpha = 0.95, 0.975$, which are set out in Table 27.1. The accuracy is $\pm 0.6\%$.

In most cases the error is less than 0.2%. The maximum error occurs at $\nu_1 = \nu_2 = 120$ for $\alpha = 0.95$ and $\nu_1 = 60$, $\nu_2 = 120$ for $\alpha = 0.975$. Ojo (1985, 1988) provides approximations to the distribution of log F by means of t-distributions, having the same mean, variance, and coefficient of kurtosis as a linear function of log F. Viveros (1990) considers power transformations. He suggests regarding

$$\left(F^c_{\nu_1,\nu_2} - d\right)g \tag{27.30c}$$

as a unit normal variable, with

$$c = \frac{\nu_2 - \nu_1}{3(\nu_1 + \nu_2) - 4},$$

$$d = \left\{\frac{\nu_2(\nu_1 - 2c)}{\nu_1(\nu_2 + 2c)}\right\}^c,$$

$$g = \frac{1}{cd}\left\{\frac{(\nu_1 - 2c)(\nu_2 + 2c)}{2(\nu_1 + \nu_2)}\right\}^{1/2}.$$

These values are obtained by considering the Taylor expansion of the pdf of log F about the mode of the distribution. Alternatively, for any given c, d and g can be chosen as the expected value and standard deviation of F^c, producing a standardized variable in (27.30c). Note that $-\frac{1}{3} \le c \le \frac{1}{3}$.

Nomograms for evaluation of incomplete beta function values can be used for calculation of values of cumulative F-distributions. There is a nomogram specifically designed for the F-distribution in Stammberger (1967) which is reproduced in Figure 27.2. From this nomogram one can evaluate any one of $\Pr[F_{\nu_1,\nu_2} \le f]$, f, ν_1, and ν_2 given the values of the other three quantities, using two straight edges (indicated by broken lines in Figure 27.2).

Table 27.1 Johnson's empirical formulas for $F_{\nu_1,\nu_2,\alpha}$ for a = 0.95, 0.975, and their accuracy

Group	ν_1	2	Approximate $F_{\nu_1,\nu_2,0.95}$	Maximum Absolute % Deviation from $F_{\nu_1,\nu_2,0.95}$
I	1–120	1	$\dfrac{\nu_1 - 0.09849}{0.0039292\nu_1 + 0.0016579}$	0.005
II	1–120	2	$\dfrac{\nu_1 - 0.03646}{0.051294\nu_1 + 0.000761}$	0.000
III	1–120	3	$\dfrac{\nu_1 + 1.094}{0.1173\nu_1 + 0.0894}$	0.005
IV	1–120	4	$\dfrac{\nu_1 + 1.349}{0.1776\nu_1 + 0.1271}$	0.009
V	1	5–120	$7.71 - \dfrac{\nu_2 - 4.032}{0.2581\nu_2 - 0.4076}$	0.000
VI	2–120	5–120	$\dfrac{\nu_1 + 1.288}{0.1751\nu_1 + 0.1129} - \dfrac{\nu_2 - 4.119}{0.2511\nu_2 - 0.4236} \times \left\{0.552 - \dfrac{6.530}{\nu_1 + 11.533} - \dfrac{3.993}{\nu_2 + 11.533} + \dfrac{88.889}{(\nu_1 + 11.533)(\nu_2 + 11.533)}\right\}$	0.6

Group	ν_1	ν_2	Approximate $F_{\nu_1,\nu_2,0.975}$	Maximum Absolute % Deviation from $F_{\nu_1,\nu_2,0.975}$
I	1–120	1	$\dfrac{\nu_1 - 0.09582}{0.0009813\nu_1 + 0.0004153}$	0.001
II	1–120	2	$\dfrac{\nu_1 - 0.00904}{0.025317\nu_1 + 0.000416}$	0.001
III	1–120	3	$\dfrac{\nu_1 + 0.9232}{0.07192\nu_1 + 0.03836}$	0.003
IV	1–120	4	$\dfrac{\nu_1 + 1.270}{0.1210\nu_1 + 0.0648}$	0.003
V	1	5–120	$12.22 - \dfrac{\nu_2 - 4.045}{0.1387\nu_2 - 0.2603}$	0.003
VI	2–120	5–120	$\dfrac{\nu_1 + 1.739}{0.1197\nu_1 + 0.1108} - \dfrac{\nu_2 - 3.986}{0.1414\nu_2 - 0.2864} \times \left\{-0.145 + 0.00170\nu_1 + \dfrac{2.706 - 0.06150\nu_l}{\nu_2 + 30}\right\}$	0.6

Source: Johnson (1973).

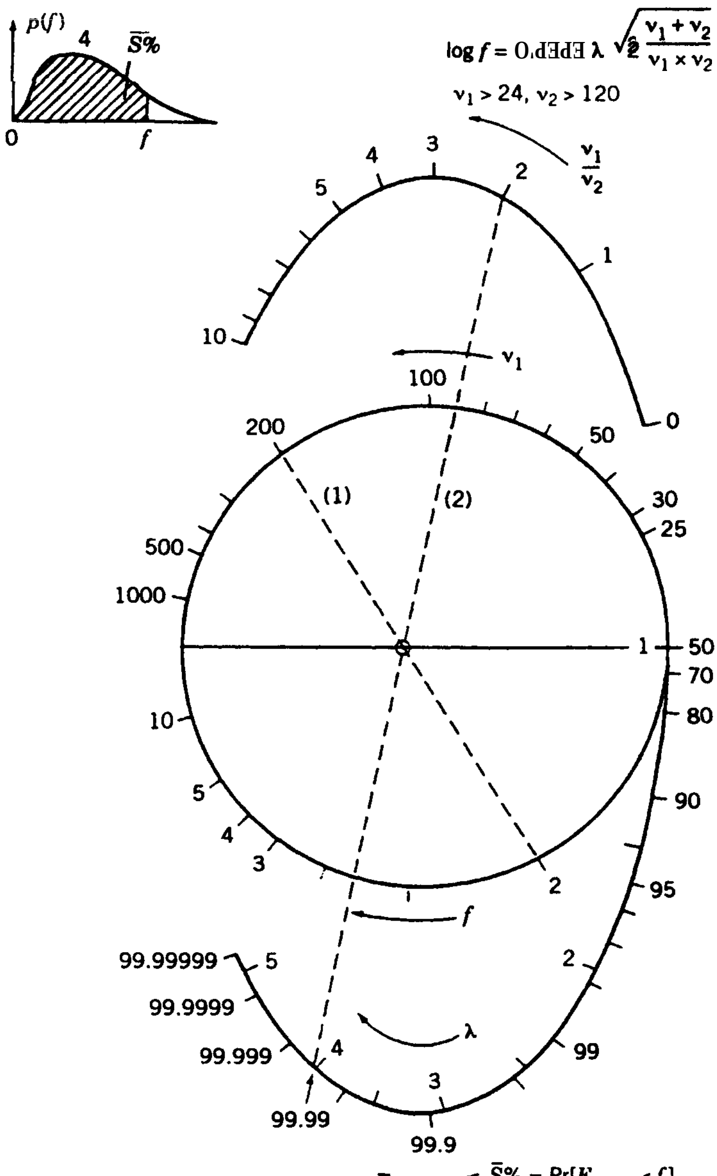


Figure 27.2 Nomogram for Cumulative F Distribution

Dion and Fridshal (1982) conjectured that

$$\frac{\chi^2_{\nu_1,(1-\gamma)/2}/\nu_1}{\chi^2_{\nu_2,(1-\gamma)/2}/\nu_2} \le F_{\nu_1,\nu_2,1-\gamma} \tag{27.31a}$$

and

$$\frac{\chi^2_{\nu_1,(1+\gamma)/2}/\nu_1}{\chi^2_{\nu_2,(1+\gamma)/2}/\nu_2} \ge F_{\nu_1,\nu_2,1+\gamma}. \tag{27.31b}$$

Burk et al. (1984) showed that this conjecture is true for $\nu_1 = \nu_2$. However, numerical studies showed that the conjecture is false if $\nu_1 \neq \nu_2$, for all γ in an interval $0 \le \gamma \le \gamma_0$, together with some "holes" in $\gamma_0 < \gamma < 1$. ($\gamma_0 \approx 0.0043$ for $\nu_1 = 4$, $\nu_2 = 2$.)

6 APPLICATIONS

The commonest application of the F-distribution in statistical work is in standard tests associated with the analysis of variance. Many of these are based on a general result of Kolodziejczyk (1935), according to which the likelihood ratio test of a general linear hypothesis (H_0) about the parameters of a general linear model with normal residuals can be expressed in terms of a statistic that has an F-distribution when H_0 is valid. The application of the F-distribution in testing equality of variances of two normal populations has been described in Section 27.1.

In some earlier tables of (upper) percentiles of F-distributions, the tables are stated to apply to the distribution of "greater divided by lesser mean square." While it is true that it is often convenient to take the ratio so that it exceeds 1, this affects the significance level of the test. If tables of upper $100\alpha\%$ percentiles are used ($F_{\nu_1,\nu_2,1-\alpha}$), then the actual significance level is not α but 2α. This can be seen by noting that "significance" is attained if the observed ratio is either greater than $F_{\nu_1,\nu_2,1-\alpha}$ or less than $(F_{\nu_2,\nu_1,1-\alpha})^{-1} = F_{\nu_1,\nu_2,\alpha}$.

The F-distribution is also used in the calculation of power functions of these tests and of confidence limits for the ratios of variances of normal populations. The relation between F-distribution and binomial distributions has already been noted in Section 27.3. A consequence of this is that the approximate confidence limits $(\underline{p}, \bar{p})$ for a binomial proportion obtained by solving the equations

$$\sum_{j=r}^{n}\binom{n}{j}\underline{p}^j(1-\underline{p})^{n-j} = \tfrac{1}{2}\alpha$$

$$\sum_{j=0}^{r}\binom{n}{j}\bar{p}^j(1-\bar{p})^{n-j} = \tfrac{1}{2}\alpha$$

[where r represents the observed value of a binomial (n, p) variable] can be expressed in terms of percentile points of an F-distribution:

$$\underline{p} = r\left[(n - r + 1)F_{2(n-r+1), 2r, \alpha/2} + r\right]^{-1},$$

$$\bar{p} = (r + 1)F_{2(r+1), 2(n-r), \alpha/2}\left[n - r + (r + 1)F_{2(r+1), 2(n-r), \alpha/2}\right]^{-1}.$$

Box (1949) expresses certain approximations to the distributions (under multinormality conditions) of multivariate test statistics in terms of F-distributions. These are just convenient ways of representing the results of fitting Pearson Type VI distributions. This family of distributions will be discussed in Section 27.7. Donner, Wells, and Eliasziw (1989) describe uses of F-distributions as approximations to distributions of certain statistics in analysis of variance for unbalanced designs [see also Satterthwaite (1946)].

The F-distribution is also used for approximating other distributions. Yip (1974) approximated the distribution of a random variable X using the first four moments. He equated the first four moments of $(X + g)/h$ to those of F_{ν_1, ν_2} [see Eq. (27.6)] and solved the resulting equations for ν_1, ν_2, g, and h. Yip (1974) also applied this method to obtain approximations of noncentral F-distribution (Chapter 30) and that of Hotelling's generalized T_0^2.

Four-moment approximations can of course be obtained from tables of the percentage points of Pearson curves, but these tables do not cover all the (β_1, β_2) points in the F-region. Whenever the (β_1, β_2) values of X lie within this F-region, the four-moment F-distribution $F \simeq (X + g)/h$ will generally provide close approximations to the distribution of X.

Wood (1989) used a three-parameter F-distribution as an approximation to the distribution of a positive linear combination of central chi-square variables. Numerical results indicate that the proposed approximation is superior over Satterthwaite's (1946) and Buckley and Eagleson's (1988) approximations (unless 0.95 or 0.05 point of the linear combination is required), but is less accurate than the iterative Gamma-Weibull approximation due to Solomon and Stephens (1977).

7 PEARSON TYPE VI DISTRIBUTIONS

We have already remarked that any F-distribution is a special form of Pearson Type VI distribution (27.3). Here we give an account of this general family, to which F-distributions belong. The most general form of equation for the probability density function of a Type VI distribution can be written as [Elderton and Johnson (1969)]

$$p_X(x) = \frac{\Gamma(q_1)(a_2 - a_1)^{q_1 - q_2 - 1}}{\Gamma(q_1 - q_2 - 1)\Gamma(q_2 + 1)} \frac{(x - a_2)^{q_2}}{(x - a_1)^{q_1}}$$

$$q_1 > q_2 > -1,\ x \geq a_2 > a_1 \qquad (27.32)$$

[assuming that the sign of X has been chosen so that $\sqrt{\beta_1(X)} \geq 0$].

The graph of $p_X(x)$ against x has a single mode at

$$x = a_2 + q_2(q_1 - q_2)^{-1}(a_2 - a_1)$$

provided that q_2 is positive. If q_2 is zero, the mode is at $x = a_2$, and if q_2 is negative, $p_X(x)$ tends to infinity as x tends to a_2. In the last two cases $p_X(x)$ decreases as x increases from a_2.

The rth moment of X about a_1 is

$$E[(X - a_1)^r] = \frac{(a_2 - a_1)^r (q_1 - 1)^{(r)}}{(q_1 - q_2 - 2)^{(r)}}. \tag{27.33}$$

Note that if $r \geq q_1 - q_2 - 1$, the rth moment is infinite. The expected value and variance of the distribution are

$$\begin{aligned} E[X] &= a_1 + \frac{(q_1 - 1)(a_2 - a_1)}{q_1 - q_2 - 2} \\ &= a_2 + \frac{(q_2 - 1)(a_2 - a_1)}{q_1 - q_2 - 2}, \qquad q_1 - q_2 > 2, \end{aligned} \tag{27.34}$$

$$\operatorname{var}(X) = (a_2 - a_1)^2 (q_1 - 1)(q_2 + 1)(q_1 + q_2 - 2)^{-2}(q_1 - q_2 - 3)^{-1}, \qquad q_1 - q_2 > 3.$$

The four parameters a_1, a_2, q_1, and q_2 can of course be expressed in terms of the first four moments of X. In fact q_2 and $-q_1$ are given by

$$\frac{1}{2}(r - 2) \pm \frac{1}{2} r(r + 2) \sqrt{\frac{\beta_1}{\beta_1(r + 2)^2 + 16(r + 1)}}, \tag{27.35}$$

where $r = -6(\beta_2 - \beta_1 - 1)/(2\beta_2 - 3\beta_1 - 6)$. Note that for Type VI curves, $2\beta_2 - 3\beta_1 - 6 > 0$ so that $r < 0$, and also

$$\frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} > 1. \tag{27.36}$$

There is a simple relation between Type VI and Type I distributions (Chapter 25): The distribution of $Y = (X - a_2)/(X - a_1)$ is of standard beta form, with parameters $q_2 + 1$ and $q_1 - q_2 - 1$. If $X_1, X_2, \cdots, X_n$ are independent random variables, each having a distribution of form (27.32), the formal equations for the maximum likelihood estimators $\hat{a}_1$, $\hat{a}_2$, $\hat{q}_1$, and $\hat{q}_2$

can be written

$$n(\hat{q}_1 - \hat{q}_2 - 1)(\hat{a}_2 - \hat{a}_1)^{-1} = \hat{q}_1 \sum_{j=1}^{n} (x_j - \hat{a}_1)^{-1} = \hat{q}_2 \sum_{j=1}^{n} (x_j - \hat{a}_2)^{-1}, \tag{27.37a}$$

$$n[\psi(\hat{q}_1) - \psi(\hat{q}_1 - \hat{q}_2 - 1) + \log(\hat{a}_2 - \hat{a}_1)] = \sum_{j=1}^{n} \log(x_j - \hat{a}_1), \tag{27.37b}$$

$$n[\psi(\hat{q}_2 + 1) - \psi(\hat{q}_1 - \hat{q}_2 - 1) + \log(\hat{a}_2 - \hat{a}_1)] = \sum_{j=1}^{n} \log(x_j - \hat{a}_2). \tag{27.37c}$$

It is necessary to check that the solutions of (27.37) do satisfy the inequalities $\hat{a}_1 < \hat{a}_2 < \min(x_1, \cdots, x_n)$. It should also be noted that the standard formulas for asymptotic variances do not apply if $q_1 - q_2 \leqq 2$. The standard formula for the asymptotic variance-covariance matrix of $\sqrt{n}\,\hat{a}_1$, $\sqrt{n}\,\hat{a}_2$, $\sqrt{n}\,\hat{q}_1$, and $\sqrt{n}\,\hat{q}_2$ is the inverse of the symmetric matrix

$$\begin{pmatrix} \dfrac{(q_2+1)(q_1-q_2-1)}{(q_1+1)(a_2-a_1)^2} & \dfrac{-(q_1-q_2-1)}{(a_2-a_1)^2} & \cdots & \cdots \\ \dfrac{-(q_1-q_2-1)}{(a_2-a_1)^2} & \dfrac{(q_1-1)(q_1-q_2-1)}{(q_2-1)(a_2-a_1)^2} & \cdots & \cdots \\ \dfrac{q_2+1}{q_1(a_2-a_1)} & \dfrac{q_1-1}{q_2(a_2-a_1)} & \begin{matrix} -\psi'(q_1) + \\ \psi'(q_1-q_2-1) \end{matrix} & \cdots \\ \dfrac{-1}{a_2-a_1} & \dfrac{-1}{a_2-a_1} & -\psi'(q_1-q_2-1) & \begin{matrix} \psi'(q_2+1) + \\ \psi'(q_1-q_2-1) \end{matrix} \end{pmatrix} \tag{27.38}$$

8 OTHER RELATED DISTRIBUTIONS

The relation between F-distributions and Type I (beta) distributions has already been described in Section 2 (and, more generally, between Type VI and Type I in Section 7). A less evident (though more specialized and less useful) relationship can be summarized as follows: The distribution of $\frac{1}{2}\sqrt{\nu}(F_{\nu,\nu}^{1/2} - F_{\nu,\nu}^{-1/2})$ is a t-distribution with ν degrees of freedom. This result has been reported by several different authors [see Aroian (1953), Cacoullos (1965)]. Kymn (1974) provides an elementary derivation.

There is also a relationship (mentioned in Chapter 3, Section 6.1) between the binomial and F-distributions. This can be summarized by the equation

$$\Pr\left[F_{2(n-r+1),2r} > \frac{1-p}{p} \cdot \frac{r}{n-r+1}\right] = \Pr[Y \geq r] \qquad (27.39)$$

(r is an integer, $0 \leq r \leq n$), where Y is a binomial variable with parameters (n, p) [Bizley (1950); Jowett (1963)]. Noncentral F-distributions are discussed in Chapter 30. Multivariate generalizations are discussed in Chapter 40.

There are several pseudo-F-distributions corresponding to replacement of either or both, of S_1 and S_2 in (27.2) by some other sample measure of dispersion such as sample range or mean deviation [see, e.g., Newman (1939)].

The works of David (1949), Gayen (1950), Horsnell (1953), Swain (1965), Tiku (1964), and Zeigler (1965) are the most comprehensive among the numerous investigations of F-distributions under nonnormal conditions, namely when the variables $X_1, X_2, \cdots, X_n$ in the definitions of S_t $(t = 1, 2)$ have nonnormal distribution. Additional references will be indicated in Chapter 28, Section 7, in connection with the similar problem for t-distributions.

The truncated Type VI distribution with density function

$$p_X(x) = \left[\frac{\beta}{\log(1+\beta)}\right](1+\beta x)^{-1}, \qquad 0 < x < 1, \beta > -1 \quad (27.40)$$

(x denotes a proportion) has been used to represent the distribution of references among different sources [Bradford (1948); Leimkuhler (1967)]. When used for this purpose it is called the Bradford distribution. Of course X would be more naturally represented as a discrete variable and the Bradford distribution may be regarded as an approximation to a Zipf or Yule distribution (Chapter 11).

In the analysis of variance one often has a situation wherein the ratios of each of a number of "mean squares" $M_1, M_2, \ldots, M_k$ to a residual mean square M_0 have to be considered. M_j can be regarded as being distributed as $\sigma^2(\chi^2_{\nu_j}/\nu_j)$ $(j = 0, 1, \ldots, k)$ and the M's as mutually independent. If one of the ratios M_j/M_0 $(j \neq 0)$ is large, it is helpful to compare it with the distribution of $\max_{1,\ldots,k}(M_j/M_0)$. In the case where $\nu_1 = \nu_2 = \cdots = \nu_k = \nu$ (not necessarily equal to ν_0), this distribution is related to that of "Studentized" χ^2—namely the ratio of the minimum of k independent χ^2_ν variables to an independent $\chi^2_{\nu_0}$ variable.

Armitage and Krishnaiah (1964) have given tables of the upper 1%, $2\frac{1}{2}\%$, 5%, and 10% points of this distribution, to two decimal places for

$$k = 1(1)12,$$

$$\nu = 1(1)19,$$

$$\nu_0 = 6(1)45 \qquad (\text{also } \nu_0 = 5\% \text{ and } 10\% \text{ points}).$$

Tables of the upper 1% and 5% points of $(\max_{1 \le j \le k} M_j)/(\min_{1 \le j \le k} M_j)$ for $k = 2(1)12$ and $\nu_1 = \nu_2 = \cdots = \nu_k = \nu = 2(1)10, 12, 15, 20, 30, 60, \infty$, are contained in Pearson and Hartley (1958); for additional comments on this, see the end of Section 8.2.

8.1 "Generalized" F-Distributions

Several "generalized" F-distributions have been described in the last two decades. We have already noted Prentice's (1975) work in connection with the z-distribution. A variety of generalizations that have been studied, with differing notations, are included within the general form of distributions of $aF^b_{\nu_1,\nu_2}$, for various values of a and b.

Ciampi, Hogg, and Kates (1986) provide a general discussion of this class of distributions, indicating special values of the four parameters (a, b, ν_1, ν_2) corresponding to certain well-known distributions. They obtain maximum likelihood estimators for the parameters using a "generalized reduced gradient method" described in Lasdon et al. (1978).

If X_1 and X_2 are mutually independent, and X_i has a gamma (α_i, β_i) distribution [Chapter 17, Eq. (17.23)] with the pdf

$$p_{X_i}(x_i) = \{\beta_i^{\alpha_i}\Gamma(\alpha_i)\}^{-1} x_i^{\alpha_i - 1} \exp(-\beta_i^{-1} x_i), \qquad x_i > 0, i = 1, 2, \quad (27.41)$$

then X_1/X_2 is distributed as

$$\frac{\alpha_1\beta_1}{\alpha_2\beta_2} F_{2\alpha_1, 2\alpha_2}.$$

Pham-Gia and Duong (1989) termed this a "corrected" F-distribution and denoted it by G3F. (It is related to the G3B distributions discussed in Chapter 25, Section 7). Dyer (1982) established the distributions of sums of independent variables having G3F distributions, and applied his results to various problems in reliability, multivariate analysis and Bayesian modeling. Shah and Rathie (1974) obtained distributions of products of G3F variables. Amaral-Turkman and Dunsmore (1985) used G3F distributions (which they called *Inbe* distributions) in studies on information in predictive distributions for a gamma model.

If $Y = aF^b_{\nu_1,\nu_2}$ $(a, b > 0)$, then

$$p_Y(y) = \frac{(\nu_1/\nu_2)^{\nu_1/2}(ab)^{-1}(y/a)^{(\nu_1/(2b))-1}}{B((\nu_1/2),(\nu_2/2))\{1+(\nu_1/\nu_2)(y/a)^{1/b}\}^{(\nu_1+\nu_2)/2}}, \qquad y > 0 \tag{27.42}$$

[Malik (1967)]. Of course the rth moment of Y is

$$\mu'_r = E\left[\left(aF^b_{\nu_1,\nu_2}\right)^r\right] = a^r \mu'_{br}(F_{\nu_1,\nu_2}) \tag{27.43}$$

$$\equiv a^r\left(\frac{\nu_2}{\nu_1}\right)^{br}\frac{\Gamma(\frac{1}{2}\nu_1 + br)\Gamma(\frac{1}{2}\nu_2 - br)}{\Gamma(\frac{1}{2}\nu_1)\Gamma(\frac{1}{2}\nu_2)}, \qquad r < \frac{1}{2}\left(\frac{\nu_2}{b}\right) \qquad \text{[cf. (27.5)]}.$$

As mentioned earlier, generalizations of the beta distribution of the second kind (see Chapter 25) also become generalizations of the F-distribution and, as a result, is often denoted by GB2. The density function of this generalized F-distribution is given by [see McDonald and Richards (1987a, b)]

$$p_X(x|a,p,q) = \frac{|a|x^{ap-1}}{B(p,q)(1+x^a)^{p+q}}, \qquad x > 0,\ p, q > 0.$$

The kth moment (about zero) of X is given by

$$E[X^k] = B\left(p + \frac{k}{a}, q - \frac{k}{a}\right)\Big/B(p,q), \qquad -p < \frac{k}{a} < q,$$

from which we readily obtain the mean of X to be

$$E[X] = \Gamma\left(p + \frac{1}{a}\right)\Gamma\left(q - \frac{1}{a}\right)\Big/\{\Gamma(p)\Gamma(q)\}.$$

McDonald and Richards (1987a) discuss various properties of this distribution and also discuss the maximum likelihood estimation of the parameters of the distribution (with an additional scale parameter b).

The cumulative distribution function corresponding to the above generalized F density may be expressed in terms of the confluent hypergeometric function ${}_2F_1$ (see Chapter 1); see, for example, McDonald and Richards (1987a). The behavior of the hazard rate of this distribution has been examined by McDonald and Richards (1987b). Bookstaber and McDonald (1987) show that the above generalized F-distribution is quite useful in the empirical estimation of security returns and in facilitating the development of option pricing models (and other models) which depend on the specification and mathematical manipulation of distributions. Mixtures of two generalized F-distributions have been considered by McDonald and Butler (1987) who have applied it in the analysis of unemployment duration. McDonald and Butler (1990) have used the generalized F-distribution while discussing regression models for positive random variables. Applications of the generalized F-distribution in modeling insurance loss processes have been illustrated by Cummins et al. (1990). McDonald and Bookstaber (1991) have developed an option pricing formula based on the generalized F-distribution, which includes the widely used Black Scholes formula that is based on the assumption of lognormally distributed returns.

8.2 Other Related Distributions

Block and Rao (1973) define the ***beta warning time distribution*** as the marginal distribution of X when the conditional distribution of X, given $Y = y$, is a beta (m, n) distribution over $(0, y)$, with the pdf

$$p_{X|Y}(x|y) = \frac{x^{m-1}(y-x)^{n-1}}{B(m,n)y^{m+n-1}}, \qquad 0 < x < y;\ m,n > 0, \quad (27.44)$$

and Y has the pdf *(27.42)* with $a = 1$. The pdf of X is

$$p_X(x) = \frac{(\nu_1/\nu_2)^{\nu_1/2}x^{m-1}}{bB(m,n)B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)}\int_x \frac{y^{(\nu_1/(2b))-m-n}(y-x)^{n-1}}{\{1+(\nu_1/\nu_2)y^{1/b}\}^{(\nu_1+\nu_2)/2}}\,dy. \quad (27.45)$$

If $n = 1$, we have

$$p_X(x) = \frac{m(\nu_1/\nu_2)^{\nu_1/2}x^{m-1}}{bB(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)}\Big[B(\tfrac{1}{2}\nu_1 - bm, \tfrac{1}{2}\nu_2 + bm)$$

$$-B_{[1+(\nu_1/\nu_2)x^{-b}]^{-1}}(\tfrac{1}{2}\nu_1 - bm, \tfrac{1}{2}\nu_2 + bm)\Big], \quad (27.46)$$

where $B_p(a, b)$ is the incomplete beta function defined in Chapter 1, Eq. (1.90). The distribution with the pdf (27.46) is called a ***distended beta distribution.*** There are further details in Block and Rao *(1973).* Mihram (1969) contains some motivations for these kinds of distributions.

Mielke and Johnson (1974) reparametrize *(27.42)* taking

$$a = \beta,$$

$$b = \theta^{-1},$$

$$\nu_1 = 2\kappa\theta^{-1},$$

$$\nu_2 = 2(\alpha + 1 - \kappa\theta^{-1}),$$

and obtain

$$p_Y(y) = \frac{\theta}{\beta^{\kappa}B(\kappa\theta^{-1}, \alpha + 1 - \kappa\theta^{-1})} \cdot \frac{y^{\kappa-1}}{\{1 + (y/\beta)^{\theta}\}^{\alpha+1}},$$

$$y > 0;\ \theta, \beta, \kappa > 0. \qquad (27.47)$$

Introducing the restriction $\kappa = \alpha\theta$ (i.e., $\nu_1 = 2\alpha$; $\nu_2 = 2$), we obtain

$$p_Y(y) = \frac{\kappa}{\beta}\frac{(y/\beta)^{\kappa-1}}{\{1 + (y/\beta)^{\theta}\}^{1+(\kappa/\theta)}}, \qquad 0 < y \qquad (27.48a)$$

[since $B(\alpha, 1) = \alpha^{-1}$] which is called ***Mielke's beta-kappa distribution.*** This distribution is applied for stream flow and precipitation data. The cdf corresponding to (27.48a) is

$$F_Y(y) = \left\{\frac{(y/\beta)^{\theta}}{1 + (y/\beta)^{\theta}}\right\}^{\kappa/\theta}, \qquad 0 < y. \qquad (27.48b)$$

From (27.48b) it follows that the quantile y_p making $F_Y(y_p) = p$, is

$$y_p = \beta\left[p^{\theta/\kappa}(1 - p^{\theta/\kappa})\right]^{1/\theta}. \qquad (27.49)$$

Inserting this formula in the expressions for distributions of order statistics produces remarkably simple results.

Mielke and Johnson (1974) provide an iterative procedure for calculation of maximum likelihood estimators of the parameters β, θ and α $(= \kappa/\theta)$ of distribution (27.47). The rth moment of (27.48) is

$$\mu'_r(Y) = \alpha\beta^r B(\alpha + r\theta^{-1}, 1 - r\theta^{-1}), \qquad r < \theta \qquad (27.50)$$

(see page 322). Finally, returning to the situation described in Section 27.1 but with the distributions of the X_{ti}s being of symmetrical Pareto-Levy form, with

$$F_X(x) = \begin{cases} \frac{1}{2}|x|^{-\alpha}\{1 + o(1)\} & \text{as } x \to -\infty, \\ 1 - \frac{1}{2}x^{-\alpha}\{1 + o(1)\} & \text{as } x \to \infty, \end{cases} \qquad 1 < \alpha < 2, \quad (27.51)$$

then the limiting distribution of

$$F = \left(\frac{n_2}{n_1}\right)^{2/\alpha} \frac{\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2}{\sum_{i=1}^{n_2}(X_{2i} - \bar{X}_2)^2}$$

Table 27.2 Quantiles $1-\theta$ of the limit distribution $F(\alpha)$ for various α

$1-\theta$	α									
	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	1.0
0.990	6.857	15.636	29.758	53.651	96.179	176.386	338.548	695.841	1573.332	4052.181
0.975	3.268	6.399	11.008	18.120	29.622	49.197	84.561	153.254	299.302	647.789
0.950	2.105	3.543	5.618	8.349	12.584	19.212	30.139	49.315	85.639	161.448
0.900	1.528	2.178	3.009	4.111	5.625	7.788	11.015	16.084	24.551	39.863

as $n_1, n_2 \to \infty$ has the cdf

$$F_F(f|\alpha) = \begin{cases} \dfrac{2}{\pi\alpha}\tan^{-1}\left\{\dfrac{f^{\alpha/2}+\cos(\frac{1}{2}\pi\alpha)}{\sin(\frac{1}{2}\pi\alpha)}\right\} - \dfrac{1}{\alpha} + 1, & f > 0, \\ 0, & f \le 0, \end{cases} \quad (27.52a)$$

[Runde (1993)]. Table 27.2 shows some percentile points of this distribution. The corresponding density is

$$p_F(f|\alpha) = \frac{\sin(\frac{1}{2}\pi\alpha)}{\pi\alpha\{f^{-\alpha/2} + f^{\alpha/2} + 2\cos(\frac{1}{2}\pi\alpha)\}}, \qquad f > 0. \quad (27.52b)$$

Figures 27.3a, b exhibit some examples of this pdf.

The distribution of the maximum of certain sets of correlated F-variables is relevant to a procedure employed in the analysis of variance, in which it is desired to test whether the maximum of a number (say, k) of ratios of mutually independent mean squares $M_1, \cdots, M_k$ to a common "residual" mean square, M_0, is significantly large.

On the assumption of a standard (linear) model with independent homoscedastic (common variance σ^2) and normally distributed residuals, the null hypothesis (no fixed effects) distribution of M_j is $\sigma^2\chi^2_{\nu_j}/\nu_j$ ($j = 0, 1, \dots, k$) and the test statistic used in this procedure is

$$T = \max\frac{(M_1, \cdots, M_k)}{M_0} = \max\left(\frac{M_1}{M_0}, \cdots, \frac{M_k}{M_0}\right). \quad (27.53)$$

Under the null hypothesis, M_j/M_0 is distributed as F_{ν_j, ν_0}, but the k variables $M_1/M_0, \dots, M_k/M_0$ are not mutually independent.

In the special case $\nu_1 = \nu_2 = \cdots = \nu_k = \nu$, say, T is distributed as $(\nu_0/\nu) \times$ (ratio of maximum of k independent χ^2_ν variables to an independent $\chi^2_{\nu_0}$ variable). The distribution of $W = (\nu/\nu_0)T$ is thus that of the maximum of k correlated G_{ν, ν_0} variables.

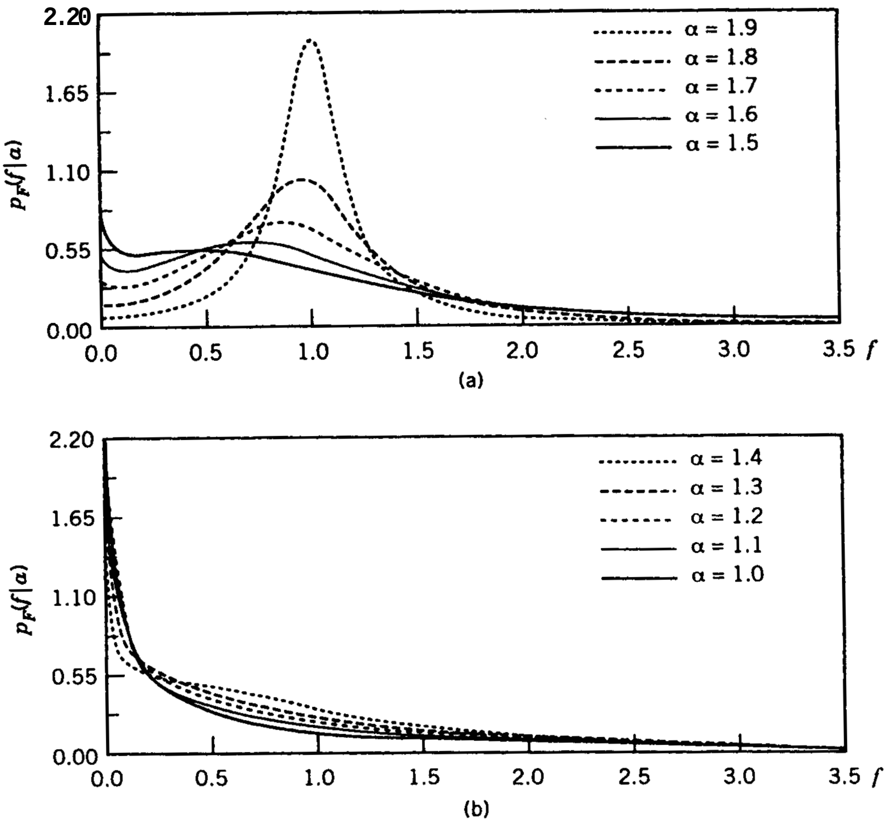


Figure 27.3 Densities $p_F(f|\alpha)$ for various a

Now if $X = \max(X_1, \cdots, X_k)$ and $X_1, \cdots, X_k$ are mutually independent with common χ^2_ν distribution, then

$$\Pr[X \le x] = \left\{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right\}^{-k}\left\{\int_0^x y^{(\nu/2)-1}e^{-y/2}\,dy\right\}^k, \qquad x \ge 0,$$

and the density

$$p_X(x) = k\left\{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right\}^{-k}\left\{\int_0^x y^{(\nu/2)-1}e^{-y/2}\,dy\right\}^{k-1} x^{(\nu/2)-1}e^{-x/2} \qquad x \ge 0,$$

whence

$$p_W(w) = k\left\{2^{\nu_0/2}\Gamma\left(\frac{\nu_0}{2}\right)\right\}^{-1}\left\{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\right)^{-k} w^{(\nu/2)-1}$$

$$\times \int_0^\infty \left\{\int_0^{uw} y^{(\nu/2)-1}e^{-y/2}\,dy\right\}^{k-1} u^{(\nu_0/2)-1}e^{-w(1+u)/2}\,du. \quad (27.54)$$

Krishnaiah and **Armitage (1964)** provided tables of upper $100\alpha\%$ points $(T_{\nu,\nu_0,1-\alpha})$ of T for $\alpha = 0.01, 0.10, 0.25$, for the choices of $\nu = 2(2)40$; $\nu_0 = 10(2)90$; $k = 1(1)12$. Hamdy, Son, and **AlMahmeed (1987)**, dealing with the special case $k = 2$, use the formula (valid for even ν and ν_0)

$$\alpha = \sum_{i=\nu/2}^{\infty} \binom{\left(\frac{\nu}{2}\right) + i - 1}{i} \left(\frac{\nu}{\nu_0}\right)^{(\nu/2)+i} \left(\frac{1}{2}\right)^{(\nu/2)+i-1}$$

$$\times \sum_{j=0}^{(\nu/2)-1} \binom{\frac{1}{2}(\nu + \nu_0) + j - 1}{j} H^j (1 - H)^{(\nu+\nu_0+i-j-1)/2}$$

$$= \sum_{i=0}^{(\nu_0/2)-1} \binom{\frac{1}{2}(\nu + \nu_0) - 1}{i} H^i (1 - H)^{((\nu+\nu_0)/2)-i-1}$$

$$- 2^{\nu/2} \sum_{i=0}^{(\nu/2)-1} \binom{\left(\frac{\nu}{2}\right) + i - 1}{i} \left(\frac{1 - H}{2 - H}\right)^{((\nu+\nu_0)/2)+i-1}$$

$$\times \sum_{j=0}^{(\nu_0/2)-1} \binom{\frac{1}{2}(\nu + \nu_0) + i - 1}{j} \left\{\frac{H}{2(1 - H)}\right\}^j, \qquad (27.55)$$

where

$$H = \left(1 + \frac{\nu}{\nu_0} T_{\nu,\nu_0,1-\alpha}\right)^{-1}.$$

These authors give values of $T_{\nu,\nu_0,1-\alpha}$ to five decimal places for $\alpha = 0.005, 0.01, 0.025, 0.1, 0.90, 0.95, 0.975, 0.995$; $\nu = 4(2)100$; $\nu_0 = 4(2)8$. (They refer to the existence of further tables, for $\nu_0 = 10(10)80$, with the same values of α and ν.)

Hartley (1950a, b) studied the distribution of a different, though similar F-type statistic, used in testing equality of k variances. This is the distribution of the ratio

$$\frac{\max(V_1, \cdots, V_k)}{\min(V_1, \cdots, V_k)},$$

where V_j's are mutually independent random variables with a common χ^2_ν distribution, as mentioned earlier on page 348.

Roy, Roy, and Ali (1993) have introduced the binomial mixture of F-distributions with density function

$$p_X(x|n,p,\nu_1,\nu_2) = \sum_{r=0}^{n}\binom{n}{r}\frac{p^r(1-p)^{n-r}(\nu_1/\nu_2)^{(\nu_1/2)+r}x^{(\nu_1/2)+r-1}}{B\left(\frac{\nu_1}{2}+r,\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}x\right)^{((\nu_1+\nu_2)/2)+r}},$$

$$0<x<\infty.$$

The kth moment about zero of X is given by

$$E[X^k] = \left(\frac{\nu_2}{\nu_1}\right)^k\sum_{r=0}^{n}\binom{n}{r}p^r(1-p)^{n-r}\frac{\Gamma\left(\frac{\nu_1}{2}+r+k\right)\Gamma\left(\frac{\nu_2}{2}-k\right)}{\Gamma\left(\frac{\nu_1}{2}+r\right)\Gamma\left(\frac{\nu_2}{2}\right)}$$

from which, in particular, we obtain the mean and variance of X to be

$$E[X] = \frac{\nu_2}{\nu_2-2}+\frac{2np\nu_2}{\nu_1(\nu_2-2)}$$

and

$$\text{var}(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} + \frac{4p\nu_2^2}{\nu_1^2(\nu_2-2)^2(\nu_2-4)} \times (n(2+5\nu_1-\nu_1\nu_2)+(n^2-n-\nu_2-4)p\}.$$

REFERENCES

Abramowitz, M., and Stegun, I. A. (eds.). (1964). ***Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,*** National Bureau of Standards, Applied Mathematics Series, 55, U.S. Department of Commerce, Washington, DC.

Amaral-Turkman, M. A., and Dunsmore, I. R. (1985). Measures of information in the predictive distribution, In ***Bayesian Statistics***, vol. *2* (eds., J. M. Bernardo, M. H. de Groot, D. V. Lindley and A. F. M. Smith), New York: Elsevier, North-Holland, pp. 603–612.

Armitage, J. V., and Krishnaiah, P. R. (1964). ***Tables of the Studentized largest chi-square distribution and their applications***, Report ARL 64-188, Aerospace Research Laboratories, Wright-Patterson Air Force Base, Dayton, OH.

Aroian, L. A. (1941). A study of R. A. Fisher's z-distribution and the related F distribution, ***Annals of Mathematical Statistics,*** **12,** 429–448.

Aroian, L. A. (1942). The relationship of Fisher's z-distribution to Student's t distribution, (Abstract), ***Annals of Mathematical Statistics,*** **13,** 451–452.

Aroian, L. A. (1947). Note on the cumulants of Fisher's z-distribution, ***Biometrika,*** **34,** 359–360.

Aroian, L. A. (1950). On the levels of significance of the incomplete beta function and the F-distributions, ***Biometrika,*** **37,** 219–223.

Aroian, L. A. (1953). A certain type of integral, ***American Mathematical Monthly,*** **50,** 382–383.

Ashby, T. (1968). A modification to Paulson's approximation to the variance ratio distribution, ***The Computer Journal,*** **11,** 209–210.

Awad, A. M. (1980). Remark on the characteristic function of the F-distribution, ***Sankhyii, Series A,*** **42,** 128–129.

Bennett, G. W., and Cornish, E. A. (1964). A comparison of the simultaneous fiducial distributions derived from the multivariate normal distribution, ***Bulletin of the International Statistical Institute,*** **46,** 918.

Bizley, M. T. L. (1950). A note on the variance-ratio distribution, ***Journal of the Institute of Actuaries Students' Society,*** **10,** 62–64.

Block, H. W., and Rao, B. R. (1973). A beta warning-time distribution and a distended beta distribution, ***Sankhyā, Series B,*** **35,** 79–84.

Bol'shev, L. N. (1960). On estimates of probabilities, ***Teoriya Veroyatnostei i ee Primeneniya, 5,*** 453–457. (In Russian. English translation: pp. 411–415.)

Bol'shev, L. N., Gladkov, B. V., and Shcheglova, M. V. (1961). Tables for the calculation of B and z-distribution functions, ***Teoriya Veroyatnostei i ee Primeneniya,*** **6,** 446–454. (English translation: pp. 410–419.)

Bookstaber, R. M., and McDonald, J. B. (1987). A general distribution for describing security price returns, ***The Journal of Business,*** **60,** 401–424.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria, ***Biometrika,*** **36,** 317–346.

Bradford, S. C. (1948). ***Documentation,*** London: Crosby Lockwood.

Buckley, M. J., and Eagleson, G. K. (1988). An approximation to the distribution of quadratic forms in normal random variables, ***Australian Journal of Statistics,*** **30,** 150–159.

Burk, F., Dion, L. R., Fridshal, D., Langford, E., O'Cinneidel, C., and Parsons, T. (1984). On a conjecture relating χ^2 and F quantiles, ***Communications in Statistics—Theory and Methods,*** **13,** 661–670.

Cacoullos, T. (1965). A relation between t and F distributions, ***Journal of the American Statistical Association, 60,*** 528–531. (Correction: **60,** 1249).

Carter, A. H. (1947). Approximation to percentage points of the z-distribution, ***Biometrika,*** **34,** 352–358.

Chen, H. J., and Makowsky, A. B. (1976). On approximations to the F-distribution and its inverse, ***Report*** 76-3, Department of Mathematical Sciences, Memphis State University, Memphis, TN.

Cheng, S. W., and Fu, J. C. **(1983).** An algorithm to obtain the critical values of the t, χ^2 and F distributions, ***Statistics & Probability Letters, 4,*** **223–227.**

Ciampi, A., Hogg, S. A., and Kates, L. (1988). Regression analysis of censored survival data with the generalized F family—An alternative to the proportional hazards model, ***Statistics in Medicine, 5,*** **85–96.**

Cochran, W. G. **(1940).** Note on an approximate formula for the significance levels of z, ***Annals of Mathematical Statistics,*** **11, 93–95.**

Colcord, C. G., and Deming, L. S. (1936). The one-tenth percent level of "z," ***Sankhyā***, 2, **423–424.**

Cornish, E. A., and Fisher, R. A. (1937). Moments and cumulants in the specification of distributions, ***Reciew of the International Statistical Institute, 5,*** **307–322.**

Cummins, J. D., Dionne, **G.,** McDonald, J. **B.,** and Pritchett, B. M. **(1990).** Applications of the **GB2** family of distributions in modeling insurance loss processes, ***Insurance: Mathematics and Economics, 9,*** **257–272.**

Curtiss, J. H. **(1953).** Convergent sequences of probability distributions, ***American Mathematical Monthly, 50,*** **100–107.**

Davenport, J. M., and Herring, T. A. **(1979).** On the use of curve fitting to model the error of the Cornish-Fisher expansion of the Pearson type-VI distribution, ***Communications in Statistics—Simulation and Computation,*** **B8, 311–333.**

David, F. N. **(1949).** The moments of the z and F distributions, ***Biometrika, 36,*** **394–403.**

Dion, L. R., and Fridshal, D. (1982). A relationship between chi-square and F critical values, ***Communications in Statistics—Simulation and Computation,*** **B11**, 233–235.

Donner, A., Wells, G. A., and Eliasziw, M. (1989). On two approximations to the F-distributions: Application to testing for intraclass correlation in family studies, ***Canadian Journal of Statistics, 17,*** **209–215.**

Donner, **E. (1968).** Algorithm **322:** F-distribution, ***Communications of the Association for Computing Machinery,*** **11, 116–117.**

Dyer, D. **(1982).** The convolution of generalized F-distributions, ***Journal of the American Statistical Association, 77,*** **189–189.**

Elderton, W. **P.,** and Johnson, N. L. **(1969).** ***Systems of Frequency Curves***, Cambridge: Cambridge University Press.

Fisher, R. A. (1924). On a distribution yielding the error functions of several well-known statistics, ***Proceedings of International Mathematical Congress, Toronto,*** **805–813.**

Fisher, R. A. (1925). ***Statistical Methods for Research Workers,*** 4th ed., Edinburgh: Oliver and Boyd.

Fisher, R. **A.,** and Cornish, E. A. **(1960).** The percentile points of distributions having known cumulants, ***Technometrics,*** 2, **209–225.**

Fisher, R. **A.,** and Yates, F. (1953). ***Statistical Tables for Biological, Agricultural, and Medical Research,*** 4th ed., London: Oliver and Boyd.

Fox, B. L. **(1963).** Generation of random samples from the beta and F distributions, ***Technometrics,*** 5, **269–270.**

Fujikoshi, **Y.** (1987). Error bounds for asymptotic expansions of scale mixtures of distributions. ***Hiroshima Mathematical Journal.*** **17. 309–324.**

Fujikoshi, Y., and Mukaihata, S. (1993). Approximations for the quantiles of Student's *t* and F distributions and their error bounds, ***Hiroshima Mathematical Journal,*** **23, 557–564.**

Gayen, A. K. (1950). The distribution of the variance ratio in random samples of any size drawn from non-normal universes, ***Biometrika,*** **37, 236–255.**

Geisser, S. (1964). Estimation in the uniform covariance case, ***Journal of the Royal Statistical Society, Series B,*** **26, 477–483.**

George, E. O., and Singh, K. P. **(1987).** An approximation of F-distribution by binomial probabilities, ***Statistics & Probability Letters,*** **5, 169–173.**

Golden, R. R., Weiss, D. J., and Davis, R. V. **(1968).** An evaluation of Jaspen's approximation of the distribution functions of the F, *t* and chi-square statistics, ***Educational and Psychological Measurement,*** **28, 163–165.**

Greenwood, J., and Hartley, H. O. **(1961).** ***Guide to Tables in Mathematical Statistics,*** Princeton: Princeton University Press.

Grzegórski, S. **(1972).** Evaluation of probability for the F-Snedecor distribution, ***Zastosowania Matematyki,*** **12, 499–502.**

Hald, A. **(1952).** ***Statistical Tables and Formulas,*** New York: Wiley.

Hamdy, H., Son, M., and AlMahmeed, M. (1987). A note on the distribution of maximum correlated F, ***Proceedings of ASA Statistical Computing Section,*** pp. **406–411.**

Hartley, H. O. (1950a). The use of range in analysis of variance, ***Biometrika,*** **37, 271–280.**

Hartley, H. O. (1950b). The maximum F-ratio as a short cut test for heterogeneity of variance, ***Biometrika,*** **37, 308–312.**

Horsnell, G. **(1953).** The effect of unequal group variances on the F-test for homogeneity of group means, ***Biometrika,*** 40, **128–136.**

Ifram, A. F. **(1970).** On the characteristic function of the F- and t-distributions, ***Sankhyii, Series A,*** **32, 350–352.**

IMSL **(1985).** Institute of Mathematical and Statistical Libraries, Packages, Edition **9.2,** Vol. **3,** Houston, Texas.

Jaspen, N. **(1965).** The calculation of probabilities corresponding to values of z, ***t, F,*** and chi-square, ***Educational and Psychological Measurement,*** **25, 877–880.**

Johnson, E. E. (1973). Empirical equation for approximating tabular F values, ***Technometrics,*** **15, 379–384.**

Johnson, N. L. **(1959).** On an extension of the connection between Poisson and χ^2 distributions, ***Biometrika,*** 46, **352–363.**

Jowett, G. H. **(1963).** The relationship between the binomial and F distributions, ***The Statistician,*** **13, 55–57.**

Kelley, T. L. (1948). ***The Kelley Statistical Tables (Revised),*** Cambridge, MA: Harvard University Press.

Kennedy, W. J., Jr., and Gentle, J. E. **(1980).** ***Statistical Computing,*** New York: Dekker.

Kolodziejczyk, S. **(1935).** On an important class of statistical hypotheses, ***Biometrika,*** **27. 161–190.**

Krishnaiah, P. K., and Armitage, J. V. (1964). Tables for the Studentized largest chi-square distribution and their applications, *Aerospace Research Laboratories*, Wright-Patterson Air Force Base, Dayton, OH, pp. 64–188.

Kullback, S. (1935). The distribution laws of the difference and quotient of variables independently distributed in Pearson Type III laws, *Annals of Mathematical Statistics*, 7, 51–53.

Kymn, K. O. (1974). A simple derivation of the relation between t- and F-distributions, *SIAM Journal of Applied Mathematics*, 27, 517–518.

Lasdon, L. S., Waren, A. D., Jain, A., and Ratner, M. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming, *Association for Computing Machinery Transactions on Mathematical Software, 4*, 34–50.

Laubscher, N. F. (1965). Interpolation in F-tables, *The American Statistician*, 19, No. 1, 28, 40.

Lee, C. M.-S. (1988). On the computation of F-cumulative probabilities, *Communications in Statistics—Simulation and Computation*, 17, 1191–1201.

Leimkuhler, F. F. (1967). The Bradford distribution, *Journal of Documentation*, 23, 197–207.

Lentner, M. (1976). On the exact calculation of cumulative F probabilities, *Communications in Statistics*, B5, 149–155.

Malik, H. J. (1967). Exact distribution of the quotient of independent generalized gamma variables, *Canadian Mathematical Bulletin*, 10, 463–465.

Mantel, N. (1966). F-ratio probabilities from binomial tables, *Biometrics*, 22, 404–407.

Mardia, K. V., and Zemroch, P. J. (1978). *Tables of the* F- *and Related Distributions with Algorithms*, New York: Academic Press.

McDonald, J. B., and Bookstaber, R. M. (1991). Option pricing for generalized distributions, *Communications in Statistics—Theory and Methods*, 20, 4053–4068.

McDonald, J. B., and Butler, R. J. (1987). Some generalized mixture distributions with an application to unemployment duration, *The Review of Economics and Statistics*, 69, 232–240.

McDonald, J. B., and Butler, R. J. (1990). Regression models for positive random variables, *Journal of Econometrics*, 43, 227–251.

McDonald, J. B., and Richards, D. O. (1987a). Model selection: Some generalized distributions, *Communications in Statistics—Theory and Methods*, 16, 1049–1074.

McDonald, J. B., and Richards, D. O. (1987b). Hazard rates and generalized beta distribution, IEEE *Transactions on Reliability*, R-36, 463–466.

McIntyre, G. A., and Ward, M. M. (1968). Estimates of percentile points of the F-distribution, *Australian Computation Journal*, 1, 113–114.

Merrington, M., and Thompson, C. M. (1943). Tables of percentage points of the inverted beta (F) distribution, *Biometrika*, 33, 73–88.

Mielke, P. W., Jr. (1973). Another family of distributions for describing and analyzing precipitation data, *Journal of Applied Meteorology*, 12, 275–280.

Mielke, P. W., Jr., and Johnson, E. S. (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology, *Water Resources Research*, 10, 223–226 (Correction: 12, [1976], 827).

Mihram, G. A. (1969). A distended Gamma distribution, *Sankhyā*, *Series* B, *31*, 421–424.

Morris, J. (1969). Algorithm 346: F-test, *Communications of the Association for Computing Machinery,* **13**, **184**.

Mudholkar, G. S., and Chaubey, Y. P. **(1976).** Some refinements of the Wise-approximation for the beta and F-distributions, ***Utilitas Mathernatica,* 10, 199–207.**

Newman, D. (1939). The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation, ***Biometrika,* 31, 20–30.**

Ojo, M. O. (1985). A new approximation to the F-distribution, ***Utilitas Mathernatica,* 28, 121–128.**

Ojo, M. O. (1988). A new approximation to the incomplete beta function, ***Communications in Statistics—Theory and Methods,*** **17**, 1423–1435.

Owen, D. B. **(1962).** ***Handbook of Statistical Tables,*** Reading, MA: Addison-Wesley.

Patil, S. A., Raghunandanan, K., and Lee, R.-Y. (1985a). Order statistics of the F-distribution, ***Proceedings of ASA Statistical Computing Section,*** pp. 337–341.

Patil, S. **A.,** Raghunandanan, K., and Lee, R.-Y. (1985b). On the moments of order statistics of the F-distribution, ***Computational Statistics Quarterly,* 2, 285–302.**

Paulson, E. (1942). An approximate normalization of the analysis of variance distribution, ***Annals of Mathematical Statistics, 13,*** 233–235.

Pearson, E. S., and Hartley, H. O. **(1958).** ***Biometrika Tables for Statisticians,*** vol. 1, *2d* ed., Cambridge: Cambridge University Press.

Pestana, D. (1977). Note on a paper of Ifram, ***Sankhyā*, *Series A*, 39, 396–397.**

Pham-Gia, T., and Duong, Q. P. **(1989).** The generalized beta- and F-distributions in statistical modelling, ***Mathematical and Computational Modelling,* 12,** 1613–1625.

Phillips, P. C. B. **(1982).** The true characteristic function of the F distribution, ***Biometrika,* 69,** 261–**264.**

Prentice, R. L. (1975). Discrimination among some parametric models, ***Biometrika,* 62, 607–614.**

Prins, H. J. **(1960).** Transforms for finding probabilities and variates of a distribution in terms of a related distribution function, ***Statistica Neerlandica,*** 14, **1–17.**

Rider, P. R. **(1931).** A note on small sample theory, ***Journal of the American Statistical Association,* 26, 172–174.**

Roy, M. **K.,** Roy, A. K., and Ali, M. Masoom **(1993).** Binomial mixtures of some standard distributions, ***Journal of Information and Optimization Sciences,* 14, 57–71.**

Runde, R. **(1993).** A note on the asymptotic distribution of the F-statistic for random variables with infinite variance, ***Statistics & Probability Letters,* 18, 9–12.**

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components, ***Biometrics*, 2, 110–114.**

Schader, M., and Schmid, F. **(1986).** Distribution function and percentage points of the central and non-central F-distributions, ***Statistische Hefte,* 27, 67–74.**

Scheffk, H. **(1942).** On the ratio of variances of two normal populations, ***Annals of Mathematical Statistics,* 13, 371–388.**

Scheffk, H., and Tukey, J. W. (1944). A formula for sample sizes for population tolerance limits, ***Annals of Mathematical Statistics,* 15, 217.**

Shah, M. C., and Rathie, P. W. (1974). Exact distribution of product of generalized F-variates, ***Canadian Journal of Statistics,* 2, 13–24.**

Smillie, K. W., and Anstey, T. H. (1964). A note on the calculation of probabilities in an F-distribution, *Communications of the Association for Computing Machinery,* **7,** 725.

Snedecor, G. W. (1934). *Calculation and Interpretation of the Analysis of Variance,* Ames, IA: Collegiate Press.

Stammberger, A. (1967). Über einige Nomogramme zür Statistik, *Wissenschaftliche Zeitschrift der Humboldt-Universitat Berlin. Mathematisch-Naturwissenschaftliche Reihe,* **16,** 86–93.

Stuart, A., and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics—Volume I,* Sixth edition, London: Edward Arnold.

Swain, A. K. P. C. (1965). A lower bound to the probability of variance-ratio, *Annals of the Institute of Statistical Mathematics,* **17,** 81–84.

Tiao, G. C., and Lochner, R. H. (1966). *Tables for the Comparison of the Spread of Two Normal Distributions,* Technical Report No. 88, Department of Statistics, University of Wisconsin, Madison.

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions, *Biometrika,* **51,** 83–95.

Viveros, R. (1990). An approximate normalizing power transformation for the F distribution, *Communications in Statistics—Simulation and Computation,* **19,** 57–69.

Vogler, L. E., and Norton, K. A. (1957). *Graphs and Tables of the Significance Levels ($F(r_1, r_2, p)$) of the Fisher-Snedecor Variance Ratio,* National Bureau of Standards Report No. 5069, Washington, DC: GPO.

Wishart, J. (1946). The variance ratio test in statistics, *Journal of the Institute of Actuaries Students' Society, 6,* 172–184.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions, *Biometrika, 34,* 170–178. (Correction: 374.)

Wishart, J. (1957). An approximate formula for the cumulative z-distribution, *Annals of Mathematical Statistics,* **28,** 504–510.

Wood, A. T. A. (1989). The F approximation to the distribution of a linear combination of chi-squared variables, *Communications in Statistics—Simulation and Computation,* **18,** 1439–1456.

Yip, D. Y. N. (1974). *Chi-square and F approximations of Hotelling's generalized T^2,* M.Sc. project, Department of Applied Mathematics, McMaster University, Hamilton, Ontario, Canada.

Zeigler, C. O. (1965). *The effect of the normality and homogeneity of cariance assumptions upon the calidity of the F table for interaction terms,* M.Sc. thesis, Department of Industrial Engineering, Texas Technological College, Lubbock.

Zelen, M., and Severo, N. C. (1964). Probability Functions, In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* (eds., M. Abramovitz and I. A. Stegun), Applied Mathematics Series, 55, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, pp. 925–955.

Zinger, A. (1964). On interpolation in tables of the F-distribution, *Applied Statistics,* 13. 51–53.

CHAPTER 28

t-Distributions

1 GENESIS AND HISTORICAL REMARKS

If $X_1, X_2, \ldots, X_n$ are independent random variables each having the same normal distribution, with expected value ξ and standard deviation σ, then, $\sqrt{n}(\bar{X} - \xi)/\sigma$ (with $\bar{X} = n^{-1}\sum_{j=1}^{n} X_j$) has a unit normal distribution. This statistic can be used in the construction of tests and confidence intervals relating to the value of ξ provided that σ is known. If σ be not known, it is reasonable to replace it by the estimator $S = [(n-1)^{-1}\sum_{j=1}^{n}(X_j - \bar{X})^2]^{1/2}$ giving the statistic $T = \sqrt{n}(\bar{X} - \xi)[(n-1)^{-1}\sum_{j=1}^{n}(X_j - \bar{X})^2]^{-1/2}$. This procedure was adopted for some time, without making allowance for differences between the distributions of $\sqrt{n}(\bar{X} - \xi)/\sigma$ and $\sqrt{n}(\bar{X} - \xi)/S$. It was realized that the two distributions are not identical but the determination of the actual distribution presented difficulties. "Student" (1908) obtained the distribution of

$$T' = \sqrt{n}(\bar{X} - \xi)\left[\sum_{j=1}^{n}(X_j - \bar{X})^2\right]^{-1/2} = \frac{1}{\sqrt{n-1}}\left\{\frac{\sqrt{n}(\bar{X} - \xi)}{S}\right\} = \frac{T}{\sqrt{n-1}}$$

and gave a short table of its cumulative distribution function.

Recall the results on the joint distribution of $\bar{X}$ and S described in Chapter 13. There it can be seen that T' is distributed as a ratio U/χ_{n-1}, the two variables U (a unit normal variable) and χ_{n-1} being mutually independent. The divisor $\sqrt{n-1}$ in the denominator was introduced by Fisher (1925a) who defined t with ν degrees of freedom as the distribution of

$$t_\nu = U\left[\frac{\chi_\nu^2}{\nu}\right]^{-1/2}. \tag{28.1}$$

This quantity is usually called Student's t and the corresponding distribution is called Student's distribution. Occasionally, they are called Fisher's statistic and distribution, respectively, but these latter terms more commonly refer to the variance-ratio F and its distribution, as were discussed in Chapter 27. The evolution of the t-statistic and its various usages are explained in Eisenhart (1979) and Box (1981).

Cacoullos (1965) showed that if X_0 and X_1 are independent χ^2 variables (Chapter 18) each with ν degrees of freedom, then $\frac{1}{2}\sqrt{\nu}(X_1 - X_0)(X_0X_1)^{-1/2}$ has a t_ν distribution. Equivalently, if Y has a $F_{\nu,\nu}$-distribution, then $\frac{1}{2}\sqrt{\nu}(Y^{1/2} - Y^{-1/2})$ has a t_ν distribution. See Chapter 27, Section 8 (page 347).

A notable characterization is due to Bondesson (1981), who has shown that for i.i.d. random variables $X_1, \cdots, X_n$ that have a common distribution with finite moments of all orders and a pdf that is continuous at zero, if

$$T = \sqrt{n}\,\bar{X}\left\{(n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right\}^{-1/2}$$

is distributed as t_{n-1} for all $n \geq 2$, then this common distribution must be normal with mean zero and positive variance. It should be noted that for $n = 2$ there is a distribution (without a finite expected value)——that of the reciprocal of a unit normal variable—for which $\sqrt{2}\bar{X}\{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2\}^{-1/2} = \sqrt{2}\bar{X}(\frac{1}{2}|X_1 - X_2|)^{-1}$ has a t_1-distribution.

For a brief recent description of some properties of the t distribution and information concerning the available tables and approximations, interested readers may refer to Stuart and Ord (1994, pp. 546–549).

2 PROPERTIES

The probability density function of $t_\nu = U[\chi_\nu/\sqrt{\nu}]^{-1}$ is

$$p_{t_\nu}(t) = \frac{1}{\sqrt{\nu}\,B(\frac{1}{2},\frac{1}{2}\nu)}\left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} = \frac{\Gamma(\frac{1}{2}(\nu+1))}{\sqrt{\pi\nu}\,\Gamma(\frac{1}{2}\nu)}\left[1 + \frac{t^2}{\nu}\right]^{-(\nu+1)/2}. \tag{28.2}$$

This is a special form of the Pearson Type VII distribution. It is symmetrical about $t = 0$ and has a single mode at $t = 0$. It is easy to show that

$$\lim_{\nu\to\infty} p_{t_\nu}(t) = (\sqrt{2\pi})^{-1}e^{-t^2/2}.$$

In fact, as $\nu \to \infty$, the distribution of t_ν tends to the unit normal distribution. (This fact is the basis of most of the methods of approximation described in Section 4).

If $t_{\nu,\alpha}$ is defined by the equation

$$\Pr[t_\nu \le t_{\nu,\alpha}] = \alpha, \tag{28.3}$$

then (from the symmetry)

$$t_{\nu,0.5} = 0 = U_{0.5},$$

where $\Phi(U_\alpha) = a$.

However, for $a > 0.5$,

$$t_{\nu,\alpha} > U_\alpha > 0,$$

and for $a < 0.5$,

$$t_{\nu,\alpha} < -U_{1-\alpha} < 0$$

(of course $t_{\nu,1-\alpha} = t_{\nu,\alpha}$).

Making the transformation $w = \nu(\nu + y^2)^{-1}$, we have for $t \ge 0$,

$$\Pr[t_\nu \le t] = \Pr[t_\nu \le 0] + \left[\sqrt{\nu}\, B\left(\frac{1}{2}, \frac{1}{2}\nu\right)\right]^{-1} \int_0^t \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2} dy$$

$$= \frac{1}{2} + \frac{1}{2}\left[B\left(\frac{1}{2}, \frac{1}{2}\nu\right)\right]^{-1} \int_{\nu/(\nu+t^2)}^{1} w^{(\nu-2)/2}(1-w)^{-1/2}\, dw$$

$$= 1 - \frac{1}{2} I_{\nu/(\nu+t^2)}\left(\frac{1}{2}\nu, \frac{1}{2}\right) = \frac{1}{2}\left[1 + I_{t^2/(\nu+t^2)}\left(\frac{1}{2}, \frac{1}{2}\nu\right)\right], \tag{28.4a}$$

where $I_x(a,b)$ is the incomplete beta function ratio, which is defined in Chapter 1. This identity can be used for computer evaluation of t-distribution functions using one of the several algorithms for evaluating the beta distribution function (see Chapter 25, Section 6). In particular, Lee and Singh (1988), starting from (28.4a), reached the formulas

$$\Pr[t_\nu \le t] = \frac{3}{4} + \pi^{-1}\sqrt{y(1-y)}\left\{1 + \sum_{i=1}^{(\nu-3)/2}\left[\prod_{j=1}^{i}\frac{2j}{2j+1}\right]y^i\right\}$$

$$- \frac{1}{2}\pi^{-1}\sin^{-1}(2y-1) \qquad \text{for } \nu \text{ odd}, \tag{28.4b}$$

$$\Pr[t_\nu \le t] = \frac{1}{2} + \frac{1}{2}\sqrt{1-y}\left\{1 + \sum_{i=1}^{(\nu-2)/2}\left[\prod_{j=1}^{i}\frac{2j-1}{2j}\right]y^i\right\} \qquad \text{for } \nu \text{ even}, \tag{28.4c}$$

where $y = \nu/(\nu + t^2)$ and $\Sigma_{i=1}^{-1} = -1$, $\Sigma_{i=1}^{0} = 0$. Alternative expressions (valid for all t) are

$$\Pr[t_1 \leq t] = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} t \qquad \text{for } \nu = 1, \tag{28.4d}$$

and, setting $\theta = \tan^{-1}(t/\sqrt{\nu})$,

$$\Pr[t_\nu \leq t] = \frac{1}{2} + \frac{1}{\pi}\left[\theta + \left\{\cos\theta + \frac{2}{3}\cos^3\theta + \cdots + \frac{(2)(4)\cdots(\nu-3)}{(3)(5)\cdots(\nu-2)}\cos^{\nu-2}\theta\right\}\sin\theta\right]$$

$$\text{for } \nu \text{ odd and greater than 1,} \tag{28.4e}$$

$$\Pr[t_\nu \leq t] = \frac{1}{2} + \frac{1}{2}\left\{1 + \frac{1}{2}\cos^2\theta + \frac{(1)(3)}{(2)(4)}\cos^4\theta + \cdots + \frac{(1)(3)\cdots(\nu-3)}{(2)(4)\cdots(\nu-2)}\cos^{\nu-2}\theta\right\}\sin\theta \qquad \text{for } \nu \text{ even} \tag{28.4f}$$

[see Zelen and Severo (1964)].

Amos (1964) obtained several expressions for $\Pr[t_\nu \leq t]$ in terms of hypergeometric functions (see Chapter 1, Section A6). For example,

$$\Pr[t_\nu \leq t] = \frac{1}{2} + \frac{t}{\sqrt{\pi\nu}}\frac{\Gamma(\frac{1}{2}(\nu+1))}{\Gamma(\frac{1}{2}\nu)} \times {}_2F_1\left[\frac{1}{2}(\nu+1); \frac{1}{2}; \frac{3}{2}; \frac{-t^2}{\nu}\right] \qquad \text{for } t^2 < \nu, \tag{28.4g}$$

which is useful when both $|t|/\nu^{1/2}$ and ν are small.

All odd moments of t_ν about zero are zero. If r is even. then the rth central moment is

$$\mu_r(t_\nu) = \nu^{r/2}\cdot\frac{\Gamma(\frac{1}{2}(r+1))\Gamma(\frac{1}{2}(\nu-r))}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}\nu)}$$

$$= \nu^{r/2}\cdot\frac{1\cdot 3\cdots(r-1)}{(\nu-r)(\nu-r+2)\cdots(\nu-2)}. \tag{28.5}$$

(Note that if $r \geq \nu$, the rth moment is infinite.) Setting $r = 1$, we obtain a formula for the mean deviation as

$$E[|t_\nu|] = \sqrt{\nu}\,\frac{\Gamma(\frac{1}{2}(\nu-1))}{\sqrt{\pi}\,\Gamma(\frac{1}{2}\nu)}. \tag{28.6}$$

From (28.5),

$$\operatorname{var}(t_\nu) = \frac{\nu}{\nu-2}, \qquad \nu \ge 2, \tag{28.7a}$$

$$\alpha_4(t_\nu) = \beta_2(t_\nu) = 3 + \frac{6}{\nu-4} = \frac{3(\nu-2)}{(\nu-4)}, \qquad \nu \ge 4, \tag{28.7b}$$

decreases from 9 for $\nu = 5$ to 3 as $\nu \to \infty$. Also $\alpha_3(t_\nu) = \sqrt{\beta_1(t_\nu)} = 0$ [Wishart (1947)].

From (28.6) and (28.7a),

$$\frac{\text{Mean deviation}}{\text{Standard deviation}} = \sqrt{\frac{2}{\pi}}\left[\sqrt{\tfrac{1}{2}\nu - 1}\,\frac{\Gamma(\frac{1}{2}(\nu-1))}{\Gamma(\frac{1}{2}\nu)}\right] \tag{28.8}$$

The multiplier of $\sqrt{2/\pi}$ tends rapidly to 1 as ν increases, as can be seen from Table 28.1.

The probability density function of t_ν has points of inflexion at

$$t_\nu = \pm\sqrt{\nu(\nu+2)}\,.$$

As ν increases, the distribution of t_ν approaches quite rapidly to the unit normal distribution. Figure 28.1 compares the t_4 and unit normal probability density functions. Even for such a small number of degrees of freedom, the two functions are not markedly dissimilar. If the standardized t_4—that is, $t_4/\sqrt{2}$—distribution is used the agreement is even closer [Weir (1960a)].

By considering the transformed random variable

$$U_\nu = \left\{\left(\nu - \frac{1}{2}\right)\log\left(1 + \frac{1}{\nu}t_\nu^2\right)\right\}^{1/2}\operatorname{sgn}(t_\nu)$$

when $\nu > \frac{1}{2}$, Fujikoshi and Mukaihata (1993) have shown that the distribution of U_ν converges rapidly to the standard normal distribution. In fact these

Table 28.1 Ratio of mean deviation to standard deviation for t_ν distribution

ν	3	4	5	6	7	8	9	10	11	12
Ratio	0.637	0.707	0.735	0.750	0.759	0.765	0.770	0.773	0.776	0.778

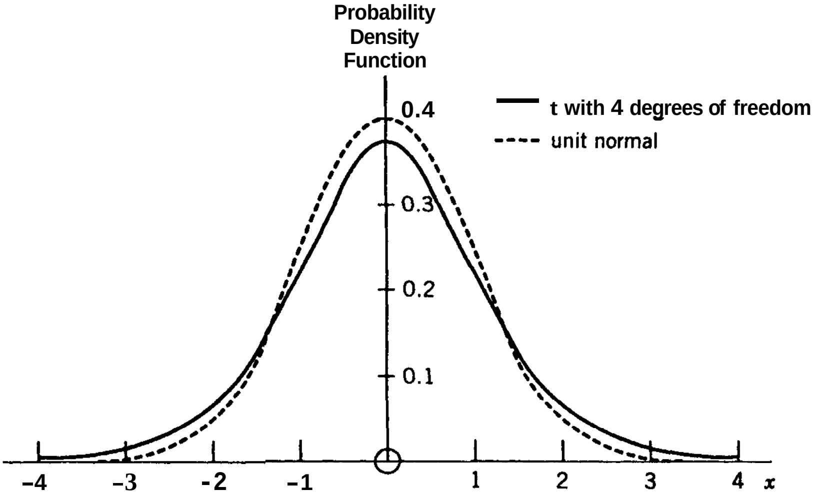


Figure 28.1 Comparison of Unit Normal and t_4 Density Functions

authors have shown that

$$\Pr[U_\nu \le x] = \Phi(x) + O(\nu^{-2})$$

for all real x. Fujikoshi and Mukaihata (1993) have also derived some approximations and bounds for the quantiles of the distribution of U_ν.

If ν is odd, the characteristic function of t_ν is

$$\varphi_{t_\nu}(t) = \left\{ \sum_{j=0}^{m-1} c_{j,m-1} |\nu^{1/2} t|^j \right\} \exp(-|\nu^{1/2} t|), \tag{28.8}$$

where $m = \frac{1}{2}(\nu + 1)$ and the $c_{j,m}$'s satisfy the recursive relations:

$$c_{o,m} = 1,\ c_{1,m} = 1,\ c_{m-1,m} = \{(2m-3)(2m-5)\cdots 3\cdot 1\}^{-1},$$

$$c_{j,m} = \{c_{j-1,m-1} + (2m-3-j)c_{j,m-1}\}(2m-3)^{-1} \qquad \text{for } 1 \le j \le m-1 \tag{28.9}$$

[Mitra (1978)]. In particular,

$$\varphi_{t_1}(t) = \exp(-|t|),$$

$$\varphi_{t_3}(t) = (1 + |t\sqrt{3}|)\exp(-|t\sqrt{3}|),$$

$$\varphi_{t_5}(t) = (1 + |t\sqrt{5}| + \tfrac{5}{3}t^2)\exp(-|t\sqrt{5}|),$$

$$\varphi_{t_7}(t) = (1 + |t\sqrt{7}| + \tfrac{14}{5}t^2 + \tfrac{1}{15}|t\sqrt{7}|^3)\exp(-|t\sqrt{7}|). \tag{28.10}$$

Ifram (1972) provides an alternative form

$$\varphi_{t_{2m+1}}(t) = \left[B\left(\tfrac{1}{2}, m + \tfrac{1}{2}\right)\right]^{-1} 2\pi i \left[(z + i)^{-m-1} \exp\left\{it(2m + 1)^{1/2} z\right\}\right]_{z=i}^{(m)} / m!. \tag{28.11}$$

He also provides (without detailed derivation) an expression for the case $\nu = 2n$, namely

$$\varphi_{t_{2n}}(t) = \left(B\left(\tfrac{1}{2}, n\right)\right)^{-1} 2\pi i \sum_{k=0}^{\infty} \binom{n - \frac{1}{2}}{k} (2i)^k \frac{\left\{\exp\left(it(2n)^{1/2} z\right)\right\}_{z=i}^{(k+2n)}}{(k + 2n)!}. \tag{28.12}$$

The Fourier transform of the t-distribution involves simple Bessel polynomials. In particular, the quotient

$$\frac{p_{n-1}(\sqrt{x})}{p_n(\sqrt{x})}, \tag{28.13}$$

where

$$p_n(z) = \sum_{k=0} \frac{(n + k)!}{(n - k)!} \frac{}{2^k k!} = z^n y_n\left(\frac{1}{z}\right)$$

and y_n is the nth Bessel polynomial $[y_n(z) = {}_2F_0(-n, n + 1; -; -z/2)]$ arises in connection with the problem of infinite divisibility of the t_ν-distribution. To show that this distribution is infinitely divisible, it is sufficient to prove that (28.13) is strictly monotonic in $[0, \infty)$. For $n = 4$, 5, and 6, this was established by Ismail and Kelker (1976), implying infinite divisibility for $\nu = 9$, 11, and 13. Later in the same year Grosswald (1976a) proved the strict monotonicity of (28.13) for odd values of n and finally Grosswald (1976b) succeeded in showing the validity of this result for all n, thus proving the general infinite divisibility of Student's t-distributions. Epstein (1977) provided, independently, a somewhat more elementary proof. As in the case of Grosswald's proof, the "even" case is more complicated. The distribution of t_ν^{-1}, for ν (≥ 3) odd, fails to be infinitely divisible.

3 TABLES AND NOMOGRAMS

3.1 Tables

The cumulative t-distributions and percentile points have been rather thoroughly tabulated. We give here a list, roughly in chronological order. There are in addition short tables in many textbooks, which are mostly derived from tables in our list. (Note that $\nu = \infty$ corresponds to the unit normal distribution.)

The earliest published tables were provided by "Student" (1908). These gave values of $\Pr[z_\nu \leq z]$, where $z_\nu = t_\nu/\sqrt{\nu+1}$. Later the same author ["Student" (1925)] gave the values of $\Pr[t_\nu \leq t]$ to four decimal places for $\nu = 1(1)20, \infty$ and $t = 0(0.1)6.0$, and also to six decimal places for $\nu = 3(1)11$ and $t = 6.0(0.5)10.0(1)12(2)16(4)28$.

In Pearson and Hartley (1958) there are tables of $\Pr[t_\nu \leq t]$ to five decimal places for

$$\nu = 1(1)24, 30, 40, 60, 120, \infty$$

and

$$t = \begin{cases} 0.0(0.1)4.0(0.2)8.0 & (\nu \leq 20), \\ 0.00(0.05)2.0(0.1)4.0, 5 & (\nu \geq 20) \end{cases},$$

and of $t_{\nu,\alpha}$ to three decimal places for

$$\nu = 1(1)30, 40, 60, 120, \infty$$

and

$$\alpha = 0.6, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975, 0.999, 0.9995.$$

Also $t_{\nu,\alpha}$ is given to at least three significant figures for

$$\nu = 1(1)10 \quad \text{and} \quad \alpha = 0.9999, 0.99999, 0.999995.$$

Parts of these tables appeared earlier in Baldwin (1946) and Hartley and Pearson (1950); parts are reproduced in Janko (1958). Rao, Mitra, and Mathai (1966) gave similar tables with the addition of $\nu = 80$ (and exclusion of $\nu = 120$) and $\alpha = 0.7, 0.8$. These tables are notable for the extreme tail percentiles included. For more extensive sets of values of ν, however, some of the following tables are more useful.

Fisher and Yates (1966) gave values of $t_{\nu,\alpha}$ to three decimal places for

$$\nu = 1(1)30, 40, 60, 120$$

and

$$\alpha = 0.55(0.05)0.95, 0.975, 0.99, 0.995, 0.9995.$$

Lampers and Lauter (1971) extended these tables by adding values for $\alpha = 0.5625(0.0625)0.9375$, excluding 0.75. Veselá (1964) gave values of $t_{\nu,\alpha}$ to four decimal places for $\nu = 30(1)120$ and $\alpha = 0.95, 0.975, 0.995$.

In Owen **(1962)** there are values of $t_{\nu,\alpha}$ to four decimal places for

$$\nu = 1(1)100, 150, 200(100)1000, \infty$$

and

$$a = 0.75, 0.90, 0.95, 0.975, 0.99, 0.995.$$

Also given are values of $t_{\nu,\alpha}$ to five decimal places for

$$\nu = 1(1)30(5)100(10)200, \infty$$

with

$$\alpha = 0.95, 0.975, 0.99, 0.995.$$

These tables are notable for the extensive series of values of ν.

Federighi (1959) concentrated on values of a very near to 1 (i.e. the extreme upper tail of the distribution). His tables give values of $t_{\nu,\alpha}$ to three decimal places for

$$\nu = 1(1)30(5)60(10)100, 200, 500, 1000, 2000, 10000, \infty$$

and

$$1 - \alpha = 0.25, 0.1, 0.05, 0.025, 0.01, 0.005, 0.0025, 10^{-3}, \tfrac{1}{2} \times 10^{-3}, \tfrac{1}{4} \times 10^{-3}$$
$$10^{-4}, \tfrac{1}{2} \times 10^{-4}, \tfrac{1}{4} \times 10^{-4}, 10^{-5}, \tfrac{1}{2} \times 10^{-5}, \tfrac{1}{4} \times 10^{-5},$$
$$10^{-6}, \tfrac{1}{2} \times 10^{-6}, \tfrac{1}{4} \times 10^{-6}, 10^{-7}.$$

Hill (1972) tabulated quantiles $t_{\nu,\alpha/2}$ of the Student t-distribution, corresponding to the two-tail probability levels

$$\alpha = 0.9(-0.1)0.1; \{5, 2, 1\} \times 10^{-2(-1)-10(-5)-30}$$

for $\nu = 1(1)30(2)50(5)100(10)150$, 200, (240, 300, 400, 600, 1200} $\times$ {1, 10, 100}, ∞ to 20 decimal places for $t_{\nu,\alpha/2} < 10^3$, 20 significant figures otherwise. He provides many ingenious approximations in the introduction of his tables.

Hald (1952) gave the values of $t_{\nu,\alpha}$ to three decimal places for

$$\nu = 1(1)30(10)60(20)100, 200, 500, \infty$$

and

$$a = 0.6(0.1)0.9, 0.95, 0.975, 0.99, 0.995, 0.999, 0.9995.$$

Values of the density function $p_{\nu}(t)$ are tabulated in Bracken and Schleifer (1964), Smirnov (1961), and Sukhatme (1938). The last reference provides seven decimal places for

$$\nu = 1(1)10, 12, 15, 20, 24, 30$$

with

$$t = 0.05(0.1)7.25.$$

In addition there are values for $\nu = 60$ with $t = 0.05(0.1)6.35$. The tables of Bracken and Schleifer (1964) cover greater ranges of values of each of the arguments (note especially the fractional value $\nu = 1.5$):

$$\nu = 1, 1.5, 2(1)10, 12, 15, 20, 24, 30, 40, 60, 120, \infty$$

and

$$t = 0.00(0.01)8.00.$$

The collection of Smirnov (1961) includes values of $p_{t_\nu}(t)$ and $\Pr[t_\nu \le t]$ to six decimal places for

$$\nu = 1(1)12 \quad \text{with} \quad t = 0.00(0.01)3.00(0.02)4.50(0.05)6.50$$

and for

$$\nu = 13(1)24 \quad \text{with} \quad t = 0.00(0.01)2.50(0.02)3.50(0.05)6.50.$$

This collection further includes values of $\Pr[t_\nu \le t]$, also to six decimal places, for

$$\nu = 1(1)10 \quad \text{with} \quad t = 6.5(0.1)9.0;$$

for

$$\nu = 25(1)35 \quad \text{with} \quad t = 0.00(0.01)2.50(0.02)3.50(0.05)5.00$$

and for

$$10^3\nu^{-1} = 30(-2)0 \quad \text{with} \quad t = 0.00(0.01)2.50(0.02)5.00.$$

(Note the extensive fractional values of ν.) There are also extensive tables of $t_{\nu,\alpha}$ to four decimal places for

$$\nu = 1(1)30(10)100, 120, 150(50)500(100)1000, 1500, 2000(1000)6000,$$
$$8000. 10.000. \infty$$

with

$$\alpha = 0.6, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975, 0.999, 0.9995.$$

(See also Section 4.) These tables include values of the multiplying constant $K_\nu = (\pi\nu)^{-1/2}\Gamma(\frac{1}{2}(\nu+1))/\Gamma(\frac{1}{2}\nu)$ and $\log K_\nu$ to ten significant figures for $\nu = 1(1)24$.

Cotterman and Knoop's (1968) tables (to five decimal places) provide boundary values $T_1(p)$, $T_2(p)$ such that, to three decimal places, $\Pr[t_\nu \geq T]$ is equal to p [$p = 0.000(0.001)0.500$] for any value of T between $T_1(p)$ and $T_2(p)$, for $\nu = 1(1)15$. Laumann (1967) gives $\Pr[t_\nu \leq t]$ to seven decimal places for $t = 0(0.01)4.50$, $\nu = 20(2)40(10)100(20)200$, 300, 500, 1000. The tables of Mardia and Zemroch (1978) include values of $t_{\nu,\alpha}$ to five significant figures for $\nu = 0.1(0.1)3(0.2)7(0.5)11(1)40$, 60, 120, ∞ and $1 - a = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.03(0.01)0.1, 0.2, 0.25, 0.3, 0.4. Note that many fractional values of ν are included, which is convenient when a t-distribution is used as an approximation.

Kafadar and Tukey (1988) note that $t_{\nu,\alpha}$ is approximately a linear function of $\log(1 - a)$, and they propose to use tables exploiting this fact to make linear interpolation more effective. They introduce the index

$$-\log_G(1-\alpha) \qquad \text{with } G = 10^{-0.1}$$

and term the units *decigalts* ("galt" in honor of Fancis Galton). The base (G) of the logarithms is chosen to make commonly used values of a correspond to nearly (or exactly) integer decigalt values (e.g., the values corresponding to a = 0.95, 0.975, and 0.99 are 13.0103, 16.0205, and 20, respectively). Kafadar and Tukey also introduce ***bidec values*** of $1 - a = 2^j 10^{-k}$ for integer values j, k and present a "bidec table" showing $t_{\nu,\alpha}$ to three decimal places, with a taking values of form $1 - 2^j 10^{-k}$, and also showing the approximate equivalent decigalt values. (The paper also includes further useful instructions on interpolation.)

Tiku and Kumra (1985) have published tables of expected values and variances and covariances of order statistics for Student's t-distribution. The values are given for $p = \frac{1}{2}(\nu + 1) = 2(0.5)10$ and n (sample size) ≤ 20. Expressions for n > 20 are presented in Tiku and Suresh (1992).

3.2 Nomograms

The preceding list shows that the t_ν distributions have been thoroughly tabulated. The available tables are more than sufficient for almost all applications. However, practical situations arise that call for quick evaluation of values of $t_{\nu,\alpha}$ or $\Pr[t_\nu \leq t]$. In such cases it is useful to have a reliable graphical method of determining the required value with sufficient accuracy.

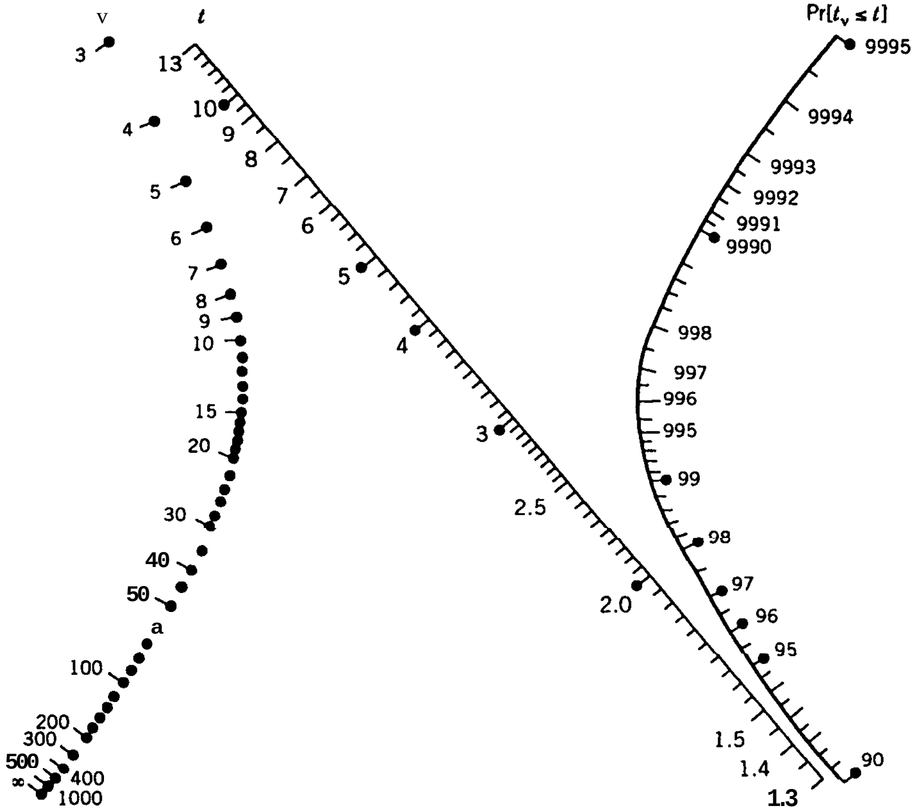


Figure 28.2 James-Levy's Nomogram. ν is the number of degrees of freedom. Given any two of ν, t or $\Pr[t_\nu \le t]$ we can determine the third.

James-Levy (1956) gave a nomogram relating ν, t, and $\Pr[t_\nu \le t]$. It is reproduced in Figure 28.2. The nomogram is used by placing a straightedge joining given values of any two of these quantities. The intersection with the third line then gives the required value of the third quantity. To find $t_{\nu,\alpha}$, for example, the appropriate points on the ν and $\Pr[t_\nu \le t]$ lines are joined, and the intersection with the t line gives $t_{\nu,\alpha}$. With $\alpha = 0.950 - 0.999$, an accuracy of about 0.001 in $t_{\nu,\alpha}$ can usually be attained.

Stammberger (1967) published a simple nomogram from which the value of any one of ν, $\Pr[t_\nu > t]$, and t, given the values of the other two quantities, can be read off using a straight edge. This nomogram is reproduced in Figure 28.3.

Babanin (1952) provided a nomogram (or abac) from which values can be read directly, without using a straightedge. However, this nomogram is not as simple as those of James-Levy and Stammberger.

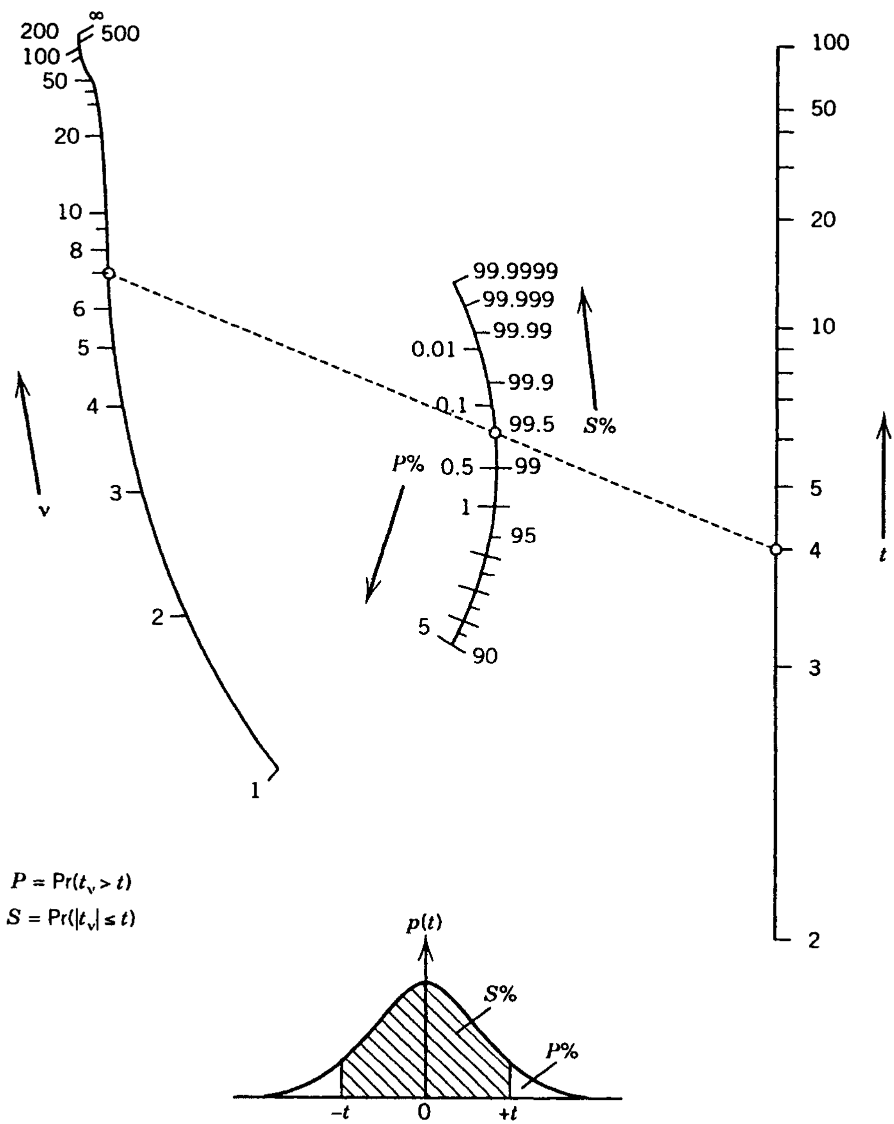


Figure 28.3 Stammberger's Nomogram

4 APPROXIMATIONS

There has been an intense study of possible approximations to t-distributions. Some of the approximations produced have very high accuracy, and some are rather complicated. For our present purpose simplicity, as well as accuracy, is an important factor in assessing the value of an approximation.

We have decided to include some complicated approximations in view of their substantial historical value.

The simplest approximation is made by regarding t_ν as a unit normal variable. This is very crude unless ν is at least 30, and it is unsatisfactory for considerably larger values of ν if extreme tails (e.g., $|t_\nu| > 4$) of the distribution are being considered. The simple modification of standardizing the t variable, and regarding $t_\nu\sqrt{1-2\nu^{-1}}$ as having a unit normal distribution [suggested by Weir (1960a)] effects a substantial improvement, but the approximation is still only moderately good (for analytical purposes) if ν is less than 20 (or if extreme tails are being considered). As in Section 3, we present a list in very rough chronological order.

Fisher (1925b) gave a direct expansion of the probability density, and hence of $\Pr[t_\nu \le t]$, as a series in ν^{-1}. For the probability density function he gives the formula

$$p_t(t) = \phi(t)\left[1 + \frac{1}{4}(t^4 - 2t^2 - 1)\nu^{-1} + \frac{1}{96}(3t^8 - 28t^6 + 30t^4 + 12t^2 + 3)\nu^{-2} + \frac{1}{384}(t^{12} - 22t^{10} + 113t^8 - 92t^6 - 33t^4 - 6t^2 + 15)\nu^{-3} + \cdots\right], \tag{28.14}$$

where $\phi(t) = (\sqrt{2\pi})^{-1}\exp(-\tfrac{1}{2}t^2)$. Integrating (28.14), we obtain

$$\Pr[t_\nu \ge t] \doteqdot 1 - \Phi(t) + \Phi(t)\left[\frac{1}{4}t(t^2+1)\nu^{-1} + \frac{1}{96}t(3t^6 - 7t^4 - 5t^2 - 3)\nu^{-2} + \frac{1}{384}t(t^{10} - 11t^8 + 14t^6 + 6t^4 - 3t^2 - 15)\nu^{-3}\right]. \tag{28.15}$$

The maximum absolute error of this approximation is shown in Table 28.2.

Fisher and Cornish (1960) inverted this series, obtaining the approximation

$$t_{\nu,\alpha} \doteqdot U_\alpha + \frac{1}{4}U_\alpha(U_\alpha^2 + 1)\nu^{-1} + \frac{1}{96}U_\alpha(5U_\alpha^4 + 16U_\alpha^2 + 3)\nu^{-2} + \frac{1}{384}U_\alpha(3U_\alpha^6 + 19U_\alpha^4 + 17U_\alpha^2 - 15)\nu^{-3}. \tag{28.16}$$

Dickey (1967) obtained an asymptotic (divergent) series approximation for

Table 28.2 **Maximum absolute error of (28.15), $c(d) \equiv c \times 10^d$**

ν	Max |Error|
5	2.8(−3)
6	1.4(−3)
7	8.2(−4)
8	5.0(−4)
9	3.2(−4)
10	2.2(−4)
11	1.5(−4)
12	1.1(−4)
13	8.0(−5)
14	6.1(−5)
15	4.7(−5)
20	1.5(−5)
25	6.5(−6)
30	3.2(−6)
35	1.7(−6)
40	1.0(−6)
45	6.4(−7)
50	4.3(−7)
60	2.1(−7)
80	6.6(−8)
100	2.7(−8)
120	1.3(−8)

Source: Ling (1978).

$t < \sqrt{\nu}$ in terms of Appell polynomials $A_r(x)$ defined by the identity

$$e^{-x}(1-ux)^{-1/u} = \sum_{r=0}^{\infty} A_r(x)u^r, \qquad |ux| < 1. \tag{28.17}$$

Setting $u = z(\nu+1)^{-1}$, $x = \frac{1}{2}(1+\nu^{-1})t^2$, we have

$$\left[1+\frac{t^2}{\nu}\right]^{-(\nu+1)/2} = e^{-(1+\nu^{-1})t^2/2}\sum_{r=0}^{\infty} A_r\left(-\frac{1}{2}(1+\nu^{-1})t^2\right)\left[\frac{2}{\nu+1}\right]^r,$$

$$t^2 < \nu. \qquad (28.17)'$$

Dickey gives the table of coefficients $B_{r,j}$ in $A_r(x) = x^r\Sigma_{j=0}^r B_{r,j}x^j$ presented in Table **28.3.**

Hendricks (1936) also approximated directly to the probability density function, obtaining

$$p_{t_\nu}(t) \doteqdot 2\nu c_\nu\sqrt{\frac{\nu+1}{\pi}}\,[t^2+2\nu]^{-3/2}\exp\left[-(\nu+1)c_\nu^2t^2(t^2+2\nu)^{-1}\right], \tag{28.18}$$

Table 28.3 Values of coefficients $B_{r,j}$ in Appell polynomials, $A_r(x) = x^r \sum_{j=0}^{r} B_{r,j} x^j$

r	j = 0	1	2	3	4	5	6
0	1						
1	2	$\frac{1}{2}$					
2	0	$\frac{1}{3}$	$\frac{1}{8}$				
3	0	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{48}$			
4	0	$\frac{1}{5}$	$\frac{13}{72}$	$\frac{1}{24}$	$\frac{1}{384}$		
5	0	$\frac{1}{6}$	$\frac{11}{60}$	$\frac{17}{288}$	$\frac{1}{144}$	$\frac{1}{3840}$	
6	0	$\frac{1}{7}$	$\frac{29}{160}$	$\frac{59}{810}$	$\frac{7}{576}$	$\frac{1}{1152}$	$\frac{1}{46080}$

Source: Dickey (1967).

where

$$c_\nu = 1 - \frac{3}{4}(\nu + 1)^{-1} - \frac{7}{32}(\nu + 1)^{-2}.$$

This gives quite good results in the "center" of the distribution ($|t| < 2$) but not in the tails.

Formula (28.18) is equivalent to the approximation $\sqrt{2(\nu + 1)}\, c_\nu t_\nu \times [t_\nu^2 + 2\nu]^{-1/2}$ has a unit normal distribution. In practice $\sqrt{2(\nu + 1)}\, c_\nu$ may be replaced by $\sqrt{2\nu - 1}$ unless ν is small. Some numerical comparisons are shown in Table *28.4*.

Another approximation, of a similar nature, was obtained by Elfving (1955):

$$\Pr[t_\nu \leq t] \doteqdot \Phi(\sigma t) + \frac{5}{96}\left(\sigma t^5 \nu^{-2}\left(1 + \tfrac{1}{2}t^2\nu^{-1}\right)^{-(\nu+4)/2} - \Phi(\sigma t/\sqrt{2}\right), \tag{28.19}$$

where

$$\sigma = \left[\frac{\nu - \frac{1}{2}}{\nu + \frac{1}{2}t^2}\right]^{1/2}.$$

Table 28.4 **Comparison of Hendricks's approximation in (28.18) with the exact value**

	α	Values of $t_{\nu,\alpha}$: Exact Value	Hendricks's Approximation
$\nu = 9$	0.55	0.129	0.129
	0.65	0.398	0.398
	0.75	0.703	0.703
	0.85	1.100	1.104
	0.95	1.833	1.844
	0.975	2.262	2.290
	0.99	2.821	2.869
	0.995	3.250	3.389
$\nu = 29$	0.55	0.127	0.127
	0.65	0.389	0.389
	0.75	0.683	0.683
	0.85	1.055	1.055
	0.95	1.699	1.700
	0.975	2.045	2.047
	0.99	2.462	2.466
	0.995	2.756	2.764

The error can be shown to be less than $\frac{1}{2}\nu^{-2}$ times the true value of $\Pr[t_\nu \le t]$, for all values of t. Hotelling and Frankel (1938) sought to find a function of t_ν with a distribution that is well approximated by the unit normal distribution. The leading terms of their series (which is in fact a Cornish-Fisher form of expansion) are

$$U = t\left\{1 - \frac{1}{4}(t^2+1)\nu^{-1} + \frac{1}{96}(13t^4 + 8t^2 + 3)\nu^{-2} - \frac{1}{384}(35t^6 + 19t^4 + t^2 - 15)\nu^{-3} + \frac{1}{92160}(6271t^8 + 3224t^6 - 102t^4 - 1680t^2 - 945)\nu^{-4}\right\}. \quad (28.20)$$

The successive terms rapidly become more complicated. Table 28.5 [taken from Hotelling and Frankel (1938)] gives values of U_α corresponding to $t_{\nu,\alpha}$ for various values of ν and α using the first two (x_1), three (x_2), four (x_3), and five (x_4) terms of (28.20). The correct values of U_α are also shown. For extreme tails poor results are obtained with $\nu = 10$, but even for extreme tails ($\alpha = 0.99995$) quite good results are obtained with $\nu \ge 30$, if five terms are used.

Among other investigations of expansions of Cornish-Fisher type we note the work of Peiser (1943) and Goldberg and Levine (1946). Peiser used the

Table 28.5 Approximations—Values of U_α using expansion (28.20)

	0.95		0.975		0.995		0.9995		0.99995		
a	$\nu = 10$	$\nu = 30$	$\nu = 10$	$\nu = 30$	$\nu = 10$	$\nu = 30$	$\nu = 10$	$\nu = 30$	$\nu = 10$	$\nu = 30$	$\nu = 100$
$t_{\nu,\alpha}$	1.812	1.697	2.228	2.042	3.169	2.750	4.587	3.646	6.22	4.482	4.052
x_1	1.618	1.642	1.896	1.954	2.294	2.554	2.059	3.212	0.05	3.69	3.88
x_2	1.650	1.645	1.980	1.960	2.754	2.579	4.981	3.313	12.86	3.98	3.89
x_3	1.643	1.645	1.953	1.960	2.446	2.575	0.896	3.283	20.44	3.85	3.89
x_4							7.163	3.293	75.66	3.91	3.89
U_α	1.645		1.960		2.576		3.291		3.891		

Table 28.6 **Approximation to $t_{\nu,\alpha}$ using (28.21)**

ν		α 0.9875	0.975	0.95	0.875	0.75
10	(28.21)	2.579	2.197	1.797	1.212	0.700
	Exact	2.634	2.228	1.813	1.221	0.700
30	(28.21)	2.354	2.039	1.696	1.171	0.683
	Exact	2.360	2.042	1.697	1.173	0.683
60	(28.21)	2.298	2.000	1.670	1.161	0.679
	Exact	2.299	2.000	1.671	1.162	0.679
120	(28.21)	2.270	1.980	1.658	1.156	0.677
	Exact	2.270	1.980	1.658	1.156	0.677

Source: Peiser (1943).

simple formula

$$t_{\nu,\alpha} \doteqdot U_\alpha + \frac{1}{4}(U_\alpha^3 + U_\alpha)\nu^{-1}. \tag{28.21}$$

Table *28.6* [taken from Peiser (1943)] shows that this gives useful results for $\nu \geq 30$.

Goldberg and Levine included one further term in the expansion, giving

$$t_{\nu,\alpha} \doteqdot U_\alpha + \frac{1}{4}(U_\alpha^3 + U_\alpha)\nu^{-1} + \frac{1}{96}(5U_\alpha^5 + 16U_\alpha^3 + 3U_\alpha)\nu^{-2}. \tag{28.22}$$

The next two terms in the series are

$$\frac{1}{384}(3U^7 + 19U^5 + 17U^3 - 15U)\nu^{-3}$$
$$+ \frac{1}{92160}(79U^9 + 776U^7 + 1482U^5 - 1920U^3 - 945U)\nu^{-4}$$

[Abramowitz and Stegun (*1964,* p. 949).] Table *28.7,* taken from Goldberg and Levine (1946), compares exact values with approximate values calculated from *(28.22).* (The original tables also give values obtained using only the first two terms, as in Peiser's work.) Inclusion of the third term considerably improves the approximation, which is now reasonably good for ν as small as *20.* Simaika *(1942)* improved the approximation $U = t_\nu\sqrt{1 - 2\nu^{-1}}$ by introducing higher powers of U. The approximation $\sinh^{-1}(t_\nu\sqrt{3\nu^{-1}/2})$ with a unit normal distribution [Anscombe (1950)] is a special case of a transformation of noncentral t (Chapter 31). It is not much used for central t.

Table 28.7 Comparative table of approximate and exact values of the percentage points of the t-distribution

Probability Integral (α)	Degrees of Freedom (ν)	Approximate Percentage Point[a]	Exact Percentage Point
0.9975	1	21.8892	127.32
	2	9.1354	14.089
	10	3.5587	3.5814
	20	3.1507	3.1534
	40	2.9708	2.9712
	60	2.9145	2.9146
	120	2.8599	2.8599
0.9950	1	16.3271	63.657
	2	7.2428	9.9248
	10	3.1558	3.1693
	20	2.8437	2.8453
	40	2.7043	2.7045
	60	2.6602	2.6603
	120	2.6174	2.6174
0.9750	1	7.1547	12.706
	2	3.8517	4.3027
	10	2.2254	2.2281
	20	2.0856	2.0860
	40	2.0210	2.0211
	60	2.0003	2.0003
	120	1.9799	1.9799
0.9500	1	4.5888	6.3138
	2	2.7618	2.9200
	10	1.8114	1.8125
	20	1.7246	1.7247
	40	1.6838	1.6839
	60	1.6706	1.6707
	120	1.6577	1.6577
0.7500	1	0.9993	1.0000
	2	0.8170	0.8165
	10	0.6998	0.6998
	20	0.6870	0.6870
	40	0.6807	0.6807
	60	0.6786	0.6786
	120	0.6765	0.6766

[a]From (28.22).

Chu (1956) obtained the following inequalities, which aid in assessing the accuracy of simple normal approximations to the distribution of t_ν, for $A \le 0$, $B \ge 0$, $\nu \ge 3$:

$$\frac{\nu}{\nu+1}\left[\Phi\left(B\sqrt{\frac{\nu+1}{\nu}}\right)-\Phi\left(A\sqrt{\frac{\nu+1}{\nu}}\right)\right] \le \Pr[A < t_\nu < B]$$
$$\le \sqrt{\frac{7\nu-3}{7\nu-14}}\left[\Phi\left(B\sqrt{\frac{\nu-2}{\nu}}\right)-\Phi\left(A\sqrt{\frac{\nu-2}{\nu}}\right)\right]. \tag{28.23}$$

He showed that for ν large, the ratio of absolute error to correct value of $\Pr[A < t_\nu < B]$ using the unit normal approximation to t_ν is less than ν^{-1}.

Wallace (1959), developing the methods used by Chu (1956), obtained bounds for the cumulative distribution function of t_ν. These are most easily expressed in terms of bounds on the corresponding normal deviate $u(t)$, defined by the equation (with $t > 0$)

$$\Phi(u(t)) = \Pr[t_\nu \le t]. \tag{28.24}$$

Wallace summarized his results as follows:

$$u(t) \le \left[\nu \log(1 + t^2\nu^{-1})\right]^{1/2}, \tag{28.25a}$$

$$u(t) \ge \left(1 - \tfrac{1}{2}\nu^{-1}\right)^{1/2}\left[\nu \log(1 + t^2\nu^{-1})\right]^{1/2} \quad \text{for } \nu > f, \tag{28.25b}$$

and also

$$u(t) \ge \left[\nu \log(1 + t^2\nu^{-1})\right]^{1/2} - 0.368\nu^{-1/2} \quad \text{for } \nu \ge \tfrac{1}{2}. \tag{28.25c}$$

From (28.25a) and (28.25b) it can be seen that $[\nu \log(1 + t^2\nu^{-1})]^{1/2}$ differs from $u(t)$ by an amount not exceeding $25\nu^{-1}\%$; (28.25c) shows that the absolute error does not exceed $0.368\nu^{-1/2}$. Usually (28.25b) gives a better (i.e., greater) lower bound than (28.25c).

Wallace further obtained two good approximations without giving precise bounds for (28.27):

$$\left(\frac{8\nu+1}{8\nu+3}\right)\left[\nu \log(1 + t^2\nu^{-1})\right]^{1/2}, \tag{28.26}$$

$$\left[1 - \frac{2}{8\nu+3}\{1 - e^{-s^2}\}^{1/2}\right]\left[\nu \log(1 + t^2\nu^{-1})\right]^{1/2}, \tag{28.27}$$

with

$$s = 0.184(8\nu + 3)\nu^{-1}\left[\log(1 + t^2\nu^{-1})\right]^{-1/2}.$$

He stated that (28.27) is within 0.02 of the true value of $u(t)$ for a wide range of values of t. Prescott (1974) advocated use of (28.26).

Wallace compared the values given by (28.25a), (28.25b), (28.25c), (28.26), and (28.27) and by the formula corresponding to the Paulson approximation to the F-distribution (Chapter 27) (setting $\nu_1 = 1$):

$$\Pr[|t_\nu| \le t] = \Pr\left[U \le \frac{1}{3\sqrt{2}}\left[(9 - 2\nu^{-1})t^{2/3} - 7\right]\left[\nu^{-1}t^{4/3} + 1\right]^{-1/2}\right]. \tag{28.28}$$

His results are shown in Table 28.8.

The accuracy of (28.27) is noteworthy, though the formula is rather complicated. Peizer and Pratt (1968) have proposed other formulas of this kind:

$$u(t) \doteqdot \left(\nu - \frac{2}{3}\right)\left[\frac{\log(1 + t^2\nu^{-1})}{\nu - (5/6)}\right]^{1/2} \tag{28.29a}$$

and

$$u(t) \doteqdot \left(\nu - \frac{2}{3} + \frac{1}{10}\nu^{-1}\right)\left[\frac{\log(1 + t^2\nu^{-1})}{\nu - (5/6)}\right]^{-1}. \tag{28.29b}$$

Gaver and Kafadar (1984) provide a simple approximation to percentage points of t-distribution very similar to (28.29b) of a somewhat superior accuracy.

Cornish (1969) reported Hill's (1969) Cornish-Fisher type expansions in terms of $u = \{a_\nu \log(1 + t^2\nu^{-1})\}^{1/2}$ and $a_\nu = \nu - \frac{1}{2}$, as follows:

$$u(t) = u + \frac{1}{48}(u^3 + 3u)a_\nu^{-2} - \frac{1}{23040}(4u^7 + 33u^5 + 240u^3 + 855u)a_\nu^{-4}, \tag{28.30a}$$

and inversely

$$t_{\nu,\alpha} = \left[\nu\left(\exp\left[u'^2 a_\nu^{-1}\right] - 1\right)\right]^{1/2}, \tag{28.30b}$$

where

$$u' = U_\alpha - \frac{1}{48}(U_\alpha^3 + 3U_\alpha)a_\nu^{-2} + \frac{1}{23040}(4U_\alpha^7 + 63U_\alpha^5 + 360U_\alpha^3 + 945U_\alpha)a_\nu^{-4}.$$

Table 28.8 Bounds on the equivalent normal deviate $u(t)$ for t_ν

			$u(t)$ Bounds			Approximations		
ν	t	Exact Value	Eq. (28.25a)	Eq. (28.25b)	Eq. (28.25c)	Eq. (28.26)	Eq. (28.27)	Paulson Eq. (28.28)
1	0.3	0.235	0.294	0.208	< 0	0.241	0.241	0.257
	1	0.674	0.832	0.589	0.465	0.680	0.681	0.674
	2	1.047	1.269	0.897	0.901	1.038	1.048	1.031
	4	1.419	1.683	1.190	1.315	1.377	1.416	1.349
	8	1.756	2.043	1.445	1.675	1.672	1.750	1.576
	12	1.935	2.231	1.577	1.863	1.825	1.927	1.670
	10^2	2.729	3.035	2.146	2.667	2.177	2.704	1.896
	10^5	4.514	4.799	3.393	4.431	3.926	4.447	1.964
3	1	0.858	0.929	0.848	0.717	0.860	0.860	0.855
	2	1.478	1.594	1.455	1.382	1.476	1.478	1.477
	4	2.197	2.353	2.148	2.141	2.179	2.197	2.160
	8	2.872	3.053	2.787	2.840	2.826	2.879	2.705
	12	3.228	3.417	3.119	3.204	3.164	3.237	2.953
	$\sqrt{3}\,10^2$	5.057	5.256	4.797	5.044	4.866	5.058	3.493
10	1	0.952	0.976	0.952	0.860	0.953	0.953	0.948
	2	1.790	1.834	1.788	1.718	1.790	1.790	1.805
	4	3.021	3.091	3.013	2.975	3.017	3.020	3.014
	8	4.382	4.474	4.361	4.357	4.366	4.384	4.279
	12	5.128	5.229	5.097	5.113	5.103	5.133	4.902
100	100	21.447	21.483	21.429	21.446	21.429	21.450	18.541

Source: Wallace (1959).

Mickey *(1975)* suggested the approximation

$$u(t) \doteq \left\{\left(\nu - \frac{1}{2}\right)\log\left(1 + \frac{t^2}{\nu}\right)\right\}^{1/2} \tag{28.31}$$

which is a modified Chu (1956) transform approximation; see page *382*.

These investigations suggest a general class of transformations of form

$$u(t) \doteq \left\{\frac{\nu + b}{\nu + c}\right\}\left[(\nu - a)\log\left\{1 + \frac{t^2}{\nu + h}\right\}\right]^{1/2} \tag{28.32}$$

If, in Wallace's *(1959)* or Mickey's (1975) transformations, t is expressed as a function of $u(t)$ and them expanded in a power series, it agrees with Fisher's expansion up to $O(\nu^{-1})$. Bailey (1980) showed that by choosing $a = -\frac{19}{12}$, $b = \frac{1}{8}$, $c = \frac{9}{8}$, and $h = \frac{1}{12}$ in (28.32), agreement up to $O(\nu^{-2})$ can be achieved, and the approximation

$$u(t) = \left\{\frac{8\nu + 1}{8\nu + 9}\right\}\left[\left(\nu + \frac{19}{12}\right)\log\left\{1 + \frac{t^2}{\nu + \frac{1}{12}}\right\}\right]^{1/2} \tag{28.33}$$

is accurate to $O(\nu^{-3})$ locally in the vicinity of $u(t) = 1.9469$. This value is likely to be of interest in many applications.

Bailey also suggested

$$u(t) \doteq \frac{4\nu^2 + 5(2z_c^2 + 3)/24}{4\nu^2 + \nu + (4z_c^2 + 9)/12}\left\{\nu\log\left(1 + \frac{t^2}{\nu}\right)\right\}^{1/2}, \tag{28.34}$$

where z_c is the unit normal $(\nu \to \infty)$ value for $u(t)$. For $z_c = 1.96$, 2.5758, and 3.2905, very high accuracy up to $O(\nu^{-3})$ is achieved and for $z_c = 2.3276$ the accuracy improves to $O(\nu^{-4})$.

Soms *(1984)* extended the bounds of Birnbaum *(1942)* and Sampford (1953) (see Chapter *13* and Chapter *33,* Section *7.1* in the first edition) for the ratio of the upper tail area of the normal distribution to the upper tail of the t-distribution by showing that

$$R_{\nu,x} = \int_x^\infty f_\nu(t)\,dt/f_\nu(x), \tag{28.35}$$

where $f_\nu(x)$ is the Student density with ν degrees of freedom satisfies the inequalities

$$\left[\frac{(\nu-1)x}{2\nu}+\left(1+\left\{\frac{(\nu+1)x}{2\nu}\right\}^2\right)^{1/2}\right]^{-1}<R_{\nu,x}$$

$$<\left[\frac{(3\nu-1)x}{4\nu}+\left(\frac{\nu-1}{2\nu}+\left\{\frac{(\nu+1)x}{4\nu}\right\}^2\right)^{1/2}\right]^{-1}, \qquad (28.36)$$

where the lower bound is valid for all $\nu \geq 1$ and the upper for $\nu \geq 2$ (not necessarily integer-valued).

An earlier result of Soms *(1976)* is

$$\frac{1}{x}-\frac{\nu}{(\nu+2)x^3}<R_{\nu,x}<\frac{1}{x} \qquad \text{for all } \nu>0. \qquad (28.37)$$

Soms *(1984)* gave

$$A(x,\gamma_{\min})<R_{\nu,x}<A(x,\gamma_{\max}), \qquad (28.38)$$

where

$$A(x,\gamma)=\frac{1+\gamma}{\{x^2+4c_\nu^2(1+\gamma)^2\}^{1/2}+\gamma x} \qquad \text{for } \nu>2,$$

$$\gamma_{\min}=\frac{\nu}{2(\nu+2)c_\nu^2}-1,$$

$$\gamma_{\max}=4c_\nu^2/(1-4c_\nu^2).$$

For $\nu<2$ the definitions of $\gamma_{\min}$ and $\gamma_{\max}$ are interchanged, and for $\nu=2$, $\gamma_{\min}=\gamma_{\max}$ and $R_{\nu,x}=A(x,\gamma_{\max})$. Soms compared his bounds with those of Peizer and Pratt *(1968)* and Wallace *(1959)* without reaching definite conclusions. In general the lower bound in *(28.38)* is the best while the picture is mixed for upper bounds.

Among formulas developed empirically, we mention results of four investigations, reported by Cucconi (1962), Gardiner and Bombay (1965), Moran (1966), and Kramer *(1966)*. Gardiner and Bombay *(1965)* gave formulas of the form

$$t_{\nu,\alpha}=(a\nu+b+c\nu^{-1})(\nu+d+e\nu^{-1})^{-1} \qquad (28.39)$$

for various percentiles. Values of a, b, c, d, and e for $\alpha=0.95, 0.975$, and 0.995 are shown in Table *28.9* (note that $a=U_\alpha$). The corresponding values of $t_{\nu,\alpha}$ are correct to four decimal places for $\nu>3$. These results are rather

Table 28.9 Values of a, *b*, c, d, and e for the approximation in (28.39)

α	a	*b*	*c*	*d*	*e*
0.95	1.6449	3.5283	0.85602	1.2209	−1.5162
0.975	1.9600	0.60033	0.95910	−0.90259	0.11588
0.995	2.5758	−0.82847	1.8745	−2.2311	1.5631

better, for smaller values of v, than those obtained with Cucconi's (1962) formulas:

$$t_{\nu,0.975} \doteqdot 1.9600\nu(\nu^2 - 2.143\nu + 1.696)^{-1/2}, \qquad \nu > 1, \quad (28.40a)$$

$$t_{\nu,0.995} \doteqdot 2.5758\nu(\nu^2 - 3.185\nu + 4.212)^{-1/2}, \qquad \nu > 2. \quad (28.40b)$$

Moran's (1966) investigations were confined to comparisons at 2.5%, 0.5% and 0.05% percentiles, and led to the following approximations:

At the 2.5% point,

$$U_{0.975} = t_{\nu,0.975}(1 - 0.0550t_{\nu,0.975}\nu^{-1}) \qquad (28.41)$$

and

$$U_{0.975} = t_{\nu,0.975}(1 - 0.6049t_{\nu,0.975}\nu^{-1} + 0.2783t^2_{\nu,0.975}\nu^{-2}). \quad (28.42)$$

At 0.5%,

$$U_{0.995} = t_{\nu,0.995}[1 + 0.613t_{\nu,0.995}\nu^{-1}]^{-1} - 0.8\nu^{-1} \qquad (28.43)$$

At 0.05%,

$$U_{0.9995} = t_{\nu,0.9995}[1 + 0.87t_{\nu,0.9995}\nu^{-1}]^{-1}. \qquad (28.44)$$

The formula

$$U_\alpha = t_{\nu,\alpha}[1 + 0.613t_{\nu,\alpha}\nu^{-1}]^{-1} \qquad (28.45)$$

was found to give fairly good results both for a = 0.975 and a = 0.995. Values given by (28.42), (28.43), and (28.45) for α = 0.975, 0.995, are shown in Table 28.10.

Table 28.10 **Comparison of approximations (28.42), (28.43), and (28.45) with exact values of U_α**

	$\alpha = 0.975$, $U_\alpha = 1.9600$		$\alpha = 0.995$, $U_\alpha = 2.5758$	
ν	Eq. (28.42)	Eq. (28.45)	Eq. (28.45)	Eq. (28.43)
3	2.1369	1.9284	2.6628	2.3961
4	1.9830	1.9477	2.6994	2.4994
5	1.9603	1.9546	2.6983	2.5383
6	1.9565	1.9575	2.6889	2.5556
8	1.9573	1.9597	2.6691	2.5691
10	1.9586	1.9604	2.6537	2.5737
15	1.9603	1.9607	2.6300	2.5767
20	1.9607	1.9606	2.6171	2.5771
30	1.9608	1.9605	2.6037	2.5770
40	1.9607	1.9604	2.5969	2.5769

Formula (28.45) gives reasonably good results for $\nu \geq 10$. Kramer's (1966) approximations are based on unpublished results obtained by Ray (1961). He stated that the following formulas have errors less than 0.001 for $3 < \nu < 120$: For $0 < t < 1$,

$$\begin{aligned}\Pr[0 < t_\nu < t] \doteqdot\ & 0.399622t - 0.068492t^3 \\ & + 0.010272t^5 - 0.111604t\nu^{-1} \\ & - 0.009310t^3\nu^{-1} + 0.02865t\nu^{-2}. \end{aligned} \tag{28.46a}$$

For $1 \leq t \leq 2$,

$$\begin{aligned}\Pr[0 < t_\nu < t] \doteqdot\ & -0.060820 + 0.585243t - 0.2089773t^3 \\ & + 0.025489t^5 + 0.082228\nu^{-1} \\ & - 0.276747t\nu^{-1} + 0.080726t^2\nu^{-1} \\ & + 0.011192t\nu^{-2}. \end{aligned} \tag{28.46b}$$

For $t > 2$,

$$\begin{aligned}\Pr[0 < t_\nu < t] \doteqdot\ & 0.503226 - 0.044928\nu^{-1} \\ & + 0.112057\nu^{-2} + 1.949790(\nu t^2)^{-1} \\ & - 5.917356(\nu^2 t^2)^{-1} - 7.549051(\nu t^3)^{-1} \\ & + 11.311627(\nu^2 t^3)^{-1} - 0.399205t^{-4} \\ & + 5.487170(\nu t^4)^{-1}. \end{aligned} \tag{28.46c}$$

Similar accuracy for $\nu = 1$ is obtained with the following formulas:
For $0 < t \leq \frac{1}{2}$,

$$\Pr[0 < t_1 < t] = \pi^{-1}\left(t - \tfrac{1}{3}t^3 + \tfrac{1}{5}t^5 - ft^7\right). \qquad (28.47a)$$

For $\frac{1}{2} < t < \frac{3}{2}$,

$$\text{Pl-}[0 < t, < t] = \tfrac{1}{4} + \pi^{-1}\Big[\tfrac{1}{2}(t-1) - \tfrac{1}{4}(t-1)2$$

$$+ \tfrac{1}{12}(t-1)^3 - \tfrac{1}{40}(t-1)^5\Big]. \qquad (28.47b)$$

For $t \geq \frac{3}{2}$,

$$\Pr[0 < \text{t}, < t] \doteq \tfrac{1}{2} - \pi^{-1}\left[t^{-1} - \tfrac{1}{3}t^{-3} + \tfrac{1}{5}t^{-5} - \tfrac{1}{7}t^{-7}\right]. \qquad (28.47c)$$

For $v = 2$ the formula

$$\Pr[0 < t_2 < r] = t(8 + 4t^2)^{-1/2} \qquad (28.48)$$

is exact for all $t \geq 0$. Formulas *(28.46)* and (28.47) are well suited for use in calculators.

For calculating extreme tail probabilities of the t_ν distribution, Pinkham and Wilk (1963) suggested using the expansion

$$\int_t^\infty \left(1 + y^2\nu^{-1}\right)^{-(\nu+1)/2} dy = \sum_{j=1}^{m} w_j + R_m(t), \qquad m < \tfrac{1}{2}(\nu + 1), \quad (28.49a)$$

where

$$w, = \nu(\nu - 1)^{-1}\left(1 + t^2\nu^{-1}\right)^{-(\nu-1)/2} t^{-1},$$

$$w_{j+1} = w_j(1 + \nu t^{-2})\frac{2j - 1}{2j + 1 - \nu}, \qquad j = 1, 2, \ldots, m - 1,$$

and the remainder term $R_m(t)$ does not exceed (in absolute value) w_m.

Table *28.11* [taken from Pinkham and Wilk (1963)] shows how good approximations can be obtained with this formula using only three terms in

Table 28.11 Exact and approximate tail areas for the t-distribution with ν degrees of freedom

Exact Tail Area[a]	Approximation (28.49a) with m = 3		
	$\nu = 7$	$\nu = 15$	$\nu = 40$
0.001	0.000 816	0.001 06	0.001 02
0.000 05	0.000 042 8	0.000 051 5	0.000 050 3
0.000 01	0.000 008 66	0.000 010 2	0.000 010 05
0.000 001	0.000 000 873	0.000 001 02	0.000 001 003
0.000 000 1	0.000 000 087 7	0.000 000 102	0.000 000 100 1

Source: Pinkham and Wilk (1963).
[a]These tail areas are exact to the extent that Federighi's (1959) tabled quantiles are exact.

the expansion (i.e., m = 3). An expansion of the integral (28.49a) in terms of $w = [1 + t^2\nu^{-1}]^{-1/2}$ was used by Hill (1970a) for $t^2 > n \geq 1$:

$$\int_t^{\infty}\left(1 + y^2\nu^{-1}\right)^{-(\nu+1)/2} dy = \nu^{1/2}\int_0^{w} x^{\nu-1}\left(1 - x^2\right)^{-1/2} dx$$

$$= \nu^{1/2}w^{\nu}\left\{\frac{1}{\nu} + \frac{w^2}{2(\nu+2)} + \frac{1\cdot 3w^4}{2\cdot 4(\nu+4)} + \cdots\right\} \qquad (28.49\text{b})$$

The inverse of this series was used to express $t^2\nu^{-1}$ in terms of $z = [\nu^{1/2}c_\nu\alpha]^{2/\nu}$, where c_ν is the normalizing constant of the probability integral, yielding the formula

$$t^2_{\nu,\alpha}\nu^{-1} \doteqdot z^{-1} + \frac{\nu+1}{\nu+2}\left[-1 + \frac{z}{2(\nu+4)} + \frac{z}{3(\nu+2)\left\{(\nu+6)(\nu z)^{-1} - 1\right\}}\right] \qquad (28.49\text{c})$$

which is exact for $\nu = 2$, and for larger ν correct to over six digits for $z < \nu^{-1}$.

Cornish (1969) reported Hill's (1969) Cornish-Fisher type expansions in terms of $u = [a_\nu \log(1 + t^2\nu^{-1})]^{1/2}$, where $a_\nu = \nu - (1/2)$ [see Eqs. (28.30a) and (28.30b)]. M. A. A. Cox (1991) has adapted Hill's (1970a) algorithm for evaluation of percentage points of the Student's t-distribution written in ALGOL code for use on a spreadsheet. The resulting code is entered via the macro facility. The package adopted is SYMPHONY©; however, the logic may be readily implemented in both LOTUS 123© and SUPER CALC©. In Zelen and Severo (1964) the following approximations are stated:

For $\nu \leq 5$, but t large,

$$\Pr[t_\nu \leq t] = 1 - a_\nu t^{-\nu} - b_\nu t^{-(\nu+1)} \tag{28.50}$$

with

$$a_1 = 0.3183,\ a_2 = 0.4991,\ a_3 = 1.1094,\ a_4 = 3.0941,\ a_5 = 9.948,$$

$$b_1 = 0.0000,\ b_2 = 0.0518,\ b_3 = -0.0460,\ b_4 = -2.756,\ b_5 = -14.05.$$

For ν large,

$$\Pr[t_\nu \leq t] \doteqdot \Phi\left[t\left(1 - \tfrac{1}{4}\nu^{-1}\right)\left(1 + \tfrac{1}{2}t^2\nu^{-1}\right)^{-1/2}\right]. \tag{28.51}$$

Gentleman and Jenkins (1968) published an approximation, suitable for computer use, of form

$$\Pr[|t_\nu| < t] \doteqdot \frac{1}{2}\left[1 + \sum_{j=1}^{5} c_j t^j\right]^{-8}, \tag{28.52}$$

where each c_j is the ratio of a quintic to a quadratic polynomial in ν^{-1}. This gives five decimal place accuracy for $\nu > 10$. Values of the c_j's (as functions of ν^{-1}) are shown in Table 28.12.

Taylor (1970) provided an algorithm for applying this method. For $\nu \geq 5$ the absolute error is always less than 0.001. Ling (1978) compared several approximations and demonstrated that for degrees of freedom from 5 to 45, the formula of Gentleman and Jenkins (1968) was best, according to his criterion of maximum absolute error, for tail areas between 0.0001 and 0.4999.

Continuing Ling's (1978) investigations, Lozy (1982) pointed out that the best approximation to Student's t-distribution is that of Hill (1970a, 1972, 1981). This involves a generalized Cornish-Fisher type expansion [Hill and Davis (1968)], in which

$$U = Z + (Z^3 + 3Z)b^{-1} - \tfrac{1}{10}(4Z^7 + 33Z^5 + 240Z^3 + 855Z)b^{-2} \tag{28.53}$$

with as above $Z = \{a \log(1 + \nu^{-1}t_\nu^2)\}^{1/2}$, $a = \nu - \frac{1}{2}$, $b = 48a^2$, is approximately a unit normal variable. Hill (1972, 1981) provided the first seven terms while Hill (1970a) pointed out that the contribution of the fourth term may be accounted for by replacing the denominator $(10b^2)$ of the third term by

$$10b(b + 0.8Z^4 + 10)$$

Lozy (1982) also compared Gentleman-Jenkins, Peizer-Pratt, Cornish-Fisher,

Table 28.12 Coefficients for the approximation in (28.52)

Coefficient	Numerator	Denominator
c_1	$0.009979441 - 0.581821\nu^{-1} + 1.390993\nu^{-2} - 1.222452\nu^{-3} + 2.151185\nu^{-1}$	$1 - 5.537409\nu^{-1} + 11.42343\nu^{-2}$
c_2	$0.04431742 - 0.2206018\nu^{-1} - 0.03317253\nu^{-2} - 5.679969\nu^{-3} - 12.96519\nu^{-4}$	$1 - 5.166733\nu^{-1} + 13.49862\nu^{-2}$
c_3	$0.009694901 - 0.1408854\nu^{-1} + 1.889930\nu^{-2} - 12.75532\nu^{-3} + 25.77532\nu^{-4}$	$1 - 4.233736\nu^{-1} + 14.39630\nu^{-2}$
c_4	$-0.00009187228 + 0.03789901\nu^{-1} - 1.280346\nu^{-2} - 9.249528\nu^{-3} - 19.08115\nu^{-4}$	$1 - 2.777816\nu^{-1} + 16.46132\nu^{-2}$
c_5	$0.0005796020 - 0.02763334\nu^{-1} + 0.4517029\nu^{-2} - 2.657697\nu^{-3} + 5.127212\nu^{-4}$	$1 - 0.5657187\nu^{-1} + 21.83269\nu^{-2}$

and various approximations of Hill (2-term, 3-term, and 3-term modified) and concluded:

> The Gentleman-Jenkins and Hill approximations are the only ones to give five correct decimal places for a small number of degrees of freedom, and one would probably opt for the simple Hill two-term approximation over the rather complicated Gentleman-Jenkins one. Since the two-term Hill approximation gives an accuracy of five decimals for eight degrees of freedom, as opposed to about 4–5 decimals for the Peizer-Pratt and Wallace approximations, and is no more complicated than they are, there would again seem to be no justification for using any of the latter.

Bukač and Burstein (1980) provided a table of coefficients for polynomial approximations of $t_{\nu,\alpha}$ for five values of α with proportional error less than 0.00005. The basis is Goldberg and Levine's (1946) approximation

$$t_{\nu,\alpha} = U_\alpha + \frac{U_\alpha^3 + U_\alpha}{4\nu} + \frac{5U_\alpha^5 + 16U_\alpha^3 + 34U_\alpha}{96\nu^2}. \qquad (28.22)'$$

For large $\nu \geq 120$ the authors use

$$t_{\nu,\alpha} = b_0 + b_1 x, \qquad \text{where } x = \nu^{-1},\ b_0 = U_\alpha, \qquad (28.54a)$$

and

$$b_1 = \frac{1}{4}(U_\alpha^3 + U_\alpha).$$

Then, for given ν and α, they approximate $t_{\nu,\alpha}$ by

$$R_N(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_N x^N, \qquad (28.54b)$$

where for specified N, $b_2, \cdots, b_N$, are computed so that the maximum relative error

$$\Delta_{\nu,\alpha} = \max \frac{|R_N(\nu^{-1}) - t_{\nu,\alpha}|}{t_{\nu,\alpha}} \qquad (28.54c)$$

is minimized for small ν. Table 28.13 provides the appropriate values of b_i $(i = 1, \dots, 6)$.

Sinclair (1980) notes that for large t, $\log \Pr[t_\nu > t]$ is approximately linear in $\log t$, with slope $-t$. For large values of t he suggests the approximation

$$\{2\Pr[t_\nu > t]\}^{-1/\nu} \doteqdot \alpha_\nu t + \beta_\nu t^{-1}, \qquad (28.55)$$

Table 28.13 **Coefficients** for polynomial approximation **(28.54b)** of Student's t percentage points ($t_{\nu,\alpha}$)

	α				
	0.900	0.950	0.975	0.990	0.995
b_0	1.28155	1.64485	1.95996	2.32635	2.57583
b_1	0.84658	1.52377	2.37227	3.72907	4.91655
b_2	0.57432	1.41902	2.80775	5.72289	8.86832
b_3	0.22086	1.00507	2.76386	6.61349	11.35729
b_4	0.15426	0.32789	0.69551	6.61683	17.92627
b_5	—	0.39338	2.10650	−0.22569	−9.45008
b_6	—	—	—	7.03691	27.46120
$\Delta_{\nu,\alpha}\times 10^4$	0.422	0.316	0.291	0.215	0.208

where

$$\alpha_\nu = 2\nu^{-1/2}\left\{\frac{1}{2}\nu B\left(\frac{1}{2}\nu, \frac{1}{2}\nu + 1\right)\right\}^{1/\nu}$$

and

$$\beta_\nu = \frac{1}{2}(\nu + 2)^{-1}\nu(\nu + 1)\alpha_\nu.$$

The values of α_ν and β_ν are chosen by equating the two leading terms in the Taylor series expansion of $\frac{1}{2}(\alpha_\nu t + \beta_\nu t^{-1})^{-\nu}$ to the two leading terms in the expansion, in negative powers of t, of $\Pr[t_\nu > t]$. The difference between the true value of $\Pr[t_\nu > t]$ and the approximation from **(28.55)** is of order $t^{-(\nu+4)}$. Since the value of $\Pr[t_\nu > t]$ is of order $t^{-\nu}$, the absolute relative error is of order t^{-4}. Comparisons between Pinkham and Wilk (1963), Mickey (1975), and Sinclair (1980) are provided in Sinclair (1980).

Richter and Gundlach (1990) have suggested the following approximations (for $\alpha > \frac{1}{2}$):

$$t_{\nu,\alpha} \doteqdot \nu^{1/2}\left[\left\{\nu^{-1/2}\gamma_1(\nu) + 1\right\}^{1/\gamma_1(\nu)}U_\alpha - 1\right] \tag{28.56}$$

and

$$t_{\nu,\alpha} \doteqdot U_\alpha + \tfrac{1}{2}\nu^{-1/2}\{\gamma_2(\nu)\}^{-1/\{1-\gamma_2(\nu)\}}\{1 - \gamma_2(\nu)\}(U_\alpha^2 - 1), \tag{28.57}$$

where $\gamma_1(\nu)$ and $\gamma_2(\nu)$ are appropriate constants and do not depend on α. They claim that these formulas are accurate to within $\pm 4 \times 10^{-5}$ (for $\nu \geq 4$). Table 28.14 [from Richter and Gundlach (1990)] presents some values of $\gamma_1(\nu)$ and $\gamma_2(\nu)$. As $\nu \to \infty$, $\gamma_j(\nu) \to 1$ ($j = 1, 2$), and it appears that, for $\nu > 33$, one might take

$$\gamma_j(\nu) \doteqdot 1 - \{1 - \gamma_j(33)\}\sqrt{(33/\nu)}\,.$$

Table 28.14 Values of $\gamma_1(\nu)$ and $\gamma_2(\nu)$ for (28.56) and (28.57)

ν	$\gamma_1(\nu)$	$\gamma_2(\nu)$	ν	$\gamma_1(\nu)$	$\gamma_2(\nu)$
4	0.3385	0.6629	19	0.7145	0.8539
5	0.4124	0.7036	20	0.7223	0.8576
6	0.4672	0.7324	21	0.7295	0.8611
7	0.5088	0.7541	22	0.7361	0.8642
8	0.5440	0.7713	23	0.7424	0.8673
9	0.5722	0.7852	24	0.7482	0.8701
10	0.5860	0.7967	25	0.7536	0.8729
11	0.6163	0.8067	26	0.7588	0.8752
12	0.6340	0.8152	27	0.7636	0.8775
13	0.6495	0.8227	28	0.7681	0.8797
14	0.6633	0.8293	29	0.7725	0.8818
15	0.6756	0.8553	30	0.7765	0.8837
16	0.6868	0.8406	31	0.7804	0.8856
17	0.6969	0.8454	32	0.7841	0.8874
18	0.7061	0.8498	33	0.7876	0.8891

Source: Richter and Gundlach (1990).

5 APPLICATIONS

The major applications of the t-distribution, construction of tests and confidence intervals relating to the expected values of normal distributions, have been discussed in Section 1.

In particular, if $X_1, X_2, \cdots, X_n$ are independent random variables, each having a normal distribution with expected value ξ and standard deviation σ, then the distribution of $\sqrt{n}(\bar{X} - \xi)/S'$ [where $\bar{X} = n^{-1}\Sigma_{j=1}^{n} X_j$ and $S'^2 = (n-1)^{-1}\Sigma_{j=1}^{n}(X_j - \bar{X})^2$] is a t-distribution with n – 1 degrees of freedom. Since

$$\Pr\left[|\sqrt{n}(\bar{X} - \xi)/S'| < t_{n-1,1-\frac{\alpha}{2}}\right] = 1 - \alpha,$$

it follows that

$$\Pr\left[\bar{X} - \left(t_{n-1,1-\frac{\alpha}{2}}/\sqrt{n}\right)S' < \xi < \quad + \left(t_{n-1,1-\frac{\alpha}{2}}/\sqrt{n}\right)S'\right] = 1 - \alpha. \quad (28.58)$$

So $\bar{X} \pm (t_{n-1,1-\frac{\alpha}{2}}/\sqrt{n})S'$ is a confidence interval for 5 with confidence coefficient $100(1 - \alpha)\%$. For practical purposes, it is convenient to have a table of multipliers $b_{n,\alpha} = t_{n-1,1-\frac{\alpha}{2}}/\sqrt{n}$ so that the limits of the interval are $\bar{X} \pm b_{n,\alpha}S'$. Table 28.15 contains a few values of $b_{n,\alpha}$.

In analysis of variance tests, when one of the sums of squares being compared has 1 degree of freedom, the appropriate null hypothesis distribution is F with 1, ν degrees of freedom, which is identical with the distribution of t_ν^2. Confidence limits for a single specified linear function of parameters in a general linear model (Chapter 27) corresponding to a single degree of freedom are constructed in a similar way to that just described.

Table 28.15 Values of $b_{n,\alpha}$

	n										
α	4	5	6	7	8	9	10	11	12	15	20
0.90	1.18	0.95	0.82	0.73	0.67	0.62	0.58	0.55	0.52	0.45	0.40
0.95	1.59	1.24	1.05	0.93	0.84	0.77	0.72	0.67	0.64	0.55	0.48
0.99	2.92	2.06	1.65	1.40	1.24	1.12	1.03	0.96	0.90	0.77	0.66

"Student's" distributions (along with stable distributions; see Chapter 12, Section 4) have been found to provide adequate models for description of changes in prices of speculative assets such as stocks. Some relevant references are Praetz (1972), Praetz and Wilson (1978), Blattberg and Gonedes (1974), McLeay (1986), and Taylor and Kingsman (1979). In the last of these papers a three-parameter Student (i.e., Pearson Type VII) distribution is fitted to daily changes in commodity prices. Recent applications of Student's distributions include the following:

1. [Eggers and Andersen (1989), Andersen, Lauritzen, and Thommesen (1990), Lauritzen, Thommesen, and Andersen (1990)] representing the distribution of the phase derivative (random frequency of a narrowband mobile channel) of air components in an urban environment.
2. [Mirza and Boyer (1992)] as part of noise models for depth map data and in the development of appropriate M-estimators.
3. [Angers (1992)] as (independent) prior distributions for expected values of multinormally distributed variables.
4. [Verdinelli and Wasserman (1991)] as an underlying model while discussing the Bayesian analysis of outlier models using the Gibbs sampling approach.

6 PEARSON TYPE MI DISTRIBUTIONS AND THEIR MODIFICATIONS

The general Type VII distribution has a probability density function that can be expressed in the form

$$p_X(x) = \frac{\Gamma(m)}{\sqrt{\pi}\,\Gamma(m-\frac{1}{2})}\frac{c^{2m-1}}{\left[c^2+(x-\xi)^2\right]^m}, \qquad m>0,\ c>0. \quad (28.59)$$

This depends on the three parameters m, c, and ξ. The t_ν distribution is obtained by putting m $= \frac{1}{2}(\nu+1)$, c $= \sqrt{\nu}$, and $\xi = 0$. Thus, if X has the distribution (28.59), then $\sqrt{2m-1}(X-\xi)/\sigma$ is distributed as t_{2m-1}. The shape of the curve represented by (28.59) is therefore the same as the shape of the curve corresponding to the t_{2m-1} distribution. This has been described in Section 2 and will not be discussed here. The present section will be devoted to discussion of the problem of estimation of the parameters m, c, and ξ, given observed values of n independent random variables X_1, $X_2, \cdots, X_n$, each having the distribution (28.59). This problem was discussed by Fisher (1922) as one of the earliest illustrations of the use of the method of maximum likelihood. Some further formulas were published by Sichel (1949) who also applied his method of frequency moments to the problem. Our discussion is based on these two papers.

The equations satisfied by the maximum likelihood estimators $\hat{m}$, $\hat{c}$, $\hat{\xi}$, can be written in the form

$$n^{-1} \sum_{j=1}^{n} \log\left[1 + \left(\frac{X_j - \hat{\xi}}{\hat{c}}\right)^2\right] = \psi(\hat{m}) - \psi\left(\hat{m} - \frac{1}{2}\right), \tag{28.60a}$$

$$n^{-1} \sum_{j=1}^{n} \log\left[1 + \left(\frac{X_j - \hat{\xi}}{\hat{c}}\right)^2\right]^{-1} = 1 - \frac{1}{2\hat{m}}, \tag{28.60b}$$

$$\sum_{j=1}^{n} \left(\frac{X_j - \hat{\xi}}{\hat{c}}\right)\left[1 + \left(\frac{X_j - \hat{\xi}}{\hat{c}}\right)^2\right]^{-1} = 0. \tag{28.60c}$$

For large values of n the standard formulas give the following approximations:

$$n \operatorname{var}(\hat{m}) \doteqdot \left[\psi^{(1)}\left(m - \frac{1}{2}\right) - \psi^{(1)}(m) - \frac{m+1}{m^2(2m-1)}\right]^{-1}, \tag{28.61a}$$

$$\begin{aligned} n \operatorname{var}(\hat{c}) &\doteqdot \frac{\left[\psi^{(1)}(m - \frac{1}{2}) - \psi^{(1)}(m)\right]c^2}{[(2m-1)/(m+1)]\left[\psi^{(1)}(m - \frac{1}{2}) - \psi^{(1)}(m)\right] - (1/m^2)} \\ &\doteqdot \frac{(m+1)c^2}{2m-1} \\ &\quad \times \frac{1}{1 - \left[(m+1)/\{m^2(2m-1)\}\right]\left[\psi^{(1)}(m - \frac{1}{2}) - \psi^{(1)}(m)\right]^{-1}}, \end{aligned} \tag{28.61b}$$

$$n \operatorname{var}(\hat{\xi}) \doteqdot \frac{(m+1)c^2}{m(2m-1)}, \tag{28.61c}$$

$$\operatorname{corr}(\hat{m}, \hat{c}) \doteqdot \frac{\sqrt{m+1}}{m\sqrt{(2m-1)\left[\psi^{(1)}(m - \frac{1}{2}) - \psi^{(1)}(m)\right]}}, \tag{28.61d}$$

$$\operatorname{corr}(\hat{m}, \hat{\xi}) \doteqdot \operatorname{corr}(\hat{c}, \hat{\xi}) \doteqdot 0. \tag{28.61e}$$

Taylor (1980) gives this analysis in terms of parameters $k = 2m - 1$ and $h = c^2(2m-1)^{-1}$. In these formulas $\psi(z) = (d/dz)\log\Gamma(z)$, $\psi^{(1)}(z) = (d/dz)\psi(z)$, and so on.

Formula (28.61c) also gives the approximate variance of the maximum likelihood estimator of ξ when the values of either or both of m and c are known. The asymptotic formulas for the variances of (and correlation between) the maximum likelihood estimators of m and c are the same whether the value of ξ is known or unknown.

If c is known, then

$$n\,\text{var}(\text{Maximum likelihood estimator of } m) \doteqdot \left[\psi^{(1)}\left(m - \tfrac{1}{2}\right) - \psi^{(1)}(m)\right]^{-1}. \tag{28.62}$$

The estimators of m and ξ are obtained by solving equations (28.60a) and (28.60b) with $\hat{c}$ replaced by c. If m is known, then

$$n\,\text{var}(\text{Maximum likelihood estimator of } c) = \frac{(m+1)c^2}{2m-1}. \tag{28.63}$$

Formulas (28.62) and (28.63) are applicable whether or not the value of ξ is known.

The parameters may also be estimated by equating sample and population values of first, second, and fourth moments. In terms of the population moments

$$m = \frac{\tfrac{1}{2}(5\beta_2 - 9)}{\beta_2 - 3},$$

$$c^2 = \frac{2\mu_2\beta_2}{\beta_2 - 3},$$

$$\xi = \mu_1'. \tag{28.64}$$

Denoting by $\tilde{m}$, $\tilde{c}$, $\tilde{\xi}$, the estimators obtained by replacing population values by sample values on the right-hand side of (28.641, we have the following approximate formulas (for n large):

$$n\,\text{var}(\tilde{m}) = \tfrac{2}{3}(m-1)(2m-5)(2m-3)^2(2m^2-5m+12) \times(2m-7)^{-1}(2m-9)^{-1}, \tag{28.65a}$$

$$n\,\text{var}(\tilde{c}) = \tfrac{1}{3}c^2(m-1)(2m-3)(8m^3-48m^2+108m-83) \times(2m-5)^{-1}(2m-7)^{-1}(2m-9)^{-1}, \tag{28.65b}$$

$$n\,\text{var}(\tilde{\xi}) = c^2(2m-3)^{-1}. \tag{28.65c}$$

Note that

Equation (28.65c) is exact and is also valid if either one or both of c and m have known values,

Equations (28.65a) and (28.65b) cannot be used unless m exceeds $4\frac{1}{2}$,

Equations (28.65a) and (28.65b) apply whether the value of ξ is known or not.

If either m or c is known, the other parameter (c or m, respectively) can be estimated by equating

$$\operatorname{var}(X) = c^2(2m-3)^{-1} \tag{28.66}$$

to the sample variance. For n large we have the following approximate formulas for the variance of estimators of m, c obtained from (28.64):

$$n(\text{Variance of estimator of } m) \doteqdot (2m-3)^2(m-1)(2m-5)^{-1}, \tag{28.67a}$$

$$n(\text{Variance of estimator of } c) = c^2(m-1)(2m-5)^{-1}. \tag{28.67b}$$

The parameter ξ may be estimated by the median. The variance of this estimator is approximately

$$\frac{c^2\pi}{4n}\left[\frac{\Gamma(m-\frac{1}{2})}{\Gamma(m)}\right]^2. \tag{28.68}$$

For m < 2.8 the median has a smaller asymptotic variance than the arithmetic mean. (The latter has infinite variance for $m \le 1.5$.) The ratio of asymptotic variance of mean to that of median decreases as m increases, tending to 0.637 (the value for normal distributions) as $m \to \infty$.

Fraser (1976) and Sprott (1980) use the distributional form

$$p_X(x;\theta,\sigma) \propto \left\{1 + \frac{(x-\theta)^2}{\lambda\sigma^2}\right\}^{-(\lambda+1)/2}, \tag{28.69}$$

where $\lambda(\ge 1)$ is assumed to be known. Sprott (1980) obtains an estimator of σ, regarding θ as a nuisance parameter, and using the method of maximum relative likelihood, maximizing

$$R_m(\theta_1;\mathbf{X}) = p_{\mathbf{X}}(\mathbf{X};\boldsymbol{\theta}^*)/p_{\mathbf{X}}(\mathbf{X};\hat{\boldsymbol{\theta}}),$$

where θ_1 is the parameter of interest and $\boldsymbol{\theta}^* = (\theta_1, \theta_2^*, \theta_3^*, \cdots, \theta_k^*)$ is the restricted maximum likelihood estimator of $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ is the unrestricted maximum likelihood estimator. Borwein and Gabor (1984) investigate the behavior of the MLE of parameter σ in this model.

This distribution appears in the following model of Bayesian inference: Let $X_1, X_2, \cdots, X_n$ be observations taken from $N(\theta, \sigma^2)$, where θ and σ^2 are unknown. The likelihood function of θ, after integrating over σ^2 with respect to the noninformative prior $\sigma^{-2}\,d\sigma^2$ (proper inverse gamma prior distributions for σ^2 could also be used here) is of form

$$f(\bar{X} - \theta) = \frac{\sqrt{n}\,K_{n-1}}{S}\left(1 + \frac{n(\theta - \bar{X})^2}{(n-1)S^2}\right)^{-n/2}, \tag{28.70}$$

where $\bar{X}$, S^2 are the usual sample mean and variance, and

$$K_j = \Gamma[(j+1)/2]\left[\sqrt{j\pi}\,\Gamma(j/2)\right]^{-1}.$$

[See Fan and Berger (1992) and Gambino and Guttman (1984), among others, for additional details.]

McDonald and Newey (1988) and Butler et al. (1990) studied distributions with pdfs of form

$$p_Y(y; \sigma, p, q) = \frac{p}{2\sigma q^{1/p} B(p^{-1}, q)(1 + |y|^p/(q\sigma^p))^{q+p^{-1}}}. \tag{28.71}$$

They term them ***GT distributions***. When $p = 2$ and $\sigma = \sqrt{2}\,\alpha$, (28.71) reduces to the pdf of a t-distribution with $2q$ degrees of freedom. In these distributions σ is a scale parameter while p and q control the shape of the density. Larger values of p and q are associated with "lighter" tails of the density. The hth moment of Y exists, provided $h < pq$, and is given by

$$E[Y^h] = \sigma^h q^{h/p}\frac{\Gamma((1+h)p^{-1})\Gamma(q - hp^{-1})}{\Gamma(p^{-1})\Gamma(q)}. \tag{28.72}$$

In an earlier paper McDonald and Butler (1987) point out that the log t distribution $[\mathrm{LT}(y; \mu, \sigma, q)]$ with the pdf

$$p_Y(y; \mu, \sigma, q) = \frac{\left\{B(q, \tfrac{1}{2})\,y\sqrt{2q\sigma^2}\right\}^{-1}}{\left\{1 + (\log y - \mu)^2/(2q\sigma^2)\right\}^{q+\frac{1}{2}}}, \qquad y > 0, \tag{28.73}$$

is a lognormal distribution mixed with an inverse gamma distribution. [See also Hogg and Klugman (1983)]. A similar result holds for the GT distributions, which are mixtures of Subbotin distributions with the pdf

$$pe^{-(|y|/\sigma)^p}/\{2\sigma\Gamma(p^{-1})\} \tag{28.74}$$

[See Eq. (24.83)′] with inverse generalized gamma distributions [Butler et al. (1990)]. (This result has applications in Bayesian inference.) The pdf of the inverse generalized gamma (IGG(y; $-a, \beta, q$)) is

$$p_Y(y; a, \beta, q) = \frac{\beta/y)^{ap+1} e^{-(\beta/y)^a}}{\beta\Gamma(p)} \tag{28.75}$$

(see Chapter 17). [Recall the relation between generalized gamma with the pdf

$$\frac{|a|(y/\beta)^{ap-1} e^{-(y/\beta)^a}}{\beta\Gamma(p)}, \qquad y \geq 0,$$

denoted by GG($y; a, \beta, p$), and inverse generalized gamma, denoted by IGG($y; a, \beta, p$):

$$\text{IGG}(y; a, \beta, p) \equiv \text{GG}(y; -a, \beta, p).]$$

Vaughan (1992) studies the Tiku-Suresh (1992) method of estimation for the (generalized) Student family:

$$p_X(x) = \frac{1}{\sigma k^{1/2} B\left(\frac{1}{2}, p - \frac{1}{2}\right)} \left\{1 + \frac{(x-\theta)^2}{k\sigma^2}\right\}^{-p} \qquad \text{[cf. (28.69)]} \tag{28.76}$$

for p, k, and $\sigma > 0$. The values of p and k satisfy $k = 2p - 3$ if $p \geq 2$ and $k = 1$ for $1 \leq p \leq 2$. (For $p = 1$ we have Cauchy distributions.)

The maximum likelihood equations, for p known,

$$\frac{\partial \log L}{\partial \theta} = \frac{2p}{k\sigma} \sum_{i=1} g(Z_i) = 0 \tag{28.77a}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \sigma} &= (2p-1)\frac{n}{\sigma} - 2p\sum_{i=1}^{n} \frac{k\sigma}{k\sigma^2 + (X_i - \theta)^2} \\ &= \frac{n}{-} + \frac{2p}{k\sigma} \sum_{i=1}^{n} Z_i g(Z_i) = 0,\end{aligned} \tag{28.77b}$$

where $Z_i = (X_i - \theta)/\sigma$ and $g(z) = z\{1 + z^2/k\}^{-1}$, have no explicit solutions except, of course, when $p = \infty$.

Tiku and Suresh (1992) and Vaughan (1992) provided a method for estimating θ and a by solving (28.77a) and (28.77b) iteratively. If no more than half of the sample values in a random sample of size n coincide, then, as

Vaughan (1992) shows, there is a unique pair of values $(\hat{\theta}, \hat{\sigma})$ satisfying (28.77a) and (28.77b), simultaneously and maximizing the likelihood function.

The modified maximum likelihood method of estimation suggested by Tiku and Suresh (1992) involves expression of these equations in terms of order statistics (X_i') with $Z_i' = (X_i' - \theta)/\sigma$:

$$\frac{\partial \log L}{\partial \theta} = \frac{2p}{k\sigma} \sum_{i=1}^{n} g(Z_i') = 0 \qquad (28.78a)$$

and

$$\frac{\partial \log L}{\partial \sigma} = \frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} Z_i' g\{Z_i'\} = 0, \qquad (28.78b)$$

and then linearizing $\partial \log L/\partial\theta$ using the Taylor series approximation

$$g(Z_i') \doteqdot g(\zeta_i) + [Z_i' - \zeta_i]\left\{\frac{d}{dz} g(z)\right\}_{z=\zeta_i} = \alpha_i + \beta_i Z_i', \qquad (28.79)$$

where

$$\zeta_i = E[Z_i'],$$

$$\alpha_i = \frac{(2/k)\zeta_i^3}{[1 + (1/k)\zeta_i^2]^2},$$

$$\beta_i = \frac{1 - (1/k)\zeta_i^2}{[1 + (1/k)\zeta_i^2]^2}, \qquad i = 1, \dots, n.$$

Note that since the distribution is symmetric, $\sum_{i=1}^{n} \alpha_i = 0$. Incorporating (28.79) into equations (28.78), they arrive at the MML estimators

$$\hat{\theta}^* = \left\{\sum_{i=1}^{n} \beta_i X_i'\right\}\left(\sum_{i=1}^{n} \beta_i\right)^{-1}, \qquad (28.80a)$$

$$\hat{\sigma}^* = \frac{\{B + \sqrt{B^2 + 4nC}\}}{2\sqrt{n(n-1)}}, \qquad (28.80b)$$

where

$$B = \frac{2p}{k} \sum_{i=1}^{n} \alpha_i X_i',$$

$$C = \frac{2p}{k} \sum_{i=1}^{n} \beta_i \{X_i' - \hat{e}^*\}^2.$$

Note that in $\hat{\sigma}^*$ the divisor n is replaced by $\sqrt{n(n-1)}$ for bias reduction; if $g(z)$ is linear in z, the estimators obtained are the ML estimators. The estimator of σ in (28.80b) may, however, take on negative values in some instances. A detailed analysis of these estimators is available in Vaughan (1992), who extends the results to censored situations.

7 OTHER RELATED DISTRIBUTIONS

We have already noted that the distribution of t_ν^2 is identical with that of F with 1, v degrees of freedom. On account of the symmetry, about zero, of the t, distribution,

$$\Pr[F_{1,\nu} < K] = \Pr[t_\nu^2 < K] = \Pr[|t_\nu| < \sqrt{K}], \tag{28.81}$$

and so

$$\sqrt{F_{1,\nu,\alpha}} = t_{\nu,(1+\alpha)/2}.$$

Other relations between the t and F distributions have been described in Chapter 27. Psarakis and Panaretos (1990) introduced the distribution of the random variable $W = |t_\nu|$, termed the ***folded t variable***. The expected value and variance are

$$E[W] = 2\sqrt{\frac{\nu}{\pi}}\,\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)(\nu-1)}, \qquad \nu > 1 \tag{28.82a}$$

and

$$\operatorname{var}(W) = \frac{\nu}{\nu-2} - \frac{4\nu}{\pi(\nu-1)^2}\left\{\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\right\}^2, \qquad \nu > 2. \tag{28.82b}$$

They show that the folded standard normal distributions with the pdf

$$\frac{1}{\sqrt{2\pi}\,\sigma}\{e^{-(x-\mu)^2/(2\sigma^2)} + e^{-(x+\mu)^2/(2\sigma^2)}\}, \qquad x > 0$$

(see Chapter 13, Section 10.3) is the limiting form of the folded t-distribution as $\nu \to \infty$.

The folded t-distribution is related to the chi-distribution by the relation

$$W = \frac{\nu^{-1/2}X}{Y}, \tag{28.83}$$

where X and Y are independent chi variables with 1, ν degrees of freedom, respectively.

W with 1 degree of freedom is the standard half-Cauchy. If X has an $F_{\nu,\nu}$ distribution, then (as noted in Section 1), $\frac{1}{2}\sqrt{\nu}(X^{1/2} - X^{-1/2})$ has a t, distribution, so $\frac{1}{2}\sqrt{\nu}|X^{1/2} - X^{-1/2}|$ has a folded t-distribution with ν degrees of freedom. Psarakis and Panaretos (1990) tabulate values of the cdf $F_W(w)$, for $w = 0.0(0.1)6.0$ and $\nu = 1(1)5$, 10, and 20.

There are a number of pseudo-t-distributions obtained by replacing $\chi_\nu/\sqrt{\nu}$ in the denominator of

$$t_\nu = \frac{U}{\chi_\nu/\sqrt{\nu}}$$

by variables with other distributions (still independent of U). These correspond to replacing S' in $\sqrt{n}(\bar{X} - \xi)/S'$ by other sample measures of dispersion, in particular by the range $[\max(X_1,\ldots,X_n) - \min(X_1,\ldots,X_n)]$ or the mean deviation $n^{-1}\sum_{j=1}^n|X_j - \bar{X}|$. These latter have the common feature that they are distributed as σT, where T is the variable corresponding to the case $\sigma = 1$. Hence the ratio is distributed as U/T and so does not depend on the value of σ [e.g., see Pillai (1951)].

The use of statistics of this kind has been described in Chapter 13. If the distribution of the denominator is approximated by a $c\chi_\nu/\sqrt{\nu}$ distribution, the ratio is approximated by a $c^{-1}t_\nu$ distribution. Usually ν is fractional, and it is necessary to interpolate if standard tables are used (in which ν is usually given for integer values only, the tables of Mardia and Zemroch (1978) are a notable exception). Alternatively, approximate formulas may be used.

Birnbaum and Vincze (1970) studied the distribution of

$$T^* = \frac{X'_{m+1} - \mu}{X'_{m+1+r} - X'_{m+1-r}} \tag{28.84}$$

from a general distribution with a continuous cdf. Here

$$\mu = \inf\{x : F(x) = \tfrac{1}{2}\}$$

is the population median, r is an integer $1 \le r \le m$ and the sample size is $2m + 1$; X'_{m+1} is the sample median.

The statistic T* is similar to the t-statistic. The denominator is an estimator (sample interquartile range) of a scale parameter (population interquartile range). This statistic is invariant under linear transformations and needs only three order statistics for calculation. Under mild regularity conditions [continuity of $p'_X(x)$ and $p'_X(\mu) = 0$] the limiting distribution of T* (for a fixed r) is

$$\lim_{m\to\infty} \Pr\left[\sqrt{\frac{2}{m}}\,T^* \le s\right] = \frac{1}{(2r-1)!}\int_0^\infty \phi(zs) z^{2r-1} e^{-z}\, dz, \tag{28.85}$$

where $\phi(\cdot)$ is the unit normal pdf. Tables (based on simulations) of $\Pr[T^* > t]$, when X is normally distributed, for $m = 1(1)10$, $r = 1(1)m$ and $t = 0.0(0.1)5$ were provided by Tague (1969).

The distribution of $(X_j - \bar{X})/S'$, where X, is randomly chosen from X,, $X_2, \cdot, \cdot, X_n$ was found by Thompson (1935) to be related to the t-distribution. In fact, setting $E[X_j] = 0$ and $\text{var}(X_j) = 1$ (which does not affect the distribution), we obtain

$$\sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2 = X_1^2 + \frac{1}{2}(X_2 - X_1)^2 + \frac{2}{3}\left(X_3 - \frac{1}{2}(X_1 + X_2)\right)^2 + \cdots$$
$$+ \frac{n-1}{n}\left[X_n - \frac{1}{n-1}(X_1 + \cdots + X_{n-1})\right]^2$$
$$= Y + \frac{n}{n-1}\left(X_n - \bar{\bar{X}}\right)^2,$$

with Y and $(X_n - \bar{X})^2$ mutually independent, and Y distributed as χ^2_{n-2}. Hence $(X_n - \bar{X})/S'$ is distributed as $[(n-1)/\sqrt{n}]U(\chi^2_{n-2} + U^2)^{-1/2}$, with U (standard normal) and χ^2_{n-2} mutually independent; that is, as beta $(\frac{1}{2}, \frac{n}{2} - 1)$ over the range $[-(n-1)/\sqrt{n}, (n-1)/\sqrt{n}]$. Equivalently $(X_n - \bar{X})/S'$ is distributed as

$$\frac{n-1}{\sqrt{n}}\sqrt{n-2}\, t_{n-2}\{1 + (n-2)t^2_{n-2}\}^{-1/2}. \tag{28.86}$$

Smith (1992) provides an elegant derivation of the distribution of the t-statistic (28.1) when U and χ^2_ν are not necessarily independent. With $\chi^2_\nu = \sum_{i=1}^{\nu} Y_i^2$, we take $Z = (U, Y)$ to have a multinormal distribution with expected value **0** and variance-covariance matrix

$$\underset{(\nu+1)\times(\nu+1)}{\Sigma} = \begin{pmatrix} 1 & \mathbf{V}' \\ \mathbf{V} & \mathbf{I}_\nu \end{pmatrix}$$

V is a $\nu \times 1$ vector and $\mathbf{I}_\nu$ denotes the identity **matrix** of order v. If V = 0, then U and χ^2_ν are independent and $t \sim$ t,. If V $\neq$ **0**, defining $w = \mathbf{V}'\mathbf{V} \geq 0$ (and since $|\Sigma| = 1 - w \geq 0$, $w \leq 1$), Smith (1992) derives the density function

$$p_T(t) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\{\pi\nu(1-w)\}^{1/2}}\left\{1 + \frac{t^2}{\nu(1-w)}\right\}^{-(\nu+1)/2}$$
$$\times \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left\{\left\{\frac{(\frac{1}{2})_{i+j}((\nu+1)/2)_{2i+j}}{(\frac{1}{2})_i(\frac{1}{2}\nu)_{i+j}\, i!j!}(-1)^j\left(\frac{w}{1-w}\right)^{i+j}\right.\right.$$
$$\times \left\{\frac{t^2}{\nu(1-w)}\right\}^i\left\{1 + \frac{t^2}{\nu(1-w)}\right\}^{-(2i+j)}, \qquad -\infty < t < \infty,$$
$$\tag{28.87}$$

where $(a)_i = a(a+1)\cdots(a+i-1)$.

The discrepancy between exact and nominal probabilities $P[T > t_\alpha]$ is small; the worst relative errors occur for small ν, large w, and small a. For 10% and smaller values of $1 - \alpha$ the exact tail probabilities under dependence are smaller than the nominal ones. Using Student's values thus gives rise to conservative inference.

The distribution of $\sqrt{n}(\bar{X} - \xi)/S'$ when $X_1, X_2, \cdots, X_n$ have identical, but not normal, distributions has been studied for the following cases:

Rectangular. By Hotelling and Frankel (1938), Perlo (1933), Rider (1929, 1931), Rietz (1939), Siddiqui (1964), Watanabe (1960–1966), and Ali (1975, 1976).

Exponential. By Geary (1936) and Hoq, Ali and Templeton (1978).

Cauchy and squared hyperbolic secant. By Bradley (1952) and Hotelling (1961).

Edgeworth series. By Bartlett (1935), Ghurye (1949), Gayen (1949, 1952), Tiku (1963), and Zackrisson (1959).

Compound (*mixture*) normal. By Hyrenius (1950) and Quensel (1943).

Various distributions. (Rectangular, Laplace, χ^2, Beta). By Watanabe (1960–1966), Sansing (1976), and Sansing and Owen (1974).

Stable distributions. By Logan et al. (1973).

Laderman (1939) gave a general formula for the distribution of t, for any continuous parent population, for sample size two. In a number of cases exact distributions have been obtained only for small sample sizes [ν usually up to 3 or 4]. For the parent rectangular distribution Siddiqui has given bounds for the probability integral of t_ν and obtained numerical values for $\nu \leq 6$. Hotelling (1961), by a geometrical method, has shown that as $t \to \infty$, the ratio of $\Pr[t_\nu > t]$ to its normal theory value tends to

$$\frac{\{\pi(\nu + 1)\}^{(\nu+1)/2}}{2^{\nu+1}\Gamma(\frac{1}{2}(\nu + 3))} \quad \text{for a parent uniform distribution,} \qquad (28.88)$$

and that for a parent Cauchy distribution [pdf $\pi^{-1}(1 + x^2)^{-1}$], the limiting ratio is

$$\frac{[(\nu + 1)/\pi]^{(\nu+1)/2}\Gamma(\frac{1}{2}(\nu + 1))}{\Gamma(\frac{1}{2}\nu)}. \qquad (28.89)$$

He also obtained similar results for Pearson Type II, exponential and double exponential (Laplace) distributions.

Of most general interest are the results for Edgeworth series. The results are given for all n and indicate the variation from the Student's t-distribution that one might expect to be associated with given nonnormal values of the moment ratios. Ghurye (1949) considered a series only up to the term in $\sqrt{\beta_1}$. Gayen (1949, 1952) included further terms. The results of Tiku (1963) appear to correspond to inclusion of still further terms in the expansion for the distribution in the parent population. Gayen obtained

$$\Pr[t_\nu > t] = (\text{Value for normal population})$$
$$-\sqrt{\beta_1}\,P_{\sqrt{\beta_1}}(t) - (\beta_2 - 3)P_{\beta_2}(t) + \beta_1 P_{\beta_1}(t) \quad (28.90)$$

and gives a table of the functions $P(\cdot)$. We reproduce part of his table as Table 28.16 [See also Chung (1946).] We should also note a formula obtained by Bradley (1952, p. 21) for a Cauchy parent population (for which we cannot use moment ratios as an index of nonnormality). It expresses the value of $\Pr[t_\nu \le t]$, for positive t, as a series in t^{-2} (up to and including the t^{-6} term).*

Ratcliffe (1968) presented results of an empirical investigation of the distribution of t for five markedly non-normal parent distributions (including rectangular, exponential, Laplace, gamma, and a U-shaped distribution). He studied, in particular, the reduction in the effect of non-normality with increasing sample size. He concluded that a sample size of 80 or more should eliminate effects of non-normality (mainly skewness) for most practical purposes. For symmetrical population distributions the necessary sample size is much less.

Efron (1968) has given a theoretical discussion of the distribution of t under general symmetry conditions.

Hoq, Ali, and Templeton (1978) obtained the distribution of $T = \sqrt{n}(\bar{X} - \theta)/S$ when the parent distribution is exponential. Closed expression are available for $n = 2$, 3 and 4 for values of the argument $t \ge ((n-1)(n-2)/2)^{1/2}$. The upper tails of the distribution are heavier when the parent distribution is exponential than when it was normal, however the half-normal parent distribution $[p_X(x) = (2/\pi)^{1/2}\exp(-x^2/2),\ x > 0]$ gives a heavier/upper tail than the exponential for all $t \ge n - 1$.

Although further progress has been made since Hoq, Ali, and Templeton (1978), explicit expressions for the distribution of T for all n in the case of exponential parent distribution are not as yet available.

Sansing and Owen (1974) consider the case of a parent standard double exponential (Laplace) distribution (Chapter 24) with pdf

$$p_X(x) = \tfrac{1}{2}\exp(-|x|).$$

They show that the pdf of the t-statistic (see Section 1) based on a random

*In all of the above cases the X's are assumed to be mutually independent. Weibull (1958) considered cases where they are normal but serially correlated.

Table 28.16 Corrective factors for distribution of t_ν in nonnormal (Edgeworth) populations

t	Normal Theory	$P_{\sqrt{\beta_1}}(t)$	$P_{\beta_2}(t)$	$P_{\beta_1}(t)$
		$\nu = 1$		
0.0	0.5000	0.0470	0.0000	0.0000
0.5	0.3524	0.0589	−0.0064	−0.0066
1.0	0.2500	0.0665	0.0000	0.0044
1.5	0.1872	0.0622	0.0047	0.0147
2.0	0.1476	0.0547	0.0064	0.0188
2.5	0.1211	0.0476	0.0066	0.0195
3.0	0.1024	0.0416	0.0064	0.0188
3.5	0.0886	0.0368	0.0059	0.0176
4.0	0.0780	0.0329	0.0055	0.0163
		$\nu = 2$		
0.0	0.5000	0.0384	0.0000	0.0000
0.5	0.3333	0.0495	−0.0069	−0.0066
1.0	0.2113	0.0597	−0.0027	0.0009
1.5	0.1362	0.0563	0.0025	0.0118
2.0	0.0918	0.0469	0.0047	0.0172
2.5	0.0648	0.0375	0.0051	0.0179
3.0	0.0477	0.0298	0.0047	0.0165
3.5	0.0364	0.0239	0.0041	0.0145
4.0	0.0286	0.0194	0.0035	0.0125
		$\nu = 3$		
0.0	0.5000	0.0332	0.0000	0.0000
0.5	0.3257	0.0431	−0.0062	−0.0056
1.0	0.1955	0.0540	−0.0034	−0.0002
1.5	0.1153	0.0513	0.0013	0.0098
2.0	0.0697	0.0413	0.0035	0.0152
2.5	0.0439	0.0310	0.0039	0.0157
3.0	0.0288	0.0229	0.0034	0.0139
3.5	0.0197	0.0169	0.0028	0.0114
4.0	0.0137	0.0126	0.0023	0.0093
		$\nu = 4$		
0.0	0.5000	0.0297	0.0000	0.0000
0.5	0.3217	0.0387	−0.0055	−0.0047
1.0	0.1870	0.0495	−0.0036	−0.0005
1.5	0.1040	0.0473	0.0006	0.0084
2.0	0.0581	0.0372	0.0028	0.0135
2.5	0.0334	0.0266	0.0031	0.0139
3.0	0.0200	0.0184	0.0027	0.0119
3.5	0.0124	0.0127	0.0021	0.0095
4.0	0.0081	0.0088	0.0016	0.0072

Table 28.16 *(Continued)*

t	Normal *Theory*	$P_{\sqrt{\beta_1}}(t)$	$P_{\beta_2}(t)$	$P_{\beta_1}(t)$
		$\nu = 5$		
0.0	0.5000	0.0271	0.0000	0.0000
0.5	0.3192	0.0355	−0.0049	−0.0041
1.0	0.1816	0.0397	−0.0035	−0.0005
1.5	0.0970	0.0440	0.0002	0.0074
2.0	0.0510	0.0340	0.0022	0.0122
2.5	0.0272	0.0234	0.0025	0.0125
3.0	0.0150	0.0154	0.0021	0.0104
3.5	0.0086	0.0099	0.0016	0.0079
4.0	0.0052	0.0065	0.0011	0.0057
		$\nu = 6$		
0.0	0.5000	0.0251	0.0000	0.0000
0.5	0.3174	0.0329	−0.0044	−0.0035
1.0	0.1780	0.0430	−0.0033	−0.0005
1.5	0.0921	0.0413	0.0000	0.0066
2.0	0.0462	0.0315	0.0019	0.0111
2.5	0.0233	0.0210	0.0021	0.0113
3.0	0.0120	0.0132	0.0017	0.0092
3.5	0.0064	0.0081	0.0012	0.0067
4.0	0.0036	0.0050	0.0008	0.0047
		$\nu = 8$		
0.0	0.5000	0.0222	0.0000	0.0000
0.5	0.3153	0.0291	−0.0037	−0.0028
1.0	0.1733	0.0384	−0.0030	−0.0005
1.5	0.0860	0.0371	−0.0002	−0.0055
2.0	0.0403	0.0277	0.0014	0.0094
2.5	0.0185	0.0177	0.0016	0.0095
3.0	0.0085	0.0103	0.0013	0.0074
3.5	0.0040	0.0058	0.0008	0.0051
4.0	0.0020	0.0032	0.0005	0.0033
		$\nu = 12$		
0.0	0.5000	0.0184	0.0000	0.0000
0.5	0.3131	0.0243	−0.0027	−0.0019
1.0	0.1685	0.0325	−0.0023	−0.0002
1.5	0.0797	0.0315	−0.0004	0.0042
2.0	0.0343	0.0230	0.0009	0.0073
2.5	0.0140	0.0137	0.0011	0.0072
3.0	0.0055	0.0072	0.0008	0.0053
3.5	0.0022	0.0036	0.0005	0.0033
4.0	0.0009	0.0017	0.0003	0.0019

Table 28.16 *(Continued)*

t	Normal Theory	$P_{\sqrt{\beta_1}}(t)$	$P_{\beta_2}(t)$	$P_{\beta_1}(t)$
		$\nu = 24$		
0.0	0.5000	0.0133	0.0000	0.0000
0.5	0.3101	0.0176	−0.0015	−0.0010
1.0	0.1636	0.0238	−0.0014	−0.0001
1.5	0.0733	0.0232	−0.0003	0.0025
2.0	0.0285	0.0164	0.0004	0.0043
2.5	0.0098	0.0090	0.0005	0.0041
3.0	0.0031	0.0041	0.0003	0.0028
3.5	0.0009	0.0016	0.0002	0.0015
4.0	0.0003	0.0006	0.0001	0.0007
		$\nu = \infty$		
0.0	0.5000			
0.5	0.3085			
1.0	0.1587			
1.5	0.0668			
2.0	0.0228			
2.5	0.0062			
3.0	0.0013			
3.5	0.0002			
4.0	0.0000			

sample of size n is of form

$$p_T(t) \propto \left(\sum_{j=1}^{n} |t + b_{jn}| \right)^{-n}.$$

The pdf is bounded by

$$c_n \left(\sum_{j=1}^{n} |t + b_{jn}| \right)^{-n} < p_T(t) < c_n(n|t|)^{-n}, \tag{28.91}$$

where

$$c_n = \frac{\pi^{(n-1)/2}\Gamma(n)}{2^{n-1}\sqrt{n}\,\Gamma(\frac{1}{2}(n-1))}$$

and

$$b_{jn} = \sqrt{n-1}\left\{\sqrt{j(n-j)} - \sqrt{(j-1)(n-j+1)}\right\}.$$

In the extreme tails, for $|t| > n - 1$, Sansing *(1976)* showed that

$$p_T(t) = c_n\left(\frac{n-1}{n}\right)^{(n-1)/2} t^{-(n-1)}. \tag{28.92}$$

We shall now discuss the distribution and approximations to the distribution of a difference of two 'Student' variables which is of great importance in connection with the Behrens–Fisher problem.

Ghosh (1975) has shown that if $\nu_1 = \nu_2 = \nu$, and $Z = T_1 - T_2$, then

$$\Pr[0 < Z < z] = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{2^\nu\sqrt{\pi}\{\Gamma(\frac{1}{2}\nu)\}^2}\sum_{i=0}^{\infty}\frac{\Gamma(i+\frac{1}{2})\Gamma(\nu+i+\frac{1}{2})}{i!\Gamma(\frac{1}{2}\nu+i+1)} \times\int_0^{\frac{1}{4}z^2\nu^{-1}}\frac{y^{i-(1/2)}}{(1+y)^{(1/2)+\nu+i}}\,dy. \tag{28.93}$$

He provides a table of values of $\Pr[0 < Z < z]$ for

$$z = 0.0(0.5)10.0; \nu = 1, 2(2)10$$

and for

$$z = 0.0(0.5)7.5; \nu = 1(1)20.$$

Guenther *(1975)* noted that (28.93) can be expressed as

$$\Pr[0 < Z < z] = \sum_{i=0}^{\infty} C_i \Pr\left[F_{2i+1,2\nu} \le \frac{1}{2}(2i+1)^{-1}z^2\right], \tag{28.93$'$}$$

where

$$C_i = \frac{\left\{\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{i+1}{2}\right)\right\}^2}{i!2\pi\Gamma(\frac{1}{2}\nu)\Gamma(\frac{1}{2}\nu+i+1)}.$$

The coefficients C_i can be calculated recursively, **from** the relatinn

$$C_i = \frac{(2i-1)^2}{2i(\nu+2i)}C_{i-1}\,(i \ge 1).$$

This form is suitable for calcuiation with a pocket calculator.

Chaubey and Mudholkar (1982) observed that in general, if $F_1(x)$ and $F_2(x)$ are two distribution functions symmetric around zero with unit second moment then the mixture

$$F(x) = \lambda F_1(x) + (1-\lambda)F_2(x) \tag{28.94}$$

has the same properties. Now

$$D = (T_1 - T_2)\sqrt{\frac{\nu_1}{\nu_1 - 2} + \frac{\nu_2}{\nu_2 - 2}} = (T_1 - T_2)\sqrt{Q}$$

is distributed symmetrically around zero with unit variance. To approximate the distribution of $Z = T_1 - T_2$ the authors chose $F_1(x)$ as $\Phi(x)$ and $F_2(x)$ as the cdf of a standardized symmetric distribution with higher kurtosis—specifically, the distribution of $\sqrt{\nu/(\nu-2)}\,t_\nu$. Equating the 4-th and 6-th cumulants of the mixture to those of Z, the values of ν and λ are

$$\nu = 6 + R/S; \lambda = 1 - (\nu - 4)R, \tag{28.95}$$

where

$$R = (A_1 + A_2)/Q^2$$

and

$$S = (B_1 + B_2)/Q^3,$$

with

$$A_j = (\nu_j - 2)^{-2}(\nu_j - 4)^{-1}\nu_j^2;$$

$$B_j = (\nu_j - 2)^{-3}(\nu_j - 4)^{-1}(\nu_j - 6)\nu_j^3 \qquad (j = 1,2).$$

We have

$$\Pr[D \le d] = \Pr\left[D^* < d/\sqrt{Q}\right] = \lambda F_1(d^*) + (1 - \lambda)F_2(d^*),$$

where $d^* = d/\sqrt{Q}$. If $F_1(x)$ and $F_2(x)$ are not too dissimilar it is proposed to approximate the $100\alpha\%$ point, D_α of D by

$$D_\alpha \doteqdot \{\lambda X_1(\alpha) + (1 - \lambda)X_2(\alpha)\}\sqrt{Q},$$

where $X_1(\alpha) = F_1^{-1}(\alpha)$ and $X_2(\alpha) = F_2^{-1}(\alpha)$ denote the upper $100\alpha\%$ points of $F_1(x)\,(\equiv \Phi(x))$ and $F_2(x)$ respectively. Explicitly

$$D_\alpha \doteqdot \left\{\lambda U_\alpha + (1 - \lambda)t_{\nu,\alpha}\sqrt{\frac{\nu - 2}{\nu}}\right\}\sqrt{\left\{\frac{\nu_1}{\nu_1 - 2} + \frac{\nu_2}{\nu_2 - 2}\right\}}, \tag{28.96}$$

where $U_\alpha = \Phi^{-1}(\alpha)$, and ν, λ are given by (28.95). Using Wallace's (1959) transform (28.26), t_ν can be approximated by

$$t_\nu^* \doteq \left[\nu\left\{\exp\left(\frac{1}{\nu}\left(\frac{8\nu+3}{8\nu+1}U\right)^2 - 1\right)\right\}\right]^{1/2}, \qquad (28.97)$$

where U is a unit normal variable.

Calculations suggest that little accuracy is lost by using t_ν^* in place of t_ν. Thus

$$\Pr[D \le d] = \lambda\Phi(d^*) + (1-\lambda)\Phi(d^{**}), \qquad (28.98)$$

where

$$d^* = d/[\{\nu_1/(\nu_1 - 2)\} + \{\nu_2/(\nu_2 - 2)\}]^{1/2},$$

$$d^{**} = \{(8\nu + 1)/(8\nu + 3)\}[\nu \log\{1 + (d^{*2}/(\nu - 2))\}].$$

Comparisons show that the Chaubey-Mudholkar approximation is far superior to Patil's (1965) approximation (e.g., for the case $\nu_1 = \nu_2 = 10$ and $\alpha = 0.99$, the exact value of d_α is 3.807, the mixture approach yields 3.815 while Patil's approximation gives 3.940; for $\nu_1 = \nu_2 = 20$, $\alpha = 0.99$ the values are 3.526, 3.527 and 3.585 respectively.)

Ghosh (1975), having derived the exact distribution of D, suggests the approximations:

$$\Pr[D \le d] = \Phi(d/\sqrt{2}) - \frac{d\phi(d/\sqrt{2})}{32\sqrt{2}} \times\left[\frac{1}{\nu_2}Q_1(d) + \frac{1}{\nu_2^2}Q_2(d) + \frac{1}{\nu_2^3}Q_3(d) + O\left(\frac{1}{\nu_2^4}\right)\right], \qquad (28.99)$$

where

$$Q_1(d) = (1+\theta)(d^2 + 10),$$

$$Q_2(d) = \frac{1+e^2}{384}(3d^6 + 98d^4 + 620d^2 + 168) + \frac{\theta}{64}(d^6 - 10d^4 + 36d^2 - 456),$$

$$Q_3(d) = \frac{1+\theta^3}{24576}(d^{10} + 66d^8 + 1016d^6 - 1296d^4 - 65328d^2 - 141408) + \frac{\theta(1+\theta)}{24576}(3d^{10} - 58d^8 - 280d^6 + 6864d^4 - 70032d^2 + 122592)$$

and

$$\theta = \nu_2/\nu_1.$$

Inverting (28.99) we get

$$d_\alpha = U_\alpha\sqrt{2}\left[1 + \frac{1}{\nu_2}R_1(U_\alpha) + \frac{1}{\nu_2^2}R_2(U_\alpha) + \frac{1}{\nu_2^3}R_3(U_\alpha) + O\left(\frac{1}{\nu_2^4}\right)\right], \tag{28.100}$$

where U_α satisfies $\Phi(U_\alpha) = \alpha$,

$$R_1(t) = \frac{1+\theta}{16}(t^2+5),$$

$$R_2(t) = \frac{1+\theta^2}{1536}(37t^4 + 200t^2 + 171) - \frac{\theta}{256}(9t^4 - 24t^2 + 7),$$

$$R_3(t) = \frac{1+\theta^3}{8192}(81t^6 + 349t^4 - 293t^2 - 1153)$$
$$- \frac{\theta(1+\theta)}{24576}(231t^6 - 773t^4 - 499t^2 - 2871).$$

Chaubey and Mudholkar's (1982) approximation is as accurate as **Ghosh's (1975)** approximation, though considerably simpler, especially when the Wallace transform is used; see Table 28.17.

The distributions of linear functions of the form $a_1T_1 - a_2T_2$, with a_1, a_2 both positive, and T_1, T_2 independent variables distributed as t_{ν_1}, t_{ν_2} respectively have been studied in connection with tests of the hypothesis that expected values of two normal distributions are equal, when it cannot be assumed that the standard deviations of the two distributions are equal. Tests of this kind were proposed by Behrens (1929) and studied later by Fisher (1935, 1941); they are said to relate to the "Behrens-Fisher problem."

Suppose that $X_{j1}, X_{j2}, \cdots, X_{jn_j}$ are independent random variables, each normally distributed with expected value ξ_j and standard deviation σ_j ($j = 1, 2$). Then for $j = 1, 2$, $\sqrt{n_j}(\bar{X}_j - \xi_j)/S'_j$ is distributed as t_{n_j-1} (using a notation similar to that employed in Section 1). According to one form of the *fiducial* argument (Chapter 13, Section 8), from

$$\frac{\sqrt{n_j}\left(\bar{X}_j - \xi_j\right)}{S'_j} \quad \text{distributed as} \quad t_{n_j-1},$$

Table 28.17 Comparison of approximations for D_α ($\nu_1 = \nu_2$)

		Approximations for D_α				
$\nu_1 = \nu_2$	a	Exact	[1]	[2]	[3]	[4]
10	0.55	0.1892	0.1892	0.1885	0.1798	0.1892
	0.75	1.022	1.022	1.020	0.9776	1.022
	0.90	1.978	1.979	1.979	1.926	1.978
	0.95	2.581	2.582	2.586	2.554	2.581
	0.99	3.807	3.815	3.820	3.940	3.814
15	0.55	0.1853	0.1853	0.1852	0.1796	0.1853
	0.75	0.9988	0.9987	0.9984	0.9719	0.9988
	0.90	1.919	1.919	1.919	1.890	1.919
	0.95	2.489	2.489	2.490	2.476	2.489
	0.99	3.616	3.616	3.617	3.697	3.616
20	0.55	0.1834	0.1834	0.1833	0.1793	0.1834
	0.75	0.9872	0.9872	0.9871	0.9683	0.9873
	0.90	1.891	1.891	1.891	1.871	1.891
	0.95	2.445	2.446	2.446	2.438	2.446
	0.99	3.526	3.527	3.527	3.585	3.526

Note: [1] Mixture (Chaubey-Mudholkar) (28.96); [2] same (with Wallace transform) (28.98); [3] Patil (1965); 141 Ghosh (1975) (28.99).

we can deduce

$$\xi_j \text{ (fiducially) distributed as } \bar{X}_j - \left(\frac{S'_j}{\sqrt{n_j}}\right) t_{n_j-1}$$

($\bar{X}_j$ and S'_j are regarded as fixed). Formally we can say that

$$\xi_2 - \xi_1 \quad \text{is (fiducially) distributed as}$$

$$\bar{X}_2 - \bar{X}_1 + \left[\left(\frac{S'_1}{\sqrt{n_1}}\right) t_{n_1-1} - \left(\frac{S'_2}{\sqrt{n_2}}\right) t_{n_2-1}\right]. \qquad (28.101)$$

The Behrens-Fisher test procedure rejects the hypothesis $\xi_1 = \xi_2$ at the level of significance a if (according to the fiducial distribution)

$$\Pr[\xi_2 - \xi_1 < 0] < \tfrac{1}{2}\alpha \quad \text{or} \quad \Pr[\xi_2 - \xi_1 > 0] < \tfrac{1}{2}\alpha.$$

Note that $\bar{X}_1$, $\bar{X}_2$, S'_1, and S'_2 are regarded as constants in the fiducial distribution. It is found convenient to use the quantity

$$\frac{X_1 - X_2}{\sqrt{(S_1'^2/n_1) + (S_2'^2/n_2)}} + t_{\nu_1}\cos\theta - t_{\nu_2}\sin\theta, \qquad (28.102)$$

with

$$\nu_i = n_i - 1, i = 1, 2, \cos\theta = \frac{S_1'}{\sqrt{n_1}}\left[\frac{S_1'^2}{n_1} + \frac{S_2'^2}{n_2}\right]^{-1/2}.$$

The distribution of

$$D_\theta = T_1\cos\theta - T_2\sin\theta$$

is needed to calculate fiducial limits for $(\xi_1 - \xi_2)$. Tables of percentile points of the distribution of D_θ are available as follows:

Sukhatme (1938) gave values of $D_{\theta,0.975}$ to three decimal places for

$$\nu_1, \nu_2 = 6, 8, 12, 24, \infty \quad \text{and} \quad \theta = 0°(15°)90°.$$

Fisher and Yates (1966) included these values together with values of $D_{\theta,0.995}$ (also to three decimal places) and also values with $\alpha = 0.95$, 0.975, 0.99, 0.995, 0.9975, and 0.999, for $\nu_1 = 10$, 12, 15, 20, 30, 60, and $\nu_2 = \infty$.

Weir (1966) gave values of $D_{\theta,0.999}$ to three decimal places for

$$\nu_1, \nu_2 = 6, 8, 12, 24, \infty$$
$$\theta = 0°(15°)90°.$$

Isaacs et al. (1974) provided values (to two decimal places) of $D_{\theta,\alpha}$ for $\alpha = 0.75$, 0.90, 0.95, 0.975, and 0.999, and for the same values of ν_1, ν_2, and θ as Weir (1966). These tables are reproduced in Novick and Jackson (1974).

Ruben (1960) showed that D_θ has the same distribution as the ratio of a $t_{\nu_1+\nu_2}$ variable to an independent variable

$$\phi(X) = \left[\frac{\nu_1 + \nu_2}{\nu_1 X^{-1}\cos\theta - \nu_2(1 - X)^{-1}\sin\theta}\right]^{1/2},$$

where X has a standard beta distribution (Chapter 25) with parameters $\frac{1}{2}\nu_1, \frac{1}{2}\nu_2$.

Patil (1965) suggested that the distribution of D be approximated by that of ct_f with

$$f = 4 + \left[\frac{\nu_1 \cos^2\theta}{\nu_1 - 2} + \frac{\nu_2 \sin^2\theta}{\nu_2 - 2}\right]^2 \left[\frac{\nu_1^2 \cos^4\theta}{(\nu_1 - 2)^2(\nu_1 - 4)} + \frac{\nu_2^2 \sin^4\theta}{(\nu_2 - 2)^2(\nu_2 - 4)}\right]^{-1},$$

$$c = \left[\frac{f - 2}{f}\left\{\frac{\nu_1 \cos^2\theta}{\nu_1 - 2} + \frac{\nu_2 \sin^2\theta}{\nu_2 - 2}\right\}\right]^{1/2}, \qquad \nu_1, \nu_2 > 4. \quad (28.103)$$

These values are chosen to make the first four moments of the two distributions agree.

The approximation is exact when $\cos\theta = 1$ or $\sin\theta = 1$. It gives satisfactory results, even for ν_1, ν_2 as small as 7, in the central part of the distribution ($|D_\theta| < 5$). The approximation is very good when both ν_i's are at least 24. The relative error is large in extreme tails.

Weir (1960b) has also provided the following approximations for the upper percentage point of a related statistic. He found that

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{[(n_1 - 1)S_1^2/\{(n_1 - 3)n_1\}] + [(n_2 - 1)S_2^2/\{(n_2 - 3)n_2\}]}}$$

has upper 2.5% points ($\alpha = 0.975$) between 1.96 and 2, provided that $n_1 \geq 6$ and $n_2 \geq 6$.

Welch (1938) suggested approximating the distribution of $n_1^{-1}S_1'^2 + n_2^{-1}S_2'^2$ by that of $c\chi_\nu^2$ with

$$c\nu = E\left[n_1^{-1}S_1'^2 + n_2^{-1}S_2'^2\right] = n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2, \qquad (28.104a)$$

$$2c^2\nu = \text{var}\left[n_1^{-1}S_1'^2 + n_2^{-1}S_2'^2\right] = 2n_1^{-2}(n_1 - 1)^{-1}\sigma_1^4 + 2n_2^{-2}(n_2 - 1)^{-1}\sigma_2^4; \qquad (28.104b)$$

that is,

$$c = \frac{n_1^{-2}(n_1 - 1)^{-1}\sigma_1^4 + n_2^{-2}(n_2 - 1)\sigma_2^4}{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2},$$

$$\nu = \frac{(n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2)^2}{n_1^{-2}(n_1 - 1)\sigma_1^4 + n_2^{-2}(n_2 - 1)\sigma_2^4}.$$

Then $(\bar{X}_1 - \bar{X}_2)(n_1^{-1}S_1'^2 + n_2^{-1}S_2'^2)^{-1/2}$ is approximately distributed as

$$\frac{U\sqrt{n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2}}{\sqrt{c\nu}\left(\chi_\nu/\sqrt{\nu}\right)} = t_\nu$$

(since cv $= n_1^{-1}\sigma_1^2 + n_2^{-1}\sigma_2^2$). Further work was done on this problem by Aspin (1948) who has provided some tables [Aspin (1949)] from which exact probabilities can be obtained [see also Welch (1949)].

Rahman and Saleh (1974) derive the distribution of D_θ for all combinations of ν_1 and ν_2. The expression is rather complicated. In the special case where $\nu_1 = \nu_2 = \nu$, we have

$$p_D(d) = \frac{B\{\frac{1}{2}(\nu+1), \frac{1}{2}(\nu+1)\}\Gamma(\nu+\frac{1}{2})}{\sqrt{(\nu\pi)}\left(\Gamma(\frac{1}{2}\nu)\right)^2} \csc 8 \cot'' 8$$

$$\times {}_2F_1\left(\nu + \frac{1}{2}, \frac{1}{2}(\nu+1); \frac{1}{2}\nu + 1; \frac{-d^2}{\nu \sin^2\theta}\right), \quad (28.105)$$

where ${}_2F_1(\cdot, . ; . ; \cdot)$ is the Gaussian hypergeometric function [Chapter 1, Eq. (1.104)].

Rahman and Saleh (1974) provide 97.5% and 95% points of the distribution of D_θ for $\nu_1 = 6(1)15$ and $\nu_2 = 6(1)9$. The numerical evaluation is accomplished by evaluating the Appell function using a Gaussian quadrature numerical integration technique. It should be noted that Behrens (1929) provided an integral form of the distributions of D_θ for various degrees of freedom. Fisher (1935) confirmed the result and extended Behrens's theory. Later Fisher and Healy (1956) gave the exact distributions of D_θ for small odd values of the degrees of freedom.

Molenaar (1977) provided inter alia, the two following approximations to $\Pr[a_1 t_{\nu_1} - a_2 t_{\nu_2}$ I $d]$. Assume that 5 I $\nu_1 \le \nu_2$ (take the complement if $\nu_1 > \nu_2$). (The case $\nu_1 < 5$ is not considered.)

Method U

Take $\Phi(u)$, where Φ denotes the unit normal cdf function and $u = d/(\omega_1 + \omega_2)^{1/2}$, with $\omega_i = a_i^2\nu_i/(\nu_i - 2)$ $(i = 1, 2)$.

Method V

Take $\Phi(v)$, where $v = \operatorname{sgn}(t)\left\{f - \frac{2}{3} + \frac{1}{10f}\right\}\left\{\frac{\log\left(1 + (t^2/f)\right)}{f - (5/6)}\right\}^{1/2}$,

$$f = 4 + \frac{(\omega_1 + \omega_2)^2}{\{\omega_1^2/(\nu_1 - 4)\} + \{\omega_2^2/(\nu_2 - 4)\}},$$

and

$$t = d\left(1 - \frac{2}{f}\right)^{-1/2}(\omega_1 + \omega_2)^{-1/2}$$

(This is a simplified Patil method.)
Including

Method* W. *Exact calculation,
Molenaar recommends the following system:

For maximum absolute error of at most:	Use method:		
0.01	U if $\nu_1 \geq 16$	V if $6 \leq \nu_1 \leq 15$	W if $\nu_1 = 5$
0.005	U if $\nu_1 \geq 30$	V if $7 \leq \nu_1 \leq 29$	W if $\nu_1 = 6$
0.002	U if $\nu_1 \geq 72$	V if $9 \leq \nu_1 \leq 71$	W if $6 \leq \nu_1 18$
0.001	U if $\nu_1 \geq 140$	V if $12 \leq \nu_1 \leq 139$	W if $7 \leq \nu_1 \leq 11$
0.0005	U if $\nu_1 \geq 273$	V if $16 \leq \nu_1 \leq 272$	W if $9 \leq \nu_1 \leq 15$
0.0002	U if $\nu_1 \geq 680$	V if $23 \leq \nu_1 \leq 679$	W if $14 \leq \nu_1 \leq 22$
0.0001	U if $\nu_1 \geq 1310$	V if $32 \leq \nu_1 \leq 1309$	W if $19 \leq \nu_1 \leq 31$

The error in method U is rather robust against changes in the parameter. The largest error occurs for probability values between 0.20 and 0.25 and between 0.75 and 0.80. The secondary minimum and maximum occur very close to 0.01 and 0.99, respectively. For method V the largest discrepancies are at the values of probability between 0.72 and 0.75 and 0.25 and 0.28, respectively, with the secondary maxima and minima in the vicinity of 0.01–0.02 and 0.98–0.99.

Nel, van der Merwe and Moser (1990) have tackled the Behrens-Fisher problem directly, and they have provided an exact distribution for

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\{n_1^{-1}S_1'^2 + n_2^{-1}S_2'^2\}^{1/2}}.$$

The distribution is a generalization of the noncentral F-distribution and will be discussed in Chapter 30. Hajek (1962) had shown that if $\sum_{j=1}^k \lambda_j = 1$, and

$$T = U\left[\sum_{j=1}^{k} \lambda_j \chi_{\nu_j}^2 \nu_j^{-1}\right]^{-1/2},$$

U being a unit normal variable and if U and the χ^2's are mutually independent, then for $t' \leq 0 \leq t''$, $\Pr[t' \leq T \leq t'']$ lies between $\Pr[t' \leq t_\nu \leq t'']$ and $\Pr[t' \leq t_m \leq t'']$ (calculated on normal theory), where $m = \sum_{j=1}^k \nu_j$ and ν is any integer not exceeding $\min(\nu_j/\lambda_j)$.

Wallgren (1980) investigated the distribution of

$$W = \frac{XY}{S^2}, \tag{28.106}$$

where the joint distribution of X and Y is bivariate normal [see Chapter 32, Eq. (32.2)], $\nu S^2/\sigma^2$ is distributed as X_ν^2 [see Chapter 18, Equation (18.5)] and (X, Y) and S are mutually independent. Also

$$\begin{aligned} E[X] &= \xi, E[Y] = \eta, \\ \operatorname{var}(X) &= \operatorname{var}(Y) = \sigma^2, \\ \operatorname{corr}(X, Y) &= \rho. \end{aligned} \tag{28.107}$$

The special case p = 0 (X and Y independent), with ξ and η equal to zero was solved by Harter (1951).

In general, W is distributed as the product of two *correlated* noncentral t variables (see Chapter 31),

$$W_1 = \frac{X}{S} \qquad \left[\text{distributed as } t'_\nu\left(\frac{\xi}{\sigma}\right)\right]$$

and

$$W_2 = \frac{Y}{S} \qquad \left[\text{distributed as } t'_\nu\left(\frac{\eta}{\sigma}\right)\right].$$

When $\xi = \eta = 0$, W_1 and W_2 are correlated random variables, each distributed as ς,. Wallgren (1980) obtained the following expressions for the cdf of W in this case:

For $w < 0$,

$$F_W(w) = \int_{\varepsilon_1}^{0} Q_\nu(\theta; \rho, w)\, d\theta, \tag{28.108a}$$

where

$$\varepsilon_1 = \begin{cases} \alpha - \pi & \text{for } \rho < 0, \\ \alpha & \text{forp} > 0, \end{cases} \quad \text{with} \quad \alpha = \tan^{-1}\left\{-(1-\rho^2)^{1/2}/\rho\right\}, \quad 0 \le \alpha \le \pi,$$

and

$$\begin{aligned} Q_\nu(\theta; \rho, w) = \pi^{-1}\Big[&\{\nu \sin 8 \sin(\theta + \cos^{-1}\rho)\} \\ &\times \{w + \nu \sin e \sin(\theta + \cos^{-1}\rho)\}\Big]^{1/2}; \end{aligned}$$

for w > 0,

$$F_W(w) = 1 - \int_0^{\varepsilon_2} Q_\nu(\theta; \rho, w)\, d\theta, \tag{28.108b}$$

where

$$\varepsilon_2 = \begin{cases} \alpha & \text{for } \rho < 0, \\ \alpha + \pi & \text{for } \rho > 0 \end{cases}$$

Roy, Roy, and Ali (1993) have introduced a binomial mixture of t-distributions with density function

$$p_X(x|n,p,\nu) = \sum_{r=0}^{n} \binom{n}{r} \frac{p^r(1-p)^{n-r} x^{2r}}{B\left(r+\frac{1}{2},\frac{\nu}{2}\right)\nu^{r+\frac{1}{2}}\left(1+\frac{x^2}{\nu}\right)^{((\nu+1)/2)+r}},$$

$$-\infty < x < \infty$$

The moments of X are given by

$$E[X^{2k+1}] = 0$$

and

$$E[X^{2k}] = \nu^k \sum_{r=0}^{n} \binom{n}{r} p^r(1-p)^{n-r} \frac{\Gamma(r+k+\frac{1}{2})\Gamma\left(\frac{\nu}{2}-k\right)}{\Gamma(r+\frac{1}{2})\Gamma\left(\frac{\nu}{2}\right)}.$$

In particular, we obtain

$$E[X] = 0, \operatorname{var}(X) = \frac{2\nu}{\nu-2}\left(np+\frac{1}{2}\right),$$

$$\beta_1(X) = 0, \quad \text{and} \quad \beta_2(X) = 3\left(\frac{\nu-2}{\nu-4}\right) - \frac{np^2(1+2n)(\nu-2)}{(np+\frac{1}{2})^2(\nu-4)}.$$

McDonald and Newey (1988) introduced a generalized t distribution with density function

$$p_X(x|p,q) = \frac{p}{2q^{1/p}B\left(\frac{1}{p},q\right)(1+|x|^p/q)^{q+(1/p)}},$$

$$-\infty < x < \infty, \quad p,q > 0,$$

which clearly includes the t density in (28.2) as a special case when $p = 2$ and $q = 2\nu$. (In fact, it is the density function of $t_\nu/\sqrt{2}$). The above generalized density also includes as a special case Box and Tiao's *power-*

Table 28.18 Values of the coefficient of kurtosis, β_2, for the generalized t distribution for various choices of p and q

	q = 1.0	2.0	5.0	10	50	100	∞
p = 0.5				635.0	35.8	29.8	25.2
1.0			36.0	10.3	6.53	6.25	6.00
1.5			6.68	4.68	3.90	3.83	3.76
2.0			4.00	3.38	3.06	3.03	3.00
3.0		4.11	2.72	2.54	2.44	2.43	2.42
5.0	4.28	2.38	2.15	2.11	2.08	2.07	2.07
10.0	2.07	1.94	1.90	1.89	1.89	1.88	1.88
50.0	1.81	1.81	1.81	1.80	1.80	1.80	1.80
100.0	1.80	1.80	1.80	1.80	1.80	1.80	1.80

exponential density or Subbotin distribution [see Eq. (24.83)', for example]

$$pe^{-|x|^p}/\{2\Gamma(p^{-1})\}, \qquad -\infty < x < \infty, \quad p > 0,$$

when $q \to \infty$. Both these density functions are symmetric about zero. While the odd order moments of X are consequently zero, the even order moments of X are given by

$$E[X^{2k}] = q^{k/p}B\left(\frac{k+1}{p}, q - \frac{k}{p}\right)\Big/B\left(\frac{1}{p}, q\right).$$

Table *28.18,* taken from McDonald (1991), presents values of the coefficient of kurtosis, $\beta_2(X)$, for various choices of p and q. McDonald (1984) has shown that the above generalized t distribution is, in fact, a mixture of the generalized gamma distribution and the power-exponential distribution of Box and Tiao.

McDonald and Newey (1988) used the generalized t distributions to develop partially adaptive estimation of regression models; also see McDonald and Nelson (1989). Butler et al. (1990) discussed the robust estimation of regression models by using the generalized t distributions. Similarly, partially adaptive estimation of ARMA time series models via the generalized t distributions has been developed by McDonald *(1989).*

REFERENCES

Abramowitz, M., and Stegun, I. A. (eds.) (1964). ***Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,*** Applied Mathematics Series 55, National Bureau of Standards, Washington, D.C.: GPO.

Albert, J., Delampady, M., and Polarsk, W. (1991). A class of distributions for robustness studies, ***Journal of Statistical Planning and Inference, 28,*** **291–304.**

Ali, M. M. (1975). Tail distributions of "Student's" ratio for $t \geq n - 1$ in samples of size n from rectangular distribution, ***Journal of Statistical Research*, 9,** 11–24.

Ali, M. M. (1976). Tail distribution of "Student's" ratio for $t \geq (n-2)/2$ in samples of size n from rectangular distribution, Journal of Statistical Research, **10,** 43–71.

Amos, D. E. (1964). Representations of the central and non-central t distributions, Biornetrika, **51,** 451–458.

Andersen, J. B., Lauritzen, S. L., and Thommesen, C. (1990). Distributions of phase derivatives in mobile communications, IEEE Proceedings, *Microwaves*, Antennas and Propagation, **137,** 197–201.

Angers, J.-F. (1992). Use of the Student-t prior for the estimation of normal means: A computational approach, In Bayesian Statistics 4 (Eds., J. M. Bernardo, J. 0. Berger, A. P. Dawid, and A. F. M. Smith), Oxford: Oxford University Press, pp. 567–575.

Anscombe, F. J. (1950). Table of the hyperbolic transformation $\sinh^{-1}\sqrt{x}$, Journal of the Royal Statistical Society, Series A, **113,** 228–229.

Aspin, A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values, Biometrika, **35,** 88–96.

Aspin, A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated, Biometrika, **36,** 290–293.

Babanin, B. V. (1952). Nomogram of basic statistical distributions and its application to some problems of sampling method, Academy of Sciences, USSR, Institute of Mechanics, Engineering Transactions, **11,** 169–180 (In Russian).

Bailey, B. J. R. (1980). Accurate normalizing transformations of a Student's t-variate, Applied Statistics, **29,** 304–306.

Baldwin, E. M. (1946). Table of percentage points of the t-distribution, Biometrika, **33,** 362.

Banerjee, S. K. (1957). A lower bound to the probability of Student's ratio, *Sankhyā*, Series B, **19,** 391–394.

Bartlett, M. S. (1935). The effect of non-normality on the t distribution, Proceedings of Cambridge Philosophical Society, **31,** 223–231.

Behrens, W. V. (1929). Ein Beitrag zur Fehlenberechnung bei wenigen Beobachtungen, Landwirtschaftliche Jahrbucher, 68, 807–837.

Bhattacharyya, A. (1952). On the uses of the t-distribution in multivariate analysis, *Sankhyā*, **12,** 89–104.

Birnbaum, Z. W. (1942). An inequality for Mills' ratio, Annals of Mathematical Statistics, **13,** 245–246.

Birnbaum , Z. W. (1970). On a statistic similar to Student's t, In *Nonparametric* Techniques in Statistical Inference (Ed., M. L. Puri), New York: Cambridge University Press, pp. 427–433.

Birnbaum, Z. W., and Vincze, I. (1973). Limiting distributions of statistics similar to Student's *t*, Annals of Statistics, 1, 958–967.

Blattberg, R. C., and Gonedes, N. J. (1974). A comparison of the stable and Student distributions as statistical models for stock prices, Journal of Business, 47, 244–280.

Bondesson, L. (1981). When is the t-statistic 1-distributed? Statistical Research Report, University of Umeå, S-90187, Umeå, Sweden.

Borwein, P., and Gabor, G. (1984). On the behavior of the MLE of the scale parameter of the Student family, *Communications in Statistics—Theory and Methods,* **13,** 3047–3057.

Box, J. F. (1981). Gosset, Fisher and the t distribution, *The American Statistician, 35,* 61–67.

Bracken, J., and Schleifer, A., Jr. (1964). *Tables for normal sampling with unknown cariances: The Student distribution and economically optimum sampling plans,* Division of Research, Harvard University.

Bradley, R. A. (1952). The distribution of the t and F statistics for a class of non-normal populations, *Virginia Journal of Sciences, 3,* 1–32.

Buehler, R. J., and Feddersen, A. P. (1963). Note on a conditional property of Student's t, *Annals of Mathematical Statistics,* **34,** 1098–1100.

Bukaf, J., and Burstein, H. (1980). Approximations of Student's t and chi-squared percentage points, *Communications in Statistics—Simulation and Computation,* **12,** 665–672.

Butler, R. J., McDonald, J. B., Nelson, R. D., and White, S. B. (1990). Robust and partially adaptive estimation of regression models, *Reciew of Economics and Statistics,* **72,** 321–326.

Cacoullos, T. (1965). A relation between t and F distributions, *Journal of the American Statistical Association, 60,* 528–531 (Correction: 60, 1249).

Chaubey, Y. P., and Mudholkar, G. S. (1982). Difference of two t-variables, *Communication in Statistics—Theory and Methods,* **11,** 2335–2342.

Cheng, S. W., and Fu, J. C. (1983). An algorithm to obtain the critical values of t, χ^2 and F distributions, *Statistics & Probability Letters, 1,* 223–227.

Chu, J. T. (1956). Errors in normal approximations to the t, r, and similar types of distribution, *Annals of Mathematical Statistics,* **27,** 780–789.

Chung, K. L. (1946). The approximate distribution of Student's statistic, *Annals of Mathematical Statistics,* **17,** 447–465.

Cornish, E. A. (1969). Fisher Memorial Lecture (37th Session, International Statistical Institute, London).

Cotterman, T. E., and Knoop, P. A. (1968). *Tables of limiting t values for probabilities to the nearest* .001 (n = 2–16), Report AMRL-TR-67-161, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Dayton, OH.

Cox, M. A. A. (1991). The implementation of functions to evaluate percentage points of the normal and Student's t distributions on a spreadsheet, *The Statistician,* 40, 87–94.

Cucconi, 0 . (1962). On a simple relation betwen the number of degrees of freedom and critical value of Student's t, *Memoire Accademia Patacina,* **74,** 179–187.

Deming, W. E., and Birge, R. T. (1934). On the statistical theory of errors, *Review* of *Modern Physics, 6,* 119–161.

Dickey, J. M. (1967). Expansions of t densities and related complete integrals, *Annals of Mathematical Statistics,* **38,** 503–510.

Dickey, J. M. (1976). A new representation of Student's t as a function of independent t's with a generalization to the matrix t, *Journal of Multicariate Analysis, 6,* 343–346.

Efron, B. (1968). ***Student's t-test under non-normal conditions,*** Technical Report No. 21, Harvard University, Department of Statistics.

Eggers, P. C. F., and Andersen, J. B. (1989). Measurements of complex envelopes of mobile scenarios at 450 MHz, IEEE ***Transactions on Vehicular Technology,*** **38,** 37–42.

Eisenhart, C. (1979). On the transition from "Student's" z to "Student's *t*", ***The American Statistician,*** **33,** 6–11.

Elfving, G. (1955). An expansion principle for distribution functions, with application to Student's statistic, ***Annales Academiae Scientiarum Fennicae, Series A,*** **204,** 1–8.

Epstein, B. (1977). Infinite divisibility of Student's *t*-distribution, ***Sankhyā, Series*** B, ***39,*** 103–120.

Fan, T.-H., and Berger, J. O. (1992). Behavior of the posterior distribution and inferences for a normal mean with *t* prior distributions, ***Statistics and Decisions,*** **10,** 99–120.

Federighi, E. T. (1959). Extended tables of the percentage points of Student's t-distribution, ***Journal of the American Statistical Association,*** **54,** 683–688.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, ***Philosophical Transactions of the Royal Society of London, Series A,*** **222,** 309–368.

Fisher, R. A. (1925a). Applications of "Student's" distribution, ***Metron,*** **5,** 90–104.

Fisher, R. A. (1925b). Expansion of "Student's" integral in powers of n^{-1}, ***Metron,*** **5,** 109–112.

Fisher, R. A. (1935). The mathematical distributions used in the common tests of significance, ***Econometrica,*** **3,** 353–365.

Fisher, R. A., (1941). The asymptotic approach to Behrens' integral with further table for the d test of significance, ***Annals of Eugenics,*** **11,** 141–172.

Fisher, R. A., and Cornish, E. A. (1960). The percentile points of distribution having known cumulants, ***Technometrics,*** **2,** 209–225.

Fisher, R. A., and Healy, M. J. R. (1956). New tables of Behrens' test of significance, ***Journal of the Royal Statistical Society, Series B,*** **18,** 212–216.

Fisher, R. A., and Yates, F. (1966). ***Statistical Tables for Biological, Agricultural and Medical Research,*** Edinburgh: Oliver and Boyd.

Fraser, D. A. S. (1976). Necessary analysis and adaptive inference (with discussion), ***Journal of the American Statistical Association, 71,*** 99–113.

Fujihara, R.. and Park, K. (1990). The probability distribution of future prices in the foreign exchange market: A comparison of candidate processes, ***Journal of Futures Market,*** **10,** 623–641.

Fujikoshi, Y., and Mukaihata, S. (1993). Approximations for the quantiles of Student's *t* and F-distributions and their bounds, ***Hiroshima Mathematical Journal,*** **23,** 557–564.

Gambino, J., and Guttman, I. (1984). A Bayesian approach to prediction in the presence of spurious observation for several models, ***Communications in Statistics —Theory and Methods,*** **13,** 791–812.

Gardiner, D. A., and Bombay, B. F. (1965). An approximation to Student's *t*, ***Technornetrics.*** 7. 71–72.

Gaver, D. P., and Kafadar, K. (1984). A retrievable recipe for inverse t, The American Statistician, 38, 308–311.

Gayen, A. K. (1949). The distribution of "Student's" t in random samples of any size drawn from non-normal universes, Biometrika, 36, 353–369.

Gayen, A. K. (1952). The inverse hyperbolic sine transformation on Student's t for non-normal samples, *Sankhyā*, **12,** 105–108.

Geary, R. C. (1936). The distribution of "Student's" ratio for non-normal samples, Journal of the Royal Statistical Society, Series B, 3, 178–184.

Gentleman, W. M., and Jenkins, M. A. (1968). An approximation for Student's t distribution, Biometrika, 55, 571–572.

Ghosh, B. K. (1975). On the distribution of the difference of two t-variables, Journal of the American Statistical Association, 70, 463–467.

Ghurye, S. G. (1949). On the use of Student's t-test in an asymmetrical population, Biometrika, **36,** 426–430.

Goldberg, H., and Levine, H. (1946). Approximate formulas for the percentage points and normalization of t and χ^2, Annals of Mathematical Statistics, **17,** 216–225.

Grosswald, E. (1976a). The Student t-distribution for odd degrees of freedom is infinitely divisible, Annals of Probability, 4, 680–683.

Grosswald, E. (1976b). The Student t-distribution for any degree of freedom is infinitely divisible, Zeitschrift *für* Wahrscheinlichkeitstheorieund *Verwandte Gebiete*, **36,** 103–109.

Guenther, W. C. (1975). Desk calculations of probabilities for the distribution of the difference of two t-variables. Research paper No. 83, S-1975-540. College of Commerce and Industry, University of Wyoming, Laramie.

Hajek, J. (1962). Inequalities for the generalized Student's distribution and their applications, Selected Translations in Mathematical Statistics and Probability, vol. 2, American Mathematical Society: Providence, RI, pp. 63–74.

Hald, A. (1952). Statistical Tables and Formulas, New York: Wiley.

Harter, H. L. (1951). On the distribution of Wald's classification statistic, Annals of Mathematical Statistics, **22,** 58–67.

Hartley, H. O., and Pearson, E. S. (1950). Table of the probability integral of the t-distribution, Biometrika, **37,** 168–172.

Hendricks, W. A. (1936). An approximation to "Student's" distribution, Annals of Mathematical Statistics, 7, 210–221.

Hill, G. W. (1969). Progress results on asymptotic approximations for Student's ratio and chi-squared, Personal communication.

Hill, G. W. (1970a). Algorithm 395: Student's t approximation, Communications of the Association for Computing Machinery, **13,** 617–619.

Hill, G. W. (1970b). Algorithm 396: Student's t-quantiles, Communications of the Assocation for Computing Machinery, **13,** 619–620.

Hill, G. W. (1970c). Student's t quantiles, *Communications* of the Association for Computing Machinery, **13,** 621–624.

Hill, G. W. (1972). Reference Table: "Student's" t-distribution quantiles to 20D. CSIRO, Australia, Division of Mathematical Statistics Technical Paper No. 35.

Hill, G. W. (1981). Remark on algorithm 395, Student's *t* distribution, ***Association for Computing Machinery, Transactions of Mathematical Software, 7***, 247–249.

Hill, G. W., and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type, ***Annals of Mathematical Statistics,*** **39**, 1264–1273.

Hogg, R. V., and Klugman, S. A. (1983). On the estimation of long-tailed skewed distributions with actuarial data, ***Journal of Econometrics,*** **23**, 91–102.

Hoq, A. K. M. S., Ali, M. M., and Templeton, J. G. C. (1978). The distribution of Student's ratio for samples from exponential population, ***Communications in Statistics—Theory and Methods, 7***, 837–850.

Hotelling, H. (1961). The behavior of some standard statistical test under non-standard conditions, ***Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability,*** *1*, 319–359.

Hotelling, H., and Frankel, L. R. (1938). The transformation of statistics to simplify their distribution, ***Annals of Mathematical Statistics,*** **9**, 87–96.

Hyrenius, H. (1950). Distribution of "Student"-Fisher *t* in samples from compound normal functions, ***Biometrika,*** **37**, 429–442.

Ifram, A. F. (1972). On the characteristic functions of the F and *t* distributions, ***Sankhyii, Series A*****, 32**, 350–352.

Isaacs, G. L., Christ, D. E., Novick, M. R., and Jackson, P. H. (1974). ***Tables for Bayesian Statisticians,*** Ames: Iowa State University.

Ismail, M. E. H., and Kelker, D. H. (1976). The Bessel polynomials and the Student *t* distribution, ***SIAM Journal on Mathematical Analysis,*** **7**, 82–91.

James-Levy, G. E. (1956). A nomogram for the integral law of Student's distribution, ***Teoriya Veroyatnostei*** i ***ee Primeneniya,*** *1*, 271–274 (In Russian). (pp. 246–248 in English translation).

Janko, J. (1958). ***Statisticke Tabulky,*** Prague.

Jeffreys, H. (1948). ***Theory of Probability,*** 2nd ed., Oxford: Clarendon.

Kafadar, K., and Tukey, J. W. (1988). A bidec *t* table, ***Journal of the American Statistical Association,*** 83, 523–529.

Kennedy, W. J., and Gentle, J. E. (1980). ***Statistical Computing,*** New York: Dekker.

Kitagawa, T. (1954–56). Some contributions to the design of sample surveys, ***Sankhyā***, 14, 317–362; **17**, 1–36.

Koehler, K. J. (1983). A simple approximation for the percentiles of the t-distribution, ***Technometrics,*** **26**, 103–106.

Kotlarski, I. (1964). On bivariate random vectors where the quotient of their coordinates follows the Student's distribution, ***Zeszyty Naukowe Politechniki Warszawskiej,*** **99**, 207–220 (In Polish).

Kramer, C. Y. (1966). Approximation to the cumulative t-distribution, ***Technometrics,*** 8, 358–359.

Krishnamoorthy, A. S. (1951). On the orthogonal polynomials associated with Student's distribution. ***Sankhyii,*** **11**, 37–44.

Laderman, J. (1939). The distribution of "Student's" ratio for samples of two items drawn from non-normal universes, ***Annals of Mathematical Statistics,*** 10,376–379.

Lampers, F. B., and Lauter, A. S. (1971). An extension of the table of the Student distribution, ***Journal of the American Statistical Association,*** **66**, 503.

Laumann, R. (1967). *Tafeln der STUDENT-oder t-Verteilung*, Deutsch-Franzo-Sisches Forschungsinstitut, Saint-Louis, Akt. N21/67.

Lauritzen, S. L., Thommesen, C., and Andersen, J. B. (1990). A stochastic model in mobile communication, ***Stochastic Processes and Their Applications,*** **36,** 165–172.

Lee, C. M.-S., and Singh, K. P. (1988). On the t cumulative probabilities, ***Communications in Statistics—Simulation and Computation,*** **17,** 129–135.

Ling, R. F. (1978). A study of the accuracy of some approximations for t, χ^2, and F tail probabilities, ***Journal of the American Statistical Association,*** **73,** 274–283.

Logan, B. F., Mallows, C. L., Rice, S. O., and Shepp, L. A. (1973). Limit distributions of self-normalized sums, ***Annals of Probability,*** *1*, 788–809.

Lozy, M. El (1982). Efficient computation of the distribution functions of Student's t, chi-squared and F to moderate accuracy, ***Journal of Statistical Computation and Simulation,*** **14,** 179–189.

Mardia, K. V., and Zemroch, P. J. (1978). ***Tables of the F- and Related Distributions with Algorithms,*** New York: Academic Press.

Mauldon, J. G. (1956). Characterizing properties of statistical distributions, ***Quarterly Journal of Mathematics,*** Oxford, 7, 155–160.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income, ***Econometrica***, 52, 647–663.

McDonald, J. B. (1989). Partially adaptive estimation of ARMA time series models, ***International Journal of Forecasting,*** *5*, 217–230.

McDonald, J. B. (1991). Parametric models for partially adaptive estimation with skewed and leptokurtic residuals, ***Economics Letters,*** **37,** 237–278.

McDonald, J. B., and Butler, R. J. (1987). Some generalized mixture distributions with an application to employment duration, ***Review of Economics and Statistics,*** 69, 232–240.

McDonald, J. B., and Nelson, R. D. (1989). Alternative beta estimation for the market model using partially adaptive techniques, ***Communications in Statistics-Theory and Methods,*** **18,** 4039–4058.

McDonald, J. B., and Newey, W. K. (1988). Partially adaptive estimation of regression models via the generalized t distribution, ***Econometric Theory,*** 4, 428–457.

McLeay, S. (1986). Student's t and the distribution of financial ratios, ***Journal of Business Finance and Accounting,*** **13,** 209–222.

McMullen, L. (1939). "Student" as a man, ***Biometrika,*** **30,** 205.

Mickey, M. R. (1975). Approximate tail probabilities for Student's t distribution, ***Biometrika,*** **62,** 210–217.

Mirza, M., and Boyer, K. L. (1992). Performance evaluation of a class of M estimators for surface parameter estimation in noise range data, ***Proceedings, SPIE—The International Society for Optical Engineering,*** **1708,** 198–209.

Mitra, S. S. (1978). Recursive formula for the characteristic function of Student t distributions for odd degrees of freedom, ***Manuscript,*** Pennsylvania State University, State College.

Molenaar, W. (1977). The Behrens-Fisher distribution and its approximation, ***Bulletin No. 15***, Rijksuniversiteit Groningen, (Sociological Institute) Netherlands.

Molina, E. C., and Wilkinson, R. I. (1929). The frequency distribution of the unknown mean of a sampled universe, Bell System Technical Journal, 8, 632–645.

Moran, P. A. P. (1966). Accurate approximations for t-tests, In Research Papers in Statistics, Festschrift for J. Neyman (Ed. F. N. David), pp. 225–230.

Mudholkar, G. S., and Chaubey, Y. P. (1976a). A simple approximation for the doubly noncentral t distribution, Communications in Statistics, B5, 85–92.

Mudholkar, G. S., and Chaubey, Y. P. (1976b). Use of logistic distribution for approximating probabilities and percentiles of Student's t distribution, Journal of Statistical Research, 9, 1–9.

Nel, D. G., van der Merwe, C. A., and Moser, B. K. (1990). The exact distributions of the univariate and multivariate Behrens-Fisher statistics with a comparison of several solutions in the univariate case, Communications in Statistics—Theory and Methods, **19,** 279–298.

Owen, D. B. (1962). Handbook of Statistical Tables, Reading, MA: Addison-Wesley.

Owen, D. B. (1965). The power of Student's t-test, Journal of the American Statistical Association, **60,** 320–333.

Patil, V. H. (1965). Approximation to the Behrens-Fisher distributions, Biometrika, 52, 267–271.

Pearson, E. S., and Hartley, H. O. (1958) Biometrika Tables for Statisticians, vol. 1, 2d ed., Cambridge: Cambridge University Press.

Peiser, A. M. (1943). Asymptotic formulas for significance levels of certain distributions, Annals of Mathematical Statistics, **14,** 56–62.

Peizer, D. B., and Pratt, J. W. (1968). A normal approximation for binomial, F, beta, and other common, related tail probabilities, I, Journal of the American Statistical Association, **63,** 1416–1456.

Perlo, V. (1933). On the distribution of Student's ratio for samples of three drawn from a rectangular distribution, Biometrika, **25,** 203–204.

Pillai, K. C. S. (1951). On the distribution of an analogue of Student's t, Annals of Mathematical Statistics, 22, 469–472.

Pinkham, R. S., and Wilk, M. B. (1963). Tail areas of the t-distribution from a Mills'-ratio-like expansion, Annals of Mathematical Statistics, **34,** 335–337.

Praetz, P. D. (1972). The distribution of share price changes, Journal of Business, **45,** 49–55.

Praetz, P. D., and Wilson, E. J. G. (1978). The distribution of stock market returns: 1958–1973, Australian Journal of Management, **3,** 79–90.

Pratt, J. W. (1968). A normal approximation for binomial, F, beta and other common, related tail probabilities, II, Journal of the American Statistical Association, **63,** 1457–1483.

Prescott, P. (1974). Normalizing transformations of Student's t distribution, Biometrika, 61, 177–180.

Psarakis, S., and Panaretos, J. (1990). The folded t-distribution, Communications in Statistics—Theory and Methods, 19, 2717–2734.

Quensel, C. E. (1943). An extension of the validity of "Student"-Fisher's law of distribution, Skandinacisk *Aktuarietidskrift*, **26,** 210–219.

Rahman, M., and Saleh, A. K. M. E. (1974). Explicit form of the distribution of the Behrens-Fisher d-statistic, Journal of the Royal Statistical Society, Series B, 36, 54–60. [Corrigendum: p. 466 [to Tables 1 and 2 and eqs. (7), (9), (10), (12), (14), (16), and (22);] see Isaacs et al. (1974).]

Rao, C. R., Mitra, S. K., and Mathai, A. (1966). Formulae and Tables for Statistical Work, Calcutta, India: Statistical Publishing Society.

Ratcliffe, J. F. (1968). The effect on the t-distribution of non-normality in the sampled population, Applied Statistics, 17, 42–48.

Ray, J. P. (1961). Unpublished thesis, Virginia Polytechnic Institute.

Richter, W.-D., and Gundlach, G. (1990). Asymptotic quantile approximation for Student's t-distribution, Rostocker Mathematisches Kolloquium, 42, 53–58.

Rider, P. R. (1929). On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes, Biometrika, 21, 124–143.

Rider, P. R. (1931). On small samples from certain non-normal universes, Annals of Mathematical Statistics, 2, 48–62.

Rietz, H. L. (1939). On the distribution of the "Student" ratio for small samples from certain non-normal populations, Annals of Mathematical Statistics, 10, 265–274.

Roy, M. K., Roy, A. K., and Ali, M. Masoom (1993). Binomial mixtures of some standard distributions, Journal of Information & Optimization Sciences, 14, 57–71.

Ruben, H. (1960). On the distribution of the weighted difference of two independent Student variables, Journal of the Royal Statistical Society, Series B, 22, 188–194.

Sampford. M. R. (1953). Some inequalities on Mills' ratio and related functions, Annals of Mathematical Statistics, 24, 130–132.

Sansing, R. C. (1976). The t-statistic for a double exponential distribution, SLAM Journal on Applied Mathematics, 31, 634–645.

Sansing, R. C., and Owen, D. B. (1974). The density of the t-statistic for non-normal distributions, Communications in Statistics, 3, 139–155.

Scheffé, H. (1943). On solutions of the Behrens-Fisher problem based on the t-distribution, Annals of Mathematical Statistics, 14, 35–44.

Sichel, H. S. (1949). The method of frequency-moments and its application to Type VII populations, Biometrika, 36, 404–425.

Siddiqui, M. M. (1964). Distribution of Student's t in samples from a rectangular universe, *Review* of the International Statistical Institute, 32, 242–250.

Simaika, J. B. (1942). Interpolation for fresh probability levels between the standard table levels of a function, Biometrika, 32, 263–276.

Sinclair, C. D. (1980). Two approximations for Student's t-distribution, Report, Department of Statistics, University of St. Andrews, Scotland.

Smirnov, N. V. (1961). Tables for the Distribution and Density Functions of t-Distribu*tion* ("Student's" Distribution), Oxford: Pcrgarnon.

Smith, M. D. (1992). Comparing the exact distribution of the t-statistic to the Student's distribution when its constituent normal and χ^2 variables are dependent, Communications in Statistics—Theory and Methods, 21, 3589–3600.

Soms, A. P. (1976). An asymptotic expansion for the tail area of the t-distribution, Journal of the American Statistical Association, 71, 728–730.

Soms, A. P. (1984). A note on an extension of rational bounds for the t-tail area to arbitrary degrees of freedom, ***Communications in Statistics—Theory and Methods,*** **13,** 887–893.

Sprott, D. A. (1980). Maximum likelihood in small samples: Estimation in the presence of nuisance parameters, ***Biometrika,*** 67, 515–523.

Stammberger, A. (1967). Über einige Nomogramme zur Statistik, ***Wissenschaftliche Zeitschrift der Humboldt-Universität Berlin, Mathematisch-Naturwissenschaftliche Reihe, 16,*** 86–93.

Stone, M. (1963). The posterior t distribution, ***Annals of Mathematical Statistics,*** **34,** 568–573.

Stuart, A., and Ord, J. K. (1994). ***Kendall's Adcanced Theory of Statistics—Vol. 1,*** 6th ed., London: Edward Arnold.

"Student" (1908). On the probable error of the mean, ***Biometrika,*** 6, 1–25.

"Student" (1925). New tables for testing the significance of observations, ***Metron,*** 5, 105–108, 114–120.

Sukhatme, P. V. (1938). On Fisher and Behrens' test of significance for difference in means of two normal samples, ***Sankhyā***, 4, 39–48.

Tague, J. (1969). Monte Carlo tables for the S-statistic, ***Unpublished Report,*** Memorial University of Newfoundland, St. Johns', Newfoundland, Canada.

Taylor, S. J. (1980). The variance of the maximum likelihood estimate of the shape parameter of the Student distribution, ***Manuscript,*** Department of Operational Research, University of Lancaster, England.

Taylor, S. J., and Kingsman, B. G. (1979). An analysis of the variance and distribution of commodity price-changes, ***Australian Journal of Management,*** **4,** 135–149.

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation, ***Annals of Mathematical Statistics, 6,*** 213–219.

Tiku, M. L. (1963). Approximation to Student's t distribution in terms of Hermite and Laguerre polynomials, ***Journal of the Indian Mathematical Society, 27,*** 91–102.

Tiku, M. L., and Kumra, S. (1985). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t), ***Selected Tables in Mathematical Statistics,*** vol. 8, Providence, RI: American Mathematical Society, pp. 141–270.

Tiku, M. L., and Suresh, R. P. (1992). A new method of estimating for location and scale parameters, ***Journal of Statistical Planning and Inference,*** **30,** 281–292.

Vaughan, D. C. (1992). On the Tiku-Suresh method of estimation, ***Communications in Statistics—Theory and Methods,*** **21,** 451–469.

Verdinelli, I., and Wasserman, L. (1991). Bayesian analysis of outlier problems using the Gibbs sampler ***Statistics and Computing,*** **1,** 105–117.

Veselá, A. (1964). Kritische Werte der Studentschen t-Verteilung für die Freiheitsgrade zwischen 30 und 120, ***Biometrische Zeitschrift,*** 6, 123–137.

Wallace, D. L. (1959). Bounds on normal approximations to Student's and the Chi-square distributions, ***Annals of Mathematical Statistics,*** **30,** 1121–1130.

Wallgren, C. M. (1980). The distribution of the product of two correlated t variates, ***Journal of the American Statistical Association, 75,*** 996–1000.

Walsh, J. E. (1947). Concerning the effect of intraclass correlation on certain significance tests, ***Annals of Mathematical Statistics,*** **18,** 88–96.

Wasow, W. (1956). On the asymptotic transformation of certain distributions in the normal distribution, ***Proceedings of the Symposium on Applied Mathematics,*** vol. 6 (Numerical Analysis), New York: McGraw-Hill, pp. 251–259.

Watanabe, Y. (1960), (1962), (1963), (1966). The Student's distribution for a universe bounded at one or both sides, ***Journal of Gakugei, Tokushima Unicersity,*** **11,** 11–51; **12,** 5–50; **13,** 1–42; **14,** 1–53; **15,** 1–35.

Weibull, C. (1958). The distribution of the Student ratio in the case of serially correlated normal variables, ***Skandinacisk Aktuarietidskrift,*** **33,** 137–167.

Weir, J. B. de V. (1960a). Standardized ***t, Nature, London,*** **185,** 558.

Weir, J. B. de V. (1960b). Significance of the difference between two means when the populations variances may be unequal, ***Nature, London,*** **187,** 438.

Weir, J. B. de V. (1966). Table of 0.1 percentage points of Behrens's ***d, Biometrika,*** **53,** 367–368.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal, ***Biometrika,*** 29, 350–362.

Welch, B. L. (1949). Further note on Mrs. Aspin's tables and on certain approximations to the tabled function, ***Biometrika,*** **36,** 293–296.

Welch, B. L. (1958). "Student" and small sample theory, ***Journal of the American Statistical Association,*** **53,** 777–788.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions, ***Biometrika,*** **34,** 170–178.

Zackrisson, U. (1959). The distribution of "Student's" t in sample form individual non-normal populations, ***Statistical Institute of the Unicersity of Goteborg, Publications, 6,*** 7–32.

Zelen, M., and Severo, N. C. (1964). Probability Functions, In ***Handbook of Mathematical Functions*** (eds., M. Abramowitz and I. A. Stegun), National Bureau of Standards, Applied Mathematics Series **55,** Washington, DC: GPO. pp. 925–995; Table 26.10, p. 990.

CHAPTER 29

Noncentral χ^2-Distributions

1 DEFINITION AND GENESIS

If $U_1, U_2, \ldots, U_\nu$ are independent unit normal variables, and $\delta_1, \delta_2, \ldots, \delta_\nu$ are constants then the distribution of

$$\sum_{j=1} (U_j + \delta_j)^2 \tag{29.1}$$

depends on $\delta_1, \delta_2, \ldots, \delta_\nu$ only through the sum of their squares. It is called the ***noncentral χ^2 distribution with v degrees of freedom and noncentrality parameter*** $\lambda = \sum_{j=1}^{\nu} \delta_j^2$.

The symbol $\chi_\nu'^2(\lambda)$ denotes a variable with this distribution. It is derived from the symbol χ_ν^2, denoting a ***central*** χ^2 variable with ν degrees of freedom (Chapter 18), which has the same distribution as $\sum_{j=1}^{\nu} U_j^2$. In fact, when $\lambda = 0$, the "noncentral" distribution becomes the central χ^2 distribution.

Whenever justified by the context, the symbols ν and λ may be omitted and the symbols χ'^2 used. (The prime is retained to denote "noncentral.") Sometimes $\sqrt{\lambda}$, and sometimes $\frac{1}{2}\lambda$, are called the noncentrality parameter. We will not use these notations.

A simple way in which the noncentral χ^2 distribution arises is in the distribution of the sum of sauares

$$S = \sum_{j=1}^{n} \left(X_j - \bar{X}\right)^2,$$

where

$$\bar{X} = n^{-1} \sum_{j=1}^{n} X_j$$

and $X_1, X_2, \ldots, X_n$ are independent random variables with X_j distributed normally with expected value ξ_j and standard deviation σ (the same for all j) for j = 1, 2, ..., n. Clearly we can write

$$X_j = \xi_j + \sigma U_j',$$

with U_j''s independent unit normal variables. Then

$$S = \sigma^2 \sum_{j=1}^{n} \left\{U_j' + \xi_j\sigma^{-1} - \left(\bar{U}' + \bar{\xi}\sigma^{-1}\right)\right\}^2,$$

where

$$\bar{U}' = n^{-1} \sum_{j=1}^{n} U_j',$$

$$\bar{\xi} = n^{-1} \sum_{j=1}^{n} \xi_j.$$

Applying a transformation from $U_1', \dots, U_n'$ to $U_1, \dots, U_{n-1}, \bar{U}'$ (Chapter 13, Section 3) such that

$$\sum_{j=1}^{n} \left(U_j' - \bar{U}'\right)^2 = \sum_{j=1}^{n-1} U_j^2,$$

$U_1, \dots, U_{n-1}$ being independent unit normal variables, we see that

$$S = \sigma^2 \sum_{j=1}^{n-1} \left(U_j + \delta_j\right)^2,$$

where the δ_j's are linear functions of the ξ_j's, and 0,'s of the U_j''s. Setting $U_j' = 0$ for all j, it follows that 0, = 0 for all j and that

$$\sum_{j=1}^{n-1} \delta_j^2 = \sum_{j=1}^{n} \left(\xi_j - \bar{\xi}\right)^2 / \sigma^2.$$

Hence S is distributed as σ^2 times a noncentral χ^2 with n – 1 degrees of freedom and noncentrality parameter

$$\sum_{j=1}^{n} \left(\xi_j - \bar{\xi}\right)^2 / \sigma^2,$$

that is, as

$$\sigma^2 \chi_{n-1}'^2\left(\sum_{j=1}^{n} \left(\xi_j - \bar{\xi}\right)^2 / \sigma^2\right).$$

2 HISTORICAL NOTES

The distribution was obtained by Fisher (1928, p. 663) as a limiting case of the distribution of the multiple correlation coefficient (Chapter 32). He gave upper 5% points of the distribution for certain values of ν and A (Section 7). The distribution has been obtained in a number of different ways, described in outline in Section 3.

Patnaik (1949) emphasized the relevance of this distribution in approximate determination of the power of the χ^2-test and also suggested approximations to the noncentral χ^2-distribution itself. The noncentral χ^2-distribution can be regarded as a generalized Rayleigh distribution [Miller, Bernstein, and Blumenson (1958); Park (1961)] (see, for example, Chapter 18) also called the Rayleigh-Rice or Rice distribution. In this form it is used in mathematical physics.

The noncentral χ^2-distribution has also arisen in communication theory. In this context the noncentral χ^2-distribution function is called the *Marcum* Q-function, and the noncentrality parameter is interpreted as a signal-to-noise ratio. Some references in this area are Marcum (1948), Helstrom (1960), Felsen (1963), Urkowitz (1967), and Rice (1968), among many others.

3 DISTRIBUTION

The cumulative distribution function of $\chi_\nu'^2(\lambda)$ is

$$\Pr\left[\chi_\nu'^2(\lambda) \le x\right] = F(x;\nu,\lambda)$$

$$= e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\lambda\right)^j / j!}{2^{(\nu/2)+j}\Gamma\left(\frac{1}{2}\nu + j\right)} \int_0^x y^{(\nu/2)+j-1} e^{-y/2}\, dy,$$

$$x > 0, \qquad (29.2)$$

while $F(x;\nu, A) = 0$ for $x < 0$ [Patnaik (1949)]. It is possible to express $F(x;\nu, A)$, for $x > 0$, in an easily remembered form as a weighted sum of central χ^2 probabilities with weights equal to the probabilities of a Poisson distribution with expected value $\frac{1}{2}\lambda$. This is

$$F(x;\nu,\lambda) = \sum_{j=0}^{\infty} \left\{ \frac{\left(\frac{1}{2}\lambda\right)^j}{j!} e^{-\lambda/2} \right\} \Pr\left[\chi_{\nu+2j}^2 \le x\right]$$

$$= \sum_{j=0}^{\infty} \left[\frac{\left(\frac{1}{2}\lambda\right)^j}{j!} e^{-\lambda/2} \right] F(x;\nu + 2j, 0). \qquad (29.3)$$

Thus $\chi_\nu'^2(\lambda)$ can be regarded as a mixture of central χ^2 variables. This interpretation is often useful in deriving the distribution of functions of random variables, some (or all) of which are noncentral χ^2's. [See, e.g., the discussion of the noncentral F-distribution, Chapter 30, Section 3.]

The probability density function can, similarly, be expressed as a mixture of central χ^2 pdfs:

$$p(x;\nu,\lambda) = \sum_{j=0}^{\infty}\left\{\frac{\left(\frac{1}{2}\lambda\right)^j}{j!}e^{-\lambda/2}\right\}p(x;\nu+2j,0)$$

$$= \frac{\exp\left\{-\frac{1}{2}(\lambda+x)\right\}}{2^{\nu/2}}\sum_{j=0}^{\infty}\left(\frac{\lambda}{4}\right)^j\frac{x^{(\nu/2)+j-1}}{j!\Gamma\left(\frac{1}{2}\nu+j\right)}$$

$$= e^{-(\lambda+x)/2}\frac{1}{2}\left(\frac{x}{\lambda}\right)^{(\nu-2)/4}I_{(\nu-2)/2}\left(\sqrt{\lambda x}\right),\ x>0,$$

[Fisher (1928)], (29.4)

where

$$I_a(y) = \left(\tfrac{1}{2}y\right)^a\sum_{j=1}^{\infty}\frac{(y^2/4)^j}{j!\Gamma(a+j+1)}$$

is the modified Bessel function of the first kind of order a [Abramowitz and Stegun (1964)].

Although ν was an integer in our account of the genesis of $\chi_\nu'^2(\lambda)$, the distribution defined by (29.3) and (29.4) is a proper distribution for any positive ν. For convenience we omit the subscript $\chi_\nu'^2(\lambda)$ from $F(\cdot)$ and $p(\cdot)$ in this chapter, but the values of the parameters ν and λ are shown explicitly, as $F(x;\nu,$ A), $p(x;\nu,$ A), respectively.

The distribution has been derived in several different ways. Fisher (1928) gave an indirect derivation (by a limiting process). The first direct derivation was given by Tang (1938). Geometric derivations have been given by Patnaik (1949), Ruben (1960), and Guenther (1964). It is also possible to derive the distribution by a process of induction, first obtaining the distribution of $\chi_1'^2(\lambda)$ and then using the relation

$$\chi_\nu'^2(\lambda) = \chi_1'^2(\lambda) + \chi_{\nu-1}^2, \qquad (29.5a)$$

the noncentral and central χ^2's in the right-hand side being mutually independent. [See, e.g., Johnson and Leone (1964, p. 245) and Kerridge (1965)].

Hjort (1989) used an alternative decomposition

$$\chi_{\nu}'^{2}(\lambda) = Z_{\lambda} + \chi_{\nu}^{2}, \tag{29.5b}$$

where Z, is the "purely eccentric part" (or "purely noncentral part") of $\chi_{\nu}'^{2}(\lambda)$ and has a noncentral χ^2 distribution with ***zero*** degrees of freedom (and noncentrality A) [Siegel **(1979)]** with the cdf

$$\Pr[Z_{\lambda} \le z] = \sum_{j=0}^{\infty} \left\{ \frac{(\frac{1}{2}\lambda)^j}{j!} e^{-\lambda/2} \right\} \Pr\left[\chi_{2j}^2 \le z\right], \qquad z \ge 0, \tag{29.5c}$$

with **Pr[**χ_0^2 **I** z**]** = 1 for all z. Jones **(1989)** indicated Torgerson (1972) to be the first reference on noncentral χ^2**-distribution** with zero degrees of freedom.

The moment-generating function may be used [Graybill **(1961)]** or the characteristic function may be inverted by contour integration **[McNolty (1962)].** Alternatively we note that the moment-generating: function **[van** der Vaart **(1967)]** is

$$\begin{aligned}
E\left[\exp\left\{\sum_{j=1}^{\nu} t(U_j + \delta_j)^2\right\}\right] &= \prod_{j=1}^{\nu} E\left[\exp\left\{t(U_j + \delta_j)^2\right\}\right] \\
&= \prod_{j=1}^{\nu}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}u^2 + t(u+\delta_j)^2\right\} du\right] \\
&= \prod_{j=1}^{\nu}\left[(1-2t)^{-1/2}\exp\left\{\delta_j^2 t(1-2t)^{-1}\right\}\right] \\
&= (1-2t)^{-\nu/2} e^{-\lambda/2} \exp\left\{\frac{1}{2}\lambda(1-2t)^{-1}\right\} \\
&= e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j (1-2t)^{-(\nu+2j)/2}}{j!}.
\end{aligned} \tag{29.6}$$

Noting that $(1 - 2t)^{-(\nu+2j)/2}$ is the moment-generating function of $\chi_{\nu+2j}^2$, we obtain the formula (29.4).

The moment-generating function can also be written

$$\frac{\exp\{\lambda t/(1-2t)\}}{(1-2t)^{(\nu/2)-1}} \cdot \frac{1}{1-2t} = \frac{\exp\{\lambda t/(1-2t)\}}{(1-2t)^{(\nu/2)-1}} + \frac{2t\exp\{\lambda t/(1-2t)\}}{(1-2t)^{\nu/2}} \tag{29.6$'$}$$

which shows that

$$F(x;\nu,\lambda) = F(x;\nu-2,\lambda) - 2p(x;\nu,\lambda) \tag{29.7}$$

[Alam and Rizvi (1967)] as can also be obtained by integration by parts. There are a number of different forms in which the cumulative distribution function and the probability density function may be presented. We have first presented those forms which seem to be the most generally useful. We now discuss some other forms.

If ν is even, the cumulative distribution function of $\chi_\nu'^2(\lambda)$ can be expressed in terms of elementary functions. Using the relation (Chapter 18) between the integral of a χ_ν^2-distribution (with ν even) and a sum of Poisson probabilities, it can be shown that

$$\Pr\left[\chi_\nu'^2(\lambda) \le x\right] = \Pr\left[X_1 - X_2 \ge \tfrac{1}{2}\nu\right], \tag{29.8}$$

where X_1, X_2 are independent Poisson variables with expected values $\frac{1}{2}x, \frac{1}{2}\lambda$, respectively [Fisher (1928); Johnson (1959)].

It follows that the probability density function of $\chi_\nu'^2(\lambda)$ also can be expressed in terms of elementary functions when ν is even. This remains true when ν is odd because the pdf (29.4) can be written in terms of elementary functions by using the formula

$$I_{m+\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi}}\, z^{m+1/2}\left(\frac{1}{z}\frac{d}{dz}\right)^m\left(\frac{\sinh z}{z}\right) \qquad (\text{m integer}). \tag{29.9}$$

Tiku (1965) obtained an expression for the pdf in terms of the generalized Laguerre polynomials:

$$L_j^{(m)}(x) = \sum_{i=0}^{j}\frac{(-x)^i}{i!(j-i)!}\cdot\frac{\Gamma(j+m+1)}{\Gamma(i+m+1)}, \qquad m > -1, \tag{29.10}$$

as defined by Tiku (1965) [see also Chapter 1, Eq. (1.173)]. Tiku showed that

$$p(x;\nu,\lambda) = \frac{1}{2}e^{-x/2}\left(\frac{1}{2}x\right)^{(\frac{\nu}{2}-1)}\sum_{j=0}^{\infty}\frac{\left(-\frac{1}{2}\lambda\right)^j}{\Gamma\left(\frac{1}{2}\nu+j\right)}L_j^{(\frac{\nu}{2}-1)}\left(\frac{1}{2}x\right), \qquad x > 0. \tag{29.11}$$

A further alternative form is [Venables (1971)]

$$p(x;\nu,\lambda) = e^{-\lambda/2}{}_0F_1\left(\frac{1}{2}\nu;\frac{1}{4}\lambda x\right)\cdot\frac{e^{-x/2}x^{(\nu/2)-1}}{2^{\nu/2}\Gamma\left(\frac{1}{2}\nu\right)}, \qquad x > 0 \quad (29.11)'$$

where ${}_0F_1$ is defined in Chapter 1.

Formula *(29.2)* for the cumulative distribution can be rearranged by expanding $e^{-\lambda/2}$ in powers of $\frac{1}{2}\lambda$, collecting together like powers of $\frac{1}{2}\lambda$ and interchanging the order of summation. The resulting expression is compactly represented as

$$F(x;\nu,\lambda) = \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j}{j!} \Delta^j g_0 \tag{29.12}$$

or $e^{\lambda\Delta/2} g_0$, symbolically, where $g, = \Pr[\chi^2_{\nu+2j} \le x]$ and A is the forward difference operator $(\Delta g_m = g_{m+1} - g_m)$. This formula was given by Bol'shev and Kuznetzov **(1963)**.

Extending Tiku's result, **Gideon** and Gurland *(1977)* provided Laguerre expansions of form

$$F(x;v,A) \sim \int_0^{x_1} \frac{\lambda'^{\alpha+1} y^{\alpha} e^{-\lambda' y}}{\Gamma(\alpha+1)}\, dy$$

$$+ e^{-\lambda' x_1}(\lambda' x_1)^{\alpha+1} \sum_{k=1}^{\infty} \frac{\lambda'^k C_k \Gamma(k)}{\Gamma(\alpha+1+k)} L_{k-1}^{(\alpha+1)}(\lambda' x,). \quad (29.13)$$

The notation L, below denotes an expansion of the form *(29.13)* in which the first i moments of the random variable associated with the leading term of the expansion (a gamma **r.v.**) have been equated to corresponding moments of $\chi'^2_\nu(\lambda)$; $L_n^{(\alpha)}(x)$ is the nth generalized Laguerre polynomial **(29.10)**, $x_1 = x + \theta$ and a, θ and λ' are chosen as follows:

For L_0, $\alpha + 1 = \nu/2, \theta = 0, \lambda' = \frac{1}{2}$.
For $L,$, $a + 1 = \nu/2, 0 = 0, A = (\alpha + 1)/\{n(1 + \delta^2)\}$.
For L_2, $a + 1 = \nu\lambda'(1 + \delta^2), \theta = 0, A = (1 + \delta^2)/\{2(1 + 2\delta^2)\}$.
For L_3, $a + 1 = 2\nu\lambda'^2(1 + 2\delta^2), \theta = 2n\lambda'(1 + 2\delta^2) - n(1 + \delta^2)$,
$A = \frac{1}{2}(1 + 2\delta^2)/(1 + 3\delta^2)$; and $\delta^2 = \lambda/\nu$.

The coefficients C_k are all calculated recursively and are rather complicated [see Gideon and Gurland (1977) for explicit expressions]. Comparisons among L_0–L_3 show that for large and small noncentrality A, the series L_0, L, are the best, though the series L_2 and L_3 produce good results in this range also. In choosing between the higher moment series L_2 and L_3 and a lower moment series such as L_0, the higher moment series usually give three- to five-place decimal accuracy with just a few terms (one to five) but may gain additional accuracy only slowly. The series L_0, however, usually produces less accuracy up to the first five to ten partial sums, but after that converges rapidly to the true probability.

Venables (1971) provided several alternative Laguerre expansions for $F(x; \nu, \lambda)$. The "most promising" was

$$F(x;\nu,\lambda) = \Gamma_\beta(\tfrac{1}{2}ax + b) - \gamma_{\beta+1}(\tfrac{1}{2}ax + b) \sum_{j=1}^{\infty} d_j L_{j-1}^{(\beta)}(\tfrac{1}{2}ax + b) \quad (29.14)$$

where

$$a = (\nu + 2\lambda)(\nu + 3\lambda)^{-1},$$

$$b = \tfrac{1}{2}\lambda^2(\nu + 2\lambda)(\nu + 3\lambda)^{-1} = \tfrac{1}{2}\lambda^2 a,$$

$$\beta = \tfrac{1}{2}(\nu + 2\lambda)^3(\nu + 3\lambda)^{-2},$$

$$\gamma_\alpha(x) = \{\Gamma(\alpha)\}^{-1}x^{\alpha-1}e^{-x}, \qquad x, \alpha > 0,$$

$$\Gamma_\alpha(x) = \int_0^x \gamma_\alpha(t)\, d \qquad \text{[as in Chapter 1, Eq. (1.85)]},$$

$$d_j = \frac{j!}{(\beta)_j}\sum_{i=0}^{j}(1 - a)^i L_i^{(\frac{\nu}{2}-1)}\left(\frac{\frac{1}{2}\lambda a}{1 - a}\right)L_{j-i}^{(\beta-3-\frac{\nu}{2})}(b),$$

with $(\beta)_j = \beta(\beta + 1)\cdots(\beta + j - 1)$. Convergence is fairly rapid if either ν or λ (or both) are large, provided that x is not too small, say, less than $\chi'^2_{\nu,0.01}(\lambda)$. [The simpler expression (29.2) also converges rapidly for large λ.]

Han (1975) gave the following formula, applicable when ν $(= 2s + 1)$ is odd:

$$F(x;2s + 1,\lambda) = F(x;1,\lambda) + \sum_{i=1}^{s}\sum_{j=1}^{i}\binom{i-1}{j-1}2^j F^{(j)}(x;1,\lambda), \quad (29.15)$$

where $F^{(j)}(\cdot)$ denotes $d^jF(\cdot)/d\lambda^j$. For $\nu = 3$ $(m = 1)$ we have

$$F(x;3,\lambda) = \Phi(\sqrt{\lambda} + \sqrt{x}) - \Phi(\sqrt{\lambda} - \sqrt{x}) + \lambda^{-1/2}\{\varphi(\sqrt{\lambda} + \sqrt{x}) - \varphi(\sqrt{\lambda} - \sqrt{x})\}, \quad (29.16)$$

where $\varphi(t) = \Phi'(t)$.

Chou et al. (1984) derived the representation

$$F(x;\nu,\lambda) = \frac{\sqrt{\pi}}{2^{\nu-1}\Gamma((\nu-1)/2)}\int_0^x y^{(\nu-3)/2}\varphi(\sqrt{y}) \times\{\Phi(\sqrt{x-y} - \sqrt{\lambda}) - \Phi(-\sqrt{x-y} - \sqrt{\lambda})\}\,dy, \quad (29.17)$$

and in particular,

$$F(x;1,\lambda) = \Phi(\sqrt{x} - \sqrt{\lambda}) - \Phi(-\sqrt{x} - \sqrt{\lambda}).$$

The pdf is

$$p(x;\nu,\lambda) = \frac{(2\pi)^{1/2}}{2^{(\nu-1)/2}\Gamma((\nu-1)/2)}\int_0^x y^{(\nu-3)/2}\varphi(\sqrt{y})\varphi(\sqrt{x-y} - \sqrt{\lambda})$$

$$- \varphi(-\sqrt{x-y} - \sqrt{\lambda})\frac{1}{2}(x-y)^{-1/2}\,dy. \qquad (29.18)$$

Integration by parts gives the simple formula (29.7).
Guenther (1964) obtained

$$p(x;\nu,\lambda)$$

$$= \frac{(2\pi)^{1/2}\varphi(\sqrt{\lambda})}{2^{(\nu-1)/2}\Gamma((\nu-1)/2)}x^{(\nu-2)/2}\varphi(\sqrt{x})\int_0^\pi \exp(\sqrt{x\lambda}\cos\theta)\sin^{n-2}\theta\,d\theta$$

$$= \pi\lambda^{-(\nu-2)/4}I_{(\nu-2)/2}(\sqrt{x\lambda})x^{(\nu-2)/4}\varphi(\sqrt{\lambda})\varphi(\sqrt{x}). \qquad (29.19)$$

Temme (1993) obtained the expressions

$$F(x;\nu,\lambda)$$

$$= \begin{cases} 1 + \left(\dfrac{x}{\lambda}\right)^{\nu/4}\left\{T_{(\nu-2)/2}(\sqrt{x\lambda},\omega) - T_{\nu/2}(\sqrt{x\lambda},\omega)\right\} & \text{for } x > \lambda, \\ \dfrac{1}{2}\left(\dfrac{x}{\lambda}\right)^{\nu/4}\left\{T_{(\nu-2)/2}(\sqrt{x\lambda},\omega) - T_{\nu/2}(\sqrt{x\lambda},\omega)\right\} & \text{for } x < \lambda, \end{cases} \qquad (29.20)$$

where

$$\omega = \tfrac{1}{2}(\sqrt{x} - \sqrt{\lambda})^2/\sqrt{x\lambda},$$

$$T_\mu(\sqrt{x\lambda},\omega) = \int_{\sqrt{x\lambda}}^\infty e^{-(\omega+1)t}I_\mu(t)\,dt$$

which are claimed to have some computational advantages.

Ennis and Johnson (1993) obtained the formula

$$F(x;\nu,\lambda) = \quad -\pi^{-1} \quad {}^{m} y^{-1}(1-y^2)^{-\nu/4}\exp\left\{-\tfrac{1}{2}\lambda y^2(1+y^2)^{-1}\right\}$$

$$\times \sin\left[\tfrac{1}{2}\left\{\nu \tan^{-1} y + \lambda y(1+y^2)^{-1}\right\} - yx\right] dy. \qquad (29.21)$$

For computational purposes this formula avoids the need to evaluate the sum of an infinite series, but it does require numerical integration over an infinite range. (The latter is, however, assisted by the fact that the integrand is of order $y^{-(\nu+4)/4}$ as y tends to infinity.)

Ruben (1974) noted the recurrence relation (for $\nu > 6$)

$$\lambda F(x;\nu,A) = (A-(\nu-4)\}F(x;\nu-2,A)$$

$$+\{x+(\nu-4)\}F(x;\nu-4,A) - xF(x;\nu-6,A). \qquad (29.22a)$$

[See also Cohen (1988) and Temme (1993)]. For $A = 0$, writing v + 2 in place of ν, (29.22a) reduces to the relation

$$(\nu-2)F(x;\nu,0) = (x+\nu-2)F(x;\nu-2,0) - xF(x;\nu-4,0) \qquad (29.22b)$$

among central χ^2 cdfs, as noted by Khamis (1965).

The following recurrence relations, among others, were obtained by Cohen *(1988):*

$$p(x;\nu,A) = p(x;\nu-2,A) + 2\frac{\partial p(x;\nu-2,A)}{\partial\lambda}. \qquad (29.23a)$$

$$p(x;\nu,A) = p(x;\nu+2,A) + 2\frac{\partial p(x;\nu+2,A)}{}. \qquad (29.23b)$$

$$p(x;\nu,\lambda) = \lambda^{-(\nu-2)/2}\exp\left\{-\frac{1}{2}(\lambda-1)(x-1)\right\}p(\lambda x;\nu;1). \qquad (29.23c)$$

The further relation

$$\frac{\partial F(x;\nu,A)}{\partial A} = \frac{1}{2}\{F(x;\nu+2,A) - F(x;\nu,\lambda)\} \qquad (29.23d)$$

is useful when interpolating with respect to A, while [cf. (28.7)]

$$\frac{\partial F(x;\nu,A)}{\partial x} = p(x;\nu,A) = \frac{1}{2}\{F(x;\nu-2,A) - F(x;\nu,A)\} \qquad (29.7)'$$

is useful for interpolating with respect to x. Note that (29.23d) and (29.7)′ together imply that

$$\frac{\partial F(x;\nu,A)}{\partial A} = -p(x;\nu+2,A). \qquad (29.23e)$$

[See Quenouille (1949), Guenther and Terragno (1964), Ruben (1974), and Schroder (1989).]

Ashour and Abdel-Samad (1990), based on Shea (1988), developed the following computational formula:

$$F(x;\nu,\lambda) = e^{-\lambda/2}p(x;\nu,0)\sum_{i=0}^{\infty}\frac{1}{i!}C_i\left(\frac{1}{4}\lambda,\frac{1}{2}\nu\right)\sum_{j=0}^{\infty}C_j\left(\frac{1}{2}x,\frac{1}{2}\nu+i\right), \qquad (29.24a)$$

where

$$C_i(a,b) = \frac{a}{b+i}C_{i-1}(a,b) \quad \text{and} \quad C_0(a,b) = 1.$$

For odd integer values of ν, they use the series *(29.3)* with

$$F(x;\nu+2j,0) = 2\{1-\Phi(\sqrt{x})\} + \left(\frac{2}{\pi}\right)^{1/2} e^{-x/2}\sum_{i=1}^{\frac{1}{2}(\nu-1)+j}\frac{x^{i-\frac{1}{2}}}{1\cdot 3\cdot 5\cdots(2i-1)} \qquad (29.24b)$$

[Abramowitz and Stegun (1964)].

Kallenberg (1990) has obtained bounds for the difference between cdf's of noncentral chi-squares with common degrees of freedom (ν) but different noncentrality parameters (A, A^*). If $A \le A^*$

$$0 < F(x;\nu,A) - F(x;\nu,A^*) \le (2\pi)^{-1/2}(\sqrt{\lambda^*} - \sqrt{\lambda})F(x;\nu-1,0). \qquad (29.25a)$$

[Note that $F(x; \text{v} - 1, 0)$ is the cdf of central $\chi^2_{\nu-1}$.] The lower limit (0) corresponds to the fact that $F(x; \text{v}, \text{A})$ is a decreasing function of A. A better lower bound, also obtained by Kallenberg, is provided by the following result: If

$$\liminf_{n \to 0} \max(\text{A}_{,}, \ \text{A*}_{,}) > 0 \qquad (29.25b)$$

and

$$\sqrt{\lambda^*} - \sqrt{\lambda} = O(1),$$

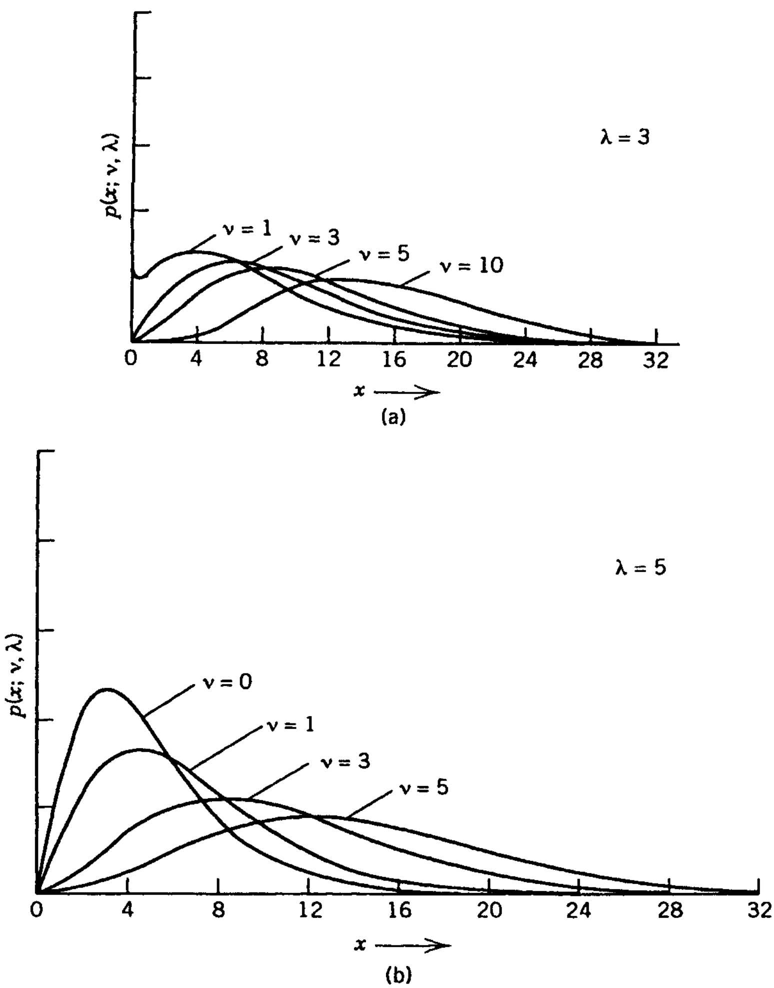
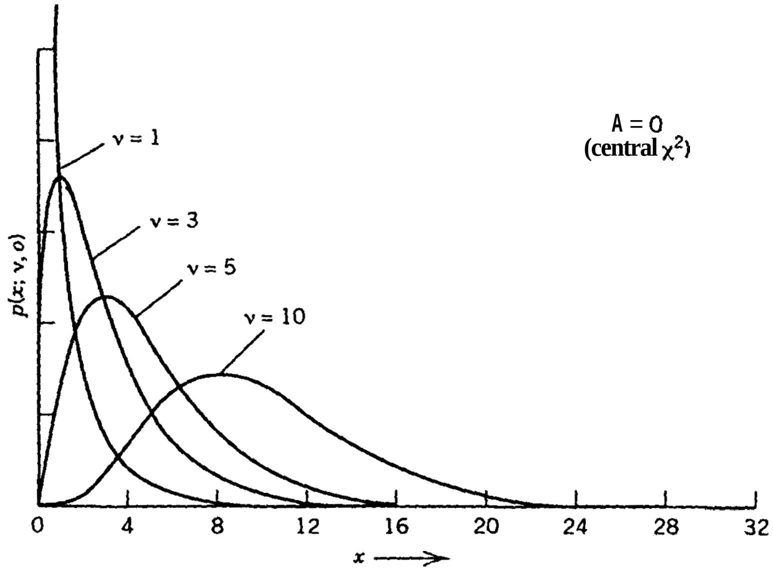


Figure 29.1*a–c* Probability density functions of noncentral and central chi-square.



(c)

Figure 29.1 *Continued*

then there is a $C(\nu)$ such that

$$\sup_{x>0}\left|F(x;\nu,\lambda_n) - F(x;\nu,\lambda_n^*)\right| \geq C(\nu)\left|\sqrt{\lambda_n^*} - \sqrt{\lambda_n}\right|.$$

Typical plots of $p(x;\nu,\lambda)$ were provided by Narula and Levy (1975). They are shown in Figures 29.1*a*, *b*. As a comparison Figure 29.1*c* exhibits $p(x;\nu,0)$, that is, pdf's of central χ^2-distributions. It is easy to see the increases in location parameters (mean, median, and mode) associated with increases in the noncentrality parameter, λ for given ν, (as also with increases in the degrees of freedom ν for given λ).

Siegel (1979) defined a $\chi_0'^2(\lambda)$ (noncentral χ^2 with zero degrees of freedom) variable as follows: Choose K from a Poisson distribution with mean $\frac{1}{2}\lambda$ so that $\Pr[K = k] = e^{-\lambda/2}(\frac{1}{2}\lambda)^k/k!$ $(k = 0, 1, \ldots)$. Then choose $Y_\lambda \sim \chi_{2K}^2$, a central chi-squared distribution. When $K = 0$, adopt the convention that the central χ_0^2 distribution is identically zero; this accounts for the discrete component of $\chi_0^2(\lambda)$. Thus $\chi_0^2(\lambda)$ is a mixture of the distributions $0, \chi_2^2, \chi_4^2, \ldots$ with Poisson weights. [See (29.5c).]

The cumulative distribution function of $Y_\lambda \sim \chi_0^2(\lambda)$ is

$$F(y;0,\lambda) = 1 - e^{-(\lambda+y)/2}\sum_{k=1}^{\infty}\frac{(\frac{1}{2}\lambda)^k}{k!}\sum_{j=0}^{k-1}\frac{(\frac{1}{2}y)^j}{j!}, \qquad (29.5c)'$$

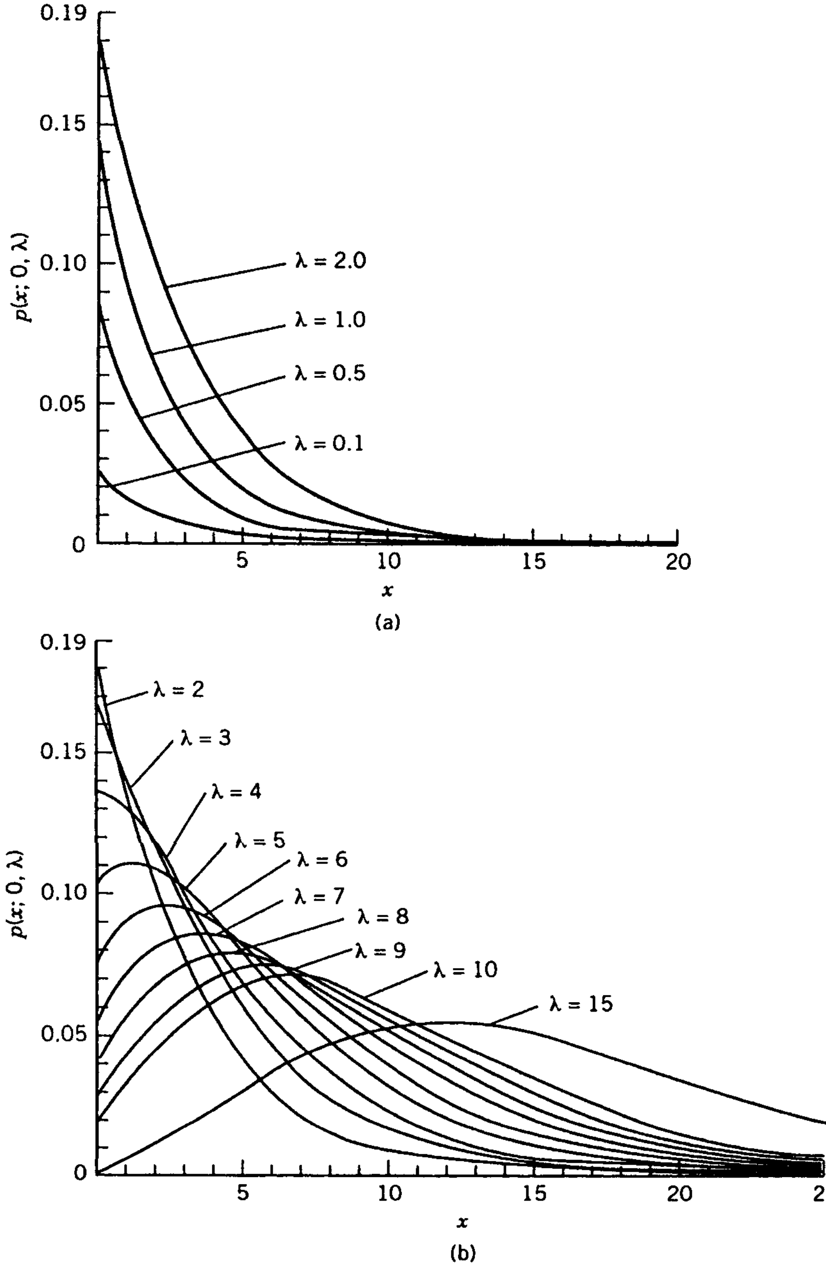


Figure 29.2*a*, b Improper density of the $\chi_0'^2(\lambda)$ distribution for some values of the noncentrality parameter λ, (a) for $\lambda \leq 2$, (b) for $\lambda \geq 2$.

where $y \geq 0$ and zero otherwise. These series converge quickly; hence this formula is convenient for computing. Figures **29.2*a*, *b***, taken from **Siegel (1979)**, show the (improper) probability density functions $p(x; 0, \lambda)$ of $\chi_0'^2(\lambda)$ for various values of λ. Clearly apparent are the reduction in area associated with the mass $e^{-\lambda/2}$ at zero, asymptotic normality when λ is Large, and asymptotic exponentiality χ_2^2 of the positive component when λ is small.
This distribution was used by **Siegel** (1979) to obtain critical values for a test of uniformity. Note that the purely noncentral part, Z_λ of $\chi_\nu'^2(\lambda)$, as defined by **Hjort** (1989) [see **(29.5b)]**, is a $\chi_0'^2(\lambda)$ random variable, as was mentioned above.

4 MOMENTS

From **(29.6)** the moment-generating function of $\chi_\nu'^2(\lambda)$ is

$$M(t;\nu,\lambda) = (1-2t)^{-\nu/2}\exp\left(\frac{\lambda t}{1-2t}\right). \tag{29.6$'$}$$

The cumulant-generating function is

$$K(t;\nu,\lambda) = \log M(t;\nu,\lambda) = -\tfrac{1}{2}\nu\log(1-2t) + \lambda t(1-2t)^{-1}. \tag{29.26}$$

Hence the *rth* cumulant is

$$\kappa_r = 2^{r-1}(r-1)!(\nu + r\lambda). \tag{29.27}$$

In particular

$$\begin{cases} \kappa_1 = \nu + \lambda = E\left[\chi'^2\right], \\ \kappa_2 = 2(\nu + 2\lambda) = \operatorname{var}(\chi'^2) = \left[\sigma(\chi'^2)\right]^2, \\ \kappa_3 = 8(\nu + 3\lambda) = \mu_3(\chi'^2), \\ \kappa_4 = 48(\nu + 4\lambda), \end{cases} \tag{29.28}$$

and hence

$$\mu_4(\chi'^2) = \kappa_4 + 3\kappa_2^2 = 48(\nu + 4\lambda) + 12(\nu + 2\lambda)^2. \tag{29.29}$$

From these formulas the values of the moment ratios can be calculated.

where y ≥ 0 and zero otherwise. These series converge quickly; hence this formula is convenient for computing. Figures **29.2*a*,** b, taken from Siegel **(1979),** show the **(improper)** probability density functions $p(x; 0, A)$ of $\chi_0^2(\lambda)$ for various values of A. Clearly apparent are the reduction in area associated with the mass $e^{-\lambda/2}$ at zero, asymptotic normality when A is large, and asymptotic exponentiaiity χ_2^2 of the positive component when A is small. This distribution was used by Siege1 **(1979)** to obtain critical values for a test of uniformity. Note that the purely noncentral part, Z_λ of $\chi_\nu'^2(\lambda)$, as defined by **Hjort (1989)** [see **(29.5b)],** is a $\chi_0'^2(\lambda)$ random variable, as was mentioned above.

4 MOMENTS

From (29.6) the moment-generating function of $\chi_\nu'^2(\lambda)$ is

$$M(t;\nu,\lambda) = (1-2t)^{-\nu/2}\exp\left(\frac{\lambda t}{1-2t}\right). \tag{29.6$'$}$$

The cumulant-generating function is

$$K(t;\nu,\lambda) = \log M(t;\nu,\lambda) = -\tfrac{1}{2}\nu\log(1-2t) + \lambda t(1-2t)^{-1}. \tag{29.26}$$

Hence the rth cumulant is

$$\kappa_r = 2^{r-1}(r-1)!(\nu + r\lambda). \tag{29.27}$$

In particular

$$\begin{cases} \kappa_1 = \nu + A = E[\chi'^2], \\ \kappa_2 = 2(\nu + 2\lambda) = \operatorname{var}(\chi'^2) = [\sigma(\chi'^2)]^2, \\ \kappa_3 = 8(\nu + 3\lambda) = \mu_3(\chi'^2), \\ \kappa_4 = 48(\nu + 4\lambda), \end{cases} \tag{29.28}$$

and hence

$$\mu_4(\chi'^2) = \kappa_4 + 3\kappa_2^2 = 48(\nu + 4A) + 12(\nu + 2\lambda)^2. \tag{29.29}$$

From these formulas the values of the moment ratios can be calculated.

These are

$$\alpha_3 = \sqrt{\beta_1} = \frac{\sqrt{8}(\nu + 3\lambda)}{(\nu + 2\lambda)^{3/2}},$$
$$\alpha_4 = \beta_2 = 3 + \frac{12(\nu + 4\lambda)}{(\nu + 2\lambda)^2}. \tag{29.30}$$

From these equations it follows that

$$\frac{\beta_2 - 3}{\beta_1} = \frac{3(\nu + 4\lambda)(\nu + 2\lambda)}{2(\nu + 3\lambda)^2} = \frac{3}{2}\left[1 - \frac{\lambda^2}{(\nu + 3\lambda)^2}\right],$$

whence

$$\frac{4}{3} \le \frac{\beta_2 - 3}{\beta_1} \le \frac{3}{2}. \tag{29.31}$$

The expressions for moments of $\chi_\nu'^2(\lambda)$ about zero are not so elegant as those for the central moments and cumulants [Park (1961)]. Sen (1989) has discussed the mean-median-mode inequality.

The following formula for the rth moment about zero was given (private communication) by D. W. Boyd:

$$\mu_r' = 2^r \Gamma\left(r + \frac{1}{2}\nu\right) \sum_{j=0}^{r} \binom{r}{j} \frac{(\lambda/2)^j}{\Gamma(j + \frac{1}{2}\nu)}.$$

The moment-generating function of $\frac{1}{2}\log[\chi'^2(\lambda)/\nu]$ was used by Bennett (1955) to evaluate the moments of this variable. It is evident that

$$E\left[\left(\frac{1}{2}\log\left[\chi'^2_{\nu+2j}(\lambda)/\nu\right]\right)^r\right] = 2^{-r}\sum_{j=0}^{\infty}\left[\frac{(\frac{1}{2}\lambda)^j}{j!}e^{-\lambda/2}\right]E\left[\left(\log\left[\chi^2_{\nu+2j}/\nu\right]\right)^r\right]$$

and that the values of $E[(\log[\chi^2_{\nu+2j}/\nu])^r]$ can be obtained from

$$\kappa_r(\log \chi_\nu^2) = \psi^{(r-1)}(\tfrac{1}{2}\nu) + \varepsilon_r \log 2,$$

with $\varepsilon_1 = \frac{1}{2}$, $\varepsilon_r = 0$ for r > 1. [See Eqs. (27.10) and (27.14)', Chapter 27.1

For integer ν we have the following expressions for reciprocal moments:

1. For $\nu > 2r$ and even,

$$E\left[\left(\chi_\nu'^2(\lambda)\right)^{-r}\right]$$

$$= \frac{(-1)^{r+(\nu/2)}}{(r-1)!}2^{-r}\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\frac{1}{2}\lambda\right)^{s+1-(\nu/2)}\Gamma\left(\frac{1}{2}\nu-s-1\right)$$

$$\times\left\{e^{-\lambda/2}-\sum_{t=0}^{(\nu/2)-s-2}\frac{\left(-\frac{1}{2}\lambda\right)^t}{t!}\right\}. \qquad (29.32a)$$

2. For $\nu > 2r$ and odd,

$$E\left[\left(\chi_\nu'^2(\lambda)\right)^{-r}\right]$$

$$= \frac{(-1)^{r+((\nu-1)/2)}}{(r-1)!}\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\frac{1}{2}\lambda\right)^{s+1-(\nu/2)}\Gamma\left(\frac{1}{2}\nu-s-1\right)$$

$$\times\left[\frac{2}{\sqrt{\pi}}D\left(\sqrt{\frac{\lambda}{2}}\right)-\sqrt{\frac{\lambda}{2}}\sum_{t=0}^{((\nu-5)/2)-s}\frac{\left(-\frac{1}{2}\lambda\right)^t}{\Gamma\left(t+\frac{3}{2}\right)}\right] \qquad (29.32b)$$

[Bock et al. (1984)]. (For $\nu \le 2r$ the rth reciprocal moment is infinite.)

Here

$$D(y) = \left\{\int_0^y \exp(t^2)\,dt\right\}e^{-y^2/2}$$

is the Dawson integral. This is a nonnegative function of y (for $y > 0$) with maximum value 0.541044... attained at $y = 0.924139\ldots$. For large y, $D(y) \sim \frac{1}{2}y^{-1}$.

An alternative expression, valid for all $\nu > 2r$, is

$$E\left[\left(\chi_\nu'^2(\lambda)\right)^{-r}\right] = 2^{-r}e^{-\lambda/2}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\lambda\right)^j}{j!}\frac{\Gamma\left(\frac{1}{2}(\nu+2j)-r\right)}{\Gamma\left(\frac{1}{2}(\nu+2j)\right)}. \qquad (29.32c)$$

Ullah (1976) noted that this can be expressed as

$$E\left[\left(\chi_\nu'^2(\lambda)\right)^{-r}\right] = 2^{-r}e^{-\lambda/2}\frac{\Gamma\left(\frac{\nu}{2}-r\right)}{\Gamma\left(\frac{\nu}{2}\right)}\,{}_1F_1\left(\tfrac{1}{2}\nu - r; \tfrac{1}{2}\nu; \tfrac{1}{2}\lambda\right), \quad (29.32d)$$

where ${}_1F_1(\cdot)$ is a confluent hypergeometric function (see Chapter 1). Yet another expression is

$$\begin{aligned} E\left[\left(\chi_\nu'^2(\lambda)\right)^{-r}\right] &= E_J\left[E\left[\left(\chi^2_{\nu+2J}\right)^{-r}\right]\right] \\ &= E_J\left[\{(\nu+2J-2)(\nu+2J-4)\cdots(\nu+2J-2r)\}^{-1}\right], \end{aligned} \quad (29.32e)$$

where J has a Poisson distribution with expected value $\frac{1}{2}\lambda$ [e.g., Egerton and Laycock (1982)].

5 PROPERTIES OF THE DISTRIBUTION

Reproductivity

From the definition it is clear that if $\chi_{\nu_1}'^2(\lambda_1)$ and $\chi_{\nu_2}'^2(\lambda_2)$ are independent, then the sum $[\chi_{\nu_1}'^2(\lambda_1) + \chi_{\nu_2}'^2(\lambda_2)]$ is distributed as $\chi_{\nu_1+\nu_2}'^2(\lambda_1 + \lambda_2)$. This may be described verbally by saying that the noncentral χ^2 distribution is reproductive under convolution and that the degrees of freedom, and also the noncentralities, are additive under convolution.

Characterization

If Y has a $\chi_\nu'^2(\lambda)$ distribution and $Y = Y_1 + Y_2 + \cdots + Y_\nu$, where the Y_j's are independent and identically distributed, then each Y_j has a $\chi_1'^2(\lambda/\nu)$ distribution. The special case $\nu = 2$ was studied by McNolty (1962). $F(x;\nu,\lambda)$ is of course an increasing function of x for $x > 0$. It is a decreasing function of ν and of λ. In fact for any fixed value of x,

$$\lim_{\nu\to\infty} F(x;\nu,\lambda) = \lim_{\lambda\to\infty} F(x;\nu,\lambda) = 0.$$

The distribution of the standardized variable

$$\frac{\chi_\nu'^2(\lambda) - (\nu+\lambda)}{[2(\nu+2\lambda)]^{1/2}}$$

tends to normality as $\nu \to \infty$, λ remaining fixed, or as $\lambda \to \infty$, ν remaining fixed.

Unimodality

The distribution of $\chi_\nu'^2(\lambda)$ is unimodal. The mode occurs at the intersection of the probability density functions of $\chi_\nu'^2(\lambda)$ and $\chi_{\nu-2}'^2(\lambda)$, that is, at the value x satisfying the equation

$$p(x;\nu,\lambda) = p(x;\nu-2,\lambda).$$

Completeness

The family of $\chi_\nu'^2(\lambda)$ distributions for a finite range A, < A < A, of values of A, is complete in the classical sense [**Marden (1982)**; Oosterhoff and **Schreiber (1987)**].

Monotonicily

As noted in Section 3, $F(x;\nu,$ A) is an increasing function of each of v and A (see Figures 29.1*b*, *c*) [**Ghosh (1973)**; Ruben (1974)].

6 ESTIMATION

The noncentral χ^2 distribution depends on two parameters: v, the degrees of freedom, and A, the noncentrality. If ν is known, then the maximum likelihood estimator $\hat{\lambda}$, of A, given values of n independent random variables $X_1, X_2, \cdots, X_n$ each having density function (29.4), must satisfy the equation

$$\sum_{i=1}^{n}\left[\frac{\sum_{j=0}^{\infty}e^{-\hat{\lambda}/2}\left[\left\{\left(\frac{1}{2}\hat{\lambda}\right)^{j-1}/(j-1)!\right\}-\left\{\left(\frac{1}{2}\hat{\lambda}\right)^{j}/j!\right\}\right]p(X_i;\nu+2j,0)}{\sum_{j=0}^{\infty}e^{-\hat{\lambda}/2}\left\{\left(\frac{1}{2}\hat{\lambda}\right)^{j}/j!\right\}p(X_i;\nu+2j,0)}\right]=0; \tag{29.33}$$

that is,

$$\sum_{i=1}^{n}\left[\frac{\sum_{j=0}^{\infty}\left\{\left(\frac{1}{2}\hat{\lambda}\right)^{j-1}/(j-1)!\right\}p(X_i;\nu+2j,0)}{\sum_{j=0}^{\infty}\left\{\left(\frac{1}{2}\hat{\lambda}\right)^{j}/j!\right\}p(X_i;\nu+2j,0)}\right]=n$$

if this equation has a positive root. This equation is usually not easy to solve. For the case v = **2** Meyer **(1967)** showed that the equation has a positive solution if $\overline{X} = n^{-1}\sum_{i=1}^{n}X_i > 2$; otherwise the maximum likelihood estimator takes the value zero. He furthermore shows that

$$\lim_{n\to\infty}\Pr[\overline{X} > 2] = 1.$$

Extending Meyer's result to values of ν exceeding **2,** Dwivedi and Pandey **(1975)** showed that the MLE of A (with ν known) is zero if $\overline{X}$ is not greater than ν and satisfies the equation

$$\sum_{i=1}^{n} h\left(\sqrt{2X_i\hat{\lambda}}\right)X_i = n \qquad \text{if } \bar{X} > \nu, \tag{29.34a}$$

where $h(z) = I_{\nu/2}(z)\{zI_{(\nu-2)/2}(z)\}^{-1}$. $I_\nu(z)$ is the modified Bessel function of order ν of purely imaginary argument; see Chapter 1 for an explicit expression.

For large n,

$$\hat{\lambda} \doteqdot \frac{1}{2}\left(n^{-1}\sum_{i=1}^{n} X_i^{1/2}\right)^2$$

[since $h(z) = z^{-1}$ for large z].

Anderson (1981a, b) considered maximum likelihood estimation of parameters a , A based on observations Y,, $Y_2, \cdots, Y_n$, each distributed as $\sigma^2\chi_\nu'^2(\lambda), \nu$ being known. (She actually considered values $\sqrt{Y_1}, \sqrt{Y_2}, \cdots, \sqrt{Y_n}$, but this did not affect the MLEs.) The MLE equations are (29.34a) with X, replaced by $Y_i/\hat{\sigma}^2$, together with

$$\hat{\sigma}^2 = (\nu n)^{-1}\sum_{i=1}^{n}\left(Y_i - \hat{\lambda}\right). \tag{29.34b}$$

Anderson (1981a) stated that these equations have unique solutions, provided that $\hat{\sigma}^2$ is not too small. The asymptotic variances and covariance of $\hat{\sigma}^2$ and $\hat{\mu} = \hat{\lambda}^{1/2}\hat{\sigma}$ (as $n \to \infty$) are

$$n\,\text{var}(\hat{\mu}) \doteqdot \Delta^{-1}\left(\tfrac{1}{2}\nu\lambda^{-1} - 1 - \text{A} + \theta\sigma^{-2}\right)\sigma^2, \tag{29.35a}$$

$$n\,\text{var}(\text{d}^2) = \Delta^{-1}(\theta\lambda^{-1}\sigma^{-2} - 1)\text{u}^4, \tag{29.35b}$$

$$n\,\text{cov}(\hat{\mu}, \hat{\sigma}^2) \doteqdot \Delta^{-1}\lambda^{1/2}(\theta\lambda^{-1}\sigma^{-2} - 1 - \lambda^{-1})\sigma^3, \tag{29.35c}$$

where

$$\Delta = (\theta\lambda^{-1}\sigma^{-2} - 1)\left(\tfrac{1}{2}\nu + \lambda^{-1}\right) - 1,$$

$$\theta = E\left[\frac{XI_{\nu/2}^2\left(\sqrt{X\lambda/\sigma}\right)}{I_{(\nu-2)/2}^2\left(\sqrt{X\lambda/\sigma}\right)}\right].$$

Anderson (1981b) provided bounds on the value of $\theta\lambda^{-1}\sigma^{-2}$, namely

$$1 + \lambda^{-1} - \frac{1}{2}\nu\lambda^{-2} + \frac{1}{4}\nu^2\lambda^{-3} \le \theta\lambda^{-1}\sigma^{-2} \le 1 + \frac{5}{4}\lambda^{-1} - \frac{20\nu - 13}{39}\lambda^{-2}. \tag{29.36a}$$

For large A a better lower bound, also given by Anderson, is

$$1 + \lambda^{-1} - \tfrac{1}{2}(\nu - 1)\lambda^{-2} + \tfrac{1}{4}(\nu - 1)(\nu - 2)\lambda^{-3}. \qquad (29.36b)$$

Also for large A,

$$n\,\mathrm{var}(\hat{\sigma}) \doteq \tfrac{1}{2}\{1 - 2(\nu - 1)\lambda^{-2} + \cdots)\,\mathrm{a}^{Z}, \qquad (29.37a)$$

$$n\,\mathrm{var}(\hat{\lambda}) \doteq 1 + \tfrac{1}{2}\lambda^{2}, \qquad (29.37b)$$

$$\mathrm{corr}(\hat{\lambda}, \hat{\sigma}) = (1 - \lambda^{-2})\{1 - 2(\nu - 2)\lambda^{-2}\}^{1/2}. \qquad (29.37c)$$

Note the high correlation between $\hat{\lambda}$ and $\hat{\sigma}$.

If a is known, an unbiased moment estimator of A is

$$\tilde{\lambda} = n^{-1}\sigma^{-2}\sum_{i=1}^{n} X_{i} - \nu. \qquad (29.38)$$

We have

$$n\,\mathrm{var}(\tilde{\lambda}) = 2\nu + 4\lambda, \qquad (29.39a)$$

while the **Cramér-Rao** lower bound for variance of unbiased estimators of $\tilde{\lambda}$ is

$$4(\theta\lambda^{-1} - 1)^{-1}n^{-1}, \qquad (29.39b)$$

where 8 is as defined in (29.35). The asymptotic relative efficiency (ARE) of $\tilde{\lambda}$ is

$$\begin{aligned} \mathrm{ARE}(\tilde{\lambda}) &= 4(\theta\lambda^{-1} - 1)^{-1}(2\nu + 4\lambda)^{-1} \\ &= 1 - \tfrac{1}{2}\lambda^{-2} + \tfrac{1}{4}(2\nu - 3)\lambda^{-6} + \cdots \qquad \text{for large A.} \end{aligned} \qquad (29.40)$$

When $\nu = 1$, results for estimating parameters of folded normal distributions (Chapter 13, Section 7.3) are applicable because $\chi_1'(\lambda)\ [= +\sqrt{\chi_1'^2(\lambda)}\,]$ has the same distribution as $|U + \sqrt{\lambda}\,|$, where U has a unit normal distribution. The general folded normal distribution [Leone, Nelson, and Nottingham (1961)] is the distribution of $|U\sigma + \xi|$ and has density function

$$p(x) = \sqrt{\frac{2}{\pi}}\,\sigma^{-1}\left[\cosh(\xi x\sigma^{-2})\right]\exp\left[-\frac{1}{2}(x^{2} + \xi^{2})\sigma^{-2}\right], \qquad 0 < x. \qquad (29.41)$$

[Evidently $|U\sigma + \xi|$ has the same distribution as $\sigma\chi_1'(\xi^2/\sigma^2)$.] See Section 7.3 of Chapter 13 for related details.

The first and second moments about zero of distribution (29.41) are

$$\mu_f = \left(\sqrt{\frac{2}{\pi}}\right)\sigma \exp\left(-\frac{1}{2}\xi^2\sigma^{-2}\right) + \xi\left[1 - 2\Phi\left(-\frac{\xi}{\sigma}\right)\right] \tag{29.42a}$$

and

$$\mu_f^2 + \sigma_f^2 = \xi^2 + \sigma^2, \tag{29.42b}$$

respectively. Leone, Nelson, and Nottingham (1961) gave tables of expected value (μ_f) and standard deviation (σ_f) for

$$\mu_f/\sigma_f = 1.33(0.01)1.50(0.02)1.70(0.05)2.40(0.1)3.0.$$

[Note that the least possible value of μ_f/σ_f is $(\frac{1}{2}\pi - 1)^{-1/2} = 1.3237$.] Leone, Nelson, and Nottingham (1961) also gave values of the cumulative distribution function to four decimal places for $\xi/\sigma = 1.4(0.1)3.0$ and arguments at intervals of 0.01. Some values of the moment ratios (β_1 and β_2) were given by Elandt (1961).

The parameters ξ and σ can be estimated by equating sample first and second moments to (29.42a), (29.42b), respectively. The tables of Leone, Nelson, and Nottingham (1961) facilitate solution. Simple explicit solutions are obtained by using second and fourth moments. Here $\theta = \mu_f/\sigma_f$ is estimated by $\tilde{\theta}$, the solution of the equation

$$\frac{\text{Sample 4th moment}}{(\text{Sample 2nd moment})^2} = \frac{3 + 6\tilde{\theta}^2 + \tilde{\theta}^4}{(1 + \tilde{\theta}^2)^2}. \tag{29.43}$$

Elandt (1961) obtained expansions to terms of order n^{-3} for the variances of the estimators of 8 by the two methods. There appears to be little difference for 8 less than 0.75; the method using first and second sample moments is about 40% more efficient when $\theta = 3$.

The maximum likelihood equations for estimators $\hat{\xi}$, 6 of ξ and σ can be expressed in the form

$$\hat{\xi}^2 + \hat{\sigma}^2 = n^{-1}\sum_{j=1}^{n} X_j^2, \tag{29.44a}$$

$$\hat{\xi} = n^{-1}\sum_{j=1}^{n} X_j \tanh\left(\frac{\hat{\xi}X_j}{\hat{\sigma}^2}\right). \tag{29.44b}$$

Johnson (1962) obtained asymptotic formulas for the variances of $\hat{\theta}$ ($= \hat{\xi}/\hat{\sigma}$) and $\hat{\sigma}$. They are rather complicated, but for large values of θ,

$$n\,\text{var}(\hat{\theta}) = 1 + \tfrac{1}{2}\theta^2,$$

$$\text{n var}(\hat{\sigma}) \doteqdot \tfrac{1}{2},$$

$$\text{corr}(\hat{\theta},\hat{\sigma}) = -8(2+\theta^2)^{-1/2}. \qquad (29.44c)$$

Relative to the maximum likelihood estimators, the efficiency of estimation of θ from first- and second-sample moments is about 95% when 8 = 1, and increases with θ. For small θ the efficiency is low. If a is known [e.g., if we have a $\chi_1'(\lambda)$ distribution and wish to estimate A], the maximum likelihood equation for $\hat{\xi}$ is (29.44b) with $\hat{\sigma}$ replaced by a.

We now consider situations in which $n = 1$ so that only a single observed value X, say, of a $\chi_\nu'^2(\lambda)$ variable is available. If ν is known, a natural estimator of A is

$$\tilde{\lambda} = X - \nu, \qquad (29.45a)$$

which is a moment estimator that can be obtained by equating observed to expected values. It is also a uniform minimum variance unbiased estimator. However, there are (biased) estimators which have a smaller mean square error (MSE).

The simplest is

$$(X-\nu)_+ = \begin{cases} X-\nu & \text{if } X \geq \nu, \\ 0 & \text{if } X \leq \nu. \end{cases} \qquad (29.45b)$$

Perlman and Rasmussen (1975) considered modified forms of $\tilde{\lambda}$, namely

$$X - \nu + \frac{b}{X}, \qquad \nu > 5,\ 0 < b < 4(\nu - 4). \qquad (29.45c)$$

Neff and Strawderman (1976) carried the analysis further, obtaining estimators

$$X - \nu + \frac{b}{X^a}, \qquad \nu > 1,\ 0 < b < \frac{2^{a+2}\Gamma(\frac{1}{2}\nu - a)}{\Gamma(\frac{1}{2}\nu - 2a)}, \qquad (29.45d)$$

$$X - \nu + \frac{r(X)}{X^a}, \qquad \nu > 8,\ r(X) \text{ monotonic increasing}, \qquad (29.45e)$$

$$E\left[\{r(X)\}^2\right] \leq \frac{4(\nu-4)(\nu-6)(\nu+8)}{\nu(\nu+2)},$$

$$X - \nu + \frac{b}{X+c}, \qquad \nu > 4,\ 0 < b \leq 4(\nu-4)\left\{1 - \left(\frac{c}{\nu+4}\right)^2\right\},$$

$$0 < c < \text{v} - 4. \qquad (29.45f)$$

These estimators all have uniformly smaller mean square errors than $\tilde{\lambda}$. Note that (29.45f) is bounded as X → 0, which is not true of (29.45c)–(29.45e).

Kubokawa, Robert, and Saleh (1993) have shown that the estimators

$$\tilde{\lambda}_1(X) = X - \nu + e^{-\nu/2}\left\{\sum_{j=0}^{\infty}\left(-\frac{1}{2}\nu\right)^j \{j!(\nu+2j)\}^{-1}\right\} \quad (29.45g)$$

and

$$\tilde{\lambda}_2(X) = \begin{cases} X - \nu & \text{if } X \ge \nu + 2, \\ 2(\nu+2)^{-1}X & \text{if } X \le \nu + 2, \end{cases} \quad (29.45h)$$

also have mean square error uniformly less than that of $\tilde{\lambda}$ and are bounded as X → 0.

Alam and Saxena (1982) defined a noncentral gamma distribution with parameters α, θ by the pdf

$$p_Y(y;\alpha,\theta) = y^{\alpha-1}e^{-\theta-y}\sum_{j=0}^{\infty}\frac{(\theta y)^j}{j!\Gamma(\alpha+j)} = e^{-\theta-y}\left(\frac{y}{\theta}\right)^{(\alpha-1)/2} I_{\alpha-1}(2\sqrt{\theta x}), \quad (29.46)$$

where

$$I_p(x) = \left(\frac{1}{2}x\right)^p \sum_{k=0}^{\infty}\frac{(x^2/4)^k}{k!\Gamma(p+k+1)}.$$

This is the distribution of $2\chi_\nu'^2(\lambda)$ with

$$\nu = 2\alpha \quad \text{and} \quad A = 20.$$

Alam and Saxena compared the MSEs of the maximum likelihood and moment estimators of θ. The MSE of the moment estimator is somewhat less than that of the maximum likelihood estimators. See Table 29.1 which shows the ratios of the MSEs (in terms of our original A and ν.) Venables (1975) suggested a novel approach to determining confidence limits for A which provides useful results if the observed value is large compared with the relevant percentage points of the (central) χ_ν^2-distribution.

He noted that the complement of the cdf of $\chi_\nu'^2(\lambda)$ can be written [cf (29.4)]

$$1 - F(x;\nu,\lambda) = 1 - F(x;\nu,0) + p(x;\nu,0)\sum_{j=1}^{\infty}\frac{(x/2)^j}{(\nu/2)^{[j]}}F(\lambda;2j,0). \quad (29.47)$$

Table 29.1 Mean-square-errorratios, MSE($\hat{\lambda}$) / MSE($(X - \nu)_+$)

$\nu \backslash \lambda$	1	2	10	20
1	1.41	1.28	1.04	1.01
2	1.35	1.26	1.04	1.02
10	1.16	1.15	1.06	1.03
20	1.10	1.10	1.06	1.03

Source: Alam and Saxena (1982).

This can be regarded as a "confidence distribution" (fiducial distribution) for λ, which is a mixture of distributions with $\Pr[\lambda = 0] = 1 - F(x;\nu,0)$ and $\{F(\lambda; 2j, 0)\}$ with weights $\{(\frac{1}{2}x)^j/(\frac{1}{2}\nu)^{[j]}\}$ $(j = 1,2,\dots)$.

In principle, one could find $100(1-\alpha)\%$ confidence limits for λ as the lower and upper $50\alpha\%$ points of (29.47). Venables proposed the following approximate method. The corresponding moment-generating function is

$$E^*[e^{t\lambda}] = 1 - F(x;\nu,0) + \exp\left\{\frac{xt}{1-2t}\right\}(1-2t)^{(\nu/2)-1}F\big(x(1-2t)^{-1};\nu,0\big). \quad (29.48)$$

If x is sufficiently large, then

$$F(x;\nu,0) \doteq F\big(x(1-2t)^{-1};\nu,0\big) = 1,$$

and the moment-generating function is approximately

$$E^*[e^{t\lambda}] \doteq (1-2t)^{(\nu/2)-1}\exp\left\{\frac{xt}{1-2t}\right\}. \quad (29.49)$$

The corresponding approximate cumulants are

$$\kappa_r^* = 2^{r-1}(r-1)!(rx - \nu + 2). \quad (29.50)$$

In particular

$$\kappa_1^* = x - \nu + 2, \quad (29.51a)$$

$$\kappa_2^* = 2(2x - \nu + 2). \quad (29.51b)$$

Venables gave a Cornish-Fisher type of expansion for the $100\alpha\%$ point of the confidence distribution in terms of the corresponding unit normal deviate u„ when $\Phi(u_\alpha) = a$. Part of this is

$$\hat{\lambda}_\alpha = X - \nu + 2 + (u_\alpha^2 - 1) + u_\alpha \sigma - u_\alpha \sigma^{-1} + \tfrac{2}{3}(\nu - 1)(u_\alpha^2 - 1)\sigma^{-2}$$
$$+ \{\tfrac{1}{6}(4\nu - 7)u_\alpha - \tfrac{2}{3}(\nu - 1)u_\alpha^3\}\sigma^{-3}$$
$$+ \tfrac{4}{15}(\nu - 1)(3u_\alpha^4 + 2u_\alpha^2 - 11)\sigma^{-4} + O(\sigma^{-5}). \tag{29.52}$$

DeWaal (1974) considered Bayesian estimation of A, minimizing expected (quadratic) loss, based on a "noninformative" prior distribution and reported the astonishing result that $(X + \nu)$ is the indicated estimator! (This estimator can never be less than v, and it always has bias ν.)

If one assumes a prior of gamma form

$$g(\lambda) = \frac{c^p \lambda^{p-1} e^{-c\lambda}}{\Gamma(p)}, \qquad c, p, \lambda > 0,$$

the Bayes estimator of A is

$$l(X) = (1 + c)^{-1} p + (1 + c)^{-2} X, \tag{29.53}$$

with mean square error

$$(1 + c)^{-r}\{p + 2\lambda + (2 + c)^2 (p - c\lambda)^2\}. \tag{29.54}$$

As c $\to$ 0, $l(X)$ approaches X + p.

7 TABLES AND COMPUTER ALGORITHMS

There has been a large number of studies on the computation of the noncentral chi squared cdf, $F(x; \text{v}, \lambda)$ in the last 20 years, including some duplications. For calculation of $F(x; \nu$, A), the simple series expansion (29.3) is convenient for small values of A. The error committed by terminating c at $(s + 1)$-th term is negative and must be less than

$$e^{-\lambda/2} \sum_{j=s+1}^{\infty} \frac{(\frac{1}{2}\lambda)^j}{j!}$$

in absolute value (for any ν or x). This in turn cannot exceed

$$1 - \Phi\left(\frac{s + 1 - (\lambda/2)}{\sqrt{(\lambda/2)}}\right) \qquad \text{[see Chapter 4, Eq. (4.49)]}$$

To ensure that the absolute error is less than $100\alpha\%$, we take $(s + 1 - \frac{1}{2}\lambda)/\sqrt{\frac{1}{2}\lambda} > u_\alpha$, where $\Phi(u_\alpha) = 1 - \alpha$. Here are a few typical values of min $s = 1 + [u_\alpha\sqrt{\frac{1}{2}\lambda} + \frac{1}{2}\lambda - 1]$:

	A =	2	8	32
a	min s =	$1 + [u_\alpha]$	$1 + [2u_\alpha + 3]$	$1 + [4u_\alpha + 15]$
0.001		4	10	28
0.0001		4	11	30
0.00001		5	12	33

([a] denotes integer part of a.) These values provide minimum guaranteed accuracy for all x and all v. Much better accuracy will be obtained for small ν and small x. Guenther (1975) found, for example, that using the series (29.3) for x = 4.60517, $\nu = 2$, and A = 2.2, he obtained five decimal place accuracy.

The most extensive tables of the noncentral χ^2-distribution are those of Haynam, Govindarajulu, and Leone (1973). These tables are especially intended to facilitate calculations involving the power of various χ^2 tests. Using $\chi^2_{\nu,1-\alpha}$ to denote the upper $100\alpha\%$ point of the central χ^2-distribution with ν degrees of freedom and

$$\beta(\nu, \lambda, \alpha) = \Pr\left[\chi'^2_\nu(\lambda) > \chi^2_{\nu,1-\alpha}\right] \tag{29.55}$$

to denote power with respect to noncentrality A, the values tabulated are the following:

Table 1. β to Four decimal places,

$$\alpha = 0.001, 0.005, 0.01, 0.025, 0.05, 0.1,$$

$$\lambda = 0(0.1)1.0(0.2)3.0(0.5)5(1)40(2)50(5)100,$$

$$\nu = 1(1)30(2)50(5)100.$$

Table 2. A to three decimal places for the same values of a and v as in Table 1,

$$1 - \beta = 0.1(0.02)0.7(0.01)0.99.$$

Table 3. v to three decimal places for the same values of a, A and β as in Tables 1 and 2.

The first tables (apart from special calculations) relating to the noncentral χ^2-distribution were compiled by Fix (1949). These tables give λ to three decimal places for

$$\alpha = 0.01, 0.05,$$
$$1 - \beta = 0.1(0.1)0.9,$$
$$\nu = 6(1)20(2)40(5)60(10)100.$$

This table is also reproduced in the Bol'shev-Smirnov tables (1965). A similar table is included in Owen's tables (1962).

Bark et al. (1964) gave tables of $\Pr[\chi_2'^2(\omega^2) \; 2 \; u^2] = Q(u, \omega)$ to six decimal places for $\omega = 0(0.02)3.00$ and $u = 0(0.02)$ until $Q(u, \omega)$ is less than 0.0000005. For cases when $\omega > 3$ and $u \leq 3$, they suggested using the formula

$$Q(u, \omega) = 1 - Q(\omega, u) + Q(\omega - u, 0)e^{-u\omega}I_0(u\omega), \quad (29.56a)$$

and gave tables of the function $e^{-x}I_0(x)$. For $u \; 2 \; \omega > 3$ the formula suggested is

$$Q(u, \omega) = q - R(q, \varepsilon), \quad (29.56b)$$

where q $= 1 - \Phi(u - \omega - (2\omega)^{-1})$, $\varepsilon = (1 + \omega^2)^{-1}$ and $R(q, \varepsilon)$ is also given in these tables. Using (29.56a), (29.56b) can also be used for $\omega > u > 3$.

Johnson (1968) gave tables of percentile points: values $x(\nu, A, \alpha)$ such that $\Pr[\chi_\nu'^2(\lambda) > x(\nu, A, a)] = a -$ to four significant figures for $\sqrt{\lambda} = 0.2(0.2)6.0$; $\nu = 1(1)12, 15, 20$; a $= 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975, 0.999$. Tables of $\sqrt{x(\nu, \lambda, \alpha)}$ to four significant figures for the same values of $\sqrt{\lambda}$ and ν, but only for a $= 0.01$, 0.025, 0.05, 0.95, 0.975, and 0.99 were given by Johnson and Pearson (1969).

A computer program for calculating $p(x; \nu, \lambda)$ and $F(x; \nu, \lambda)$ was published by Bargmann and Ghosh (1964) [also by Robertson (1969)]. They used formulas (29.2) and (29.4) and provided parameters that are in the range 10^{-8} to 10^{+8}, obtaining accuracy to five significant figures. More detailed tables for the cases $\nu = 2, 3$, calculated in connection with "coverage" problems, are described in Section 9.

Narula and Desu (1981) have developed a rapid algorithm for computing $F(x; \nu, A)$ from (29.13). The algorithm, written in FORTRAN 66, uses Lau's (1980) algorithm for the incomplete gamma function and Pike and Hill's (1966) algorithm for calculating the log gamma function.

Wiener (1975) provided a simple computer program (LAMBDA) written in FORTRAN, for use in calculating the power of a test of the hypothesis (H_0) that X has a χ_ν^2 distribution against alternatives that it has a $\chi_\nu'^2(\lambda)$ distribution (A $\neq 0$). If the critical region (leading to rejection of H_0) is

$$X > \chi^2_{\nu, 1-\alpha}$$

in absolute value (for any ν or x). This in turn cannot exceed

$$1 - \Phi\left(\frac{s + 1 - (\lambda/2)}{\sqrt{(\lambda/2)}}\right) \qquad \text{[see Chapter 4, Eq. (4.49)]}.$$

To ensure that the absolute error is less than $100\alpha\%$, we take $(s + 1 - \frac{1}{2}\lambda)/\sqrt{\frac{1}{2}\lambda} > u_\alpha$, where $\Phi(u_\alpha) = 1 - \alpha$. Here are a few typical values of min $s = 1 + [u_\alpha\sqrt{\frac{1}{2}\lambda} + \frac{1}{2}\lambda - 1]$:

α	$\lambda =$ min $s =$	2 $1 + [u_\alpha]$	8 $1 + [2u_\alpha + 3]$	32 $1 + [4u_\alpha + 15]$
0.001		4	10	28
0.0001		4	11	30
0.00001		5	12	33

($[a]$ denotes integer part of a.) These values provide minimum ***guaranteed*** accuracy for all x and all ν. Much better accuracy will be obtained for small ν and small x. Guenther (1975) found, for example, that using the series (29.3) for $x = 4.60517$, $\nu = 2$, and $\lambda = 2.2$, he obtained five decimal place accuracy.

The most extensive tables of the noncentral χ^2-distribution are those of Haynam, Govindarajulu, and Leone (1973). These tables are especially intended to facilitate calculations involving the power of various χ^2 tests. Using $\chi^2_{\nu,1-\alpha}$ to denote the upper $100\alpha\%$ point of the *central* χ^2-distribution with ν degrees of freedom and

$$\beta(\nu, \lambda, \alpha) = \Pr\left[\chi'^2_\nu(\lambda) > \chi^2_{\nu,1-\alpha}\right] \tag{29.55}$$

to denote power with respect to noncentrality λ, the values tabulated are the following:

Table 1. β to four decimal places,

$$\alpha = 0.001, 0.005, 0.01, 0.025, 0.05, 0.1,$$
$$\lambda = 0(0.1)1.0(0.2)3.0(0.5)5(1)40(2)50(5)100,$$
$$\nu = 1(1)30(2)50(5)100.$$

Table 2. λ to three decimal places for the same values of α and ν as in Table 1,

$$1 - \beta = 0.1(0.02)0.7(0.01)0.99.$$

Table 3. ν to three decimal places for the same values of α, λ and β as in Tables 1 and 2.

u_α, when $\Phi(u_\alpha) = a$. Part of this is

$$\hat{\lambda}_\alpha = X - \nu + 2 + (u_\alpha^2 - 1) + u_\alpha\sigma - u_\alpha\sigma^{-1} + \tfrac{2}{3}(\nu - 1)(u_\alpha^2 - 1)\sigma^{-2}$$
$$+ \{\tfrac{1}{6}(4\nu - 7)u_\alpha - \tfrac{2}{3}(\nu - 1)u_\alpha^3\}\sigma^{-3}$$
$$+ \tfrac{4}{15}(\nu - 1)(3u_\alpha^4 + 2u_\alpha^2 - 11)\sigma^{-4} + O(\sigma^{-5}). \tag{29.52}$$

DeWaal (1974) considered Bayesian estimation of A, minimizing expected (quadratic) loss, based on a "**noninformative**" prior distribution and reported the astonishing result that $(X + \nu)$ is the indicated estimator! (This estimator can never be less than v, and it always has bias ν.)

If one assumes a prior of gamma form

$$g(\lambda) = \frac{c^p\lambda^{p-1}e^{-c\lambda}}{\Gamma(p)}, \qquad c, p, \lambda > 0,$$

the Bayes estimator of A is

$$l(X) = (1 + c)^{-1}p + (1 + c)^{-2}X, \tag{29.53}$$

with mean square error

$$(1 + c)^{-r}\{p + 2A + (2 + c)^2(p - c\lambda)^2\}. \tag{29.54}$$

As c $\to$ 0, $l(X)$ approaches X + p.

7 TABLES AND COMPUTER ALGORITHMS

There has been a large number of studies on the computation of the noncentral chi squared cdf, $F(x;\nu$, A) in the last 20 years, including some duplications. For calculation of $F(x;$ v, A), the simple series expansion **(29.3)** is convenient for small values of A. The error committed by terminating c at (s + **1)-th** term is negative and must be less than

$$e^{-\lambda/2}\sum_{j=s+1}^{\infty}\frac{(\frac{1}{2}\lambda)^j}{j!}$$

in absolute value (for any ν or x). This in turn cannot exceed

$$1 - \Phi\left(\frac{s + 1 - (\lambda/2)}{\sqrt{(\lambda/2)}}\right) \qquad \text{[see Chapter 4, Eq. (4.49)]}$$

To ensure that the absolute error is less than $100\alpha\%$, we take $(s + 1 - \frac{1}{2}\lambda)/\sqrt{\frac{1}{2}\lambda} > u_\alpha$, where $\Phi(u_\alpha) = 1 - \alpha$. Here are a few typical values of min $s = 1 + [u_\alpha\sqrt{\frac{1}{2}\lambda} + \frac{1}{2}\lambda - 1]$:

	A =	2	8	32
a	min s =	$1 + [u,]$	$1 + [2u_\alpha + 3]$	$1 + [4u_\alpha + 15]$
0.001		4	10	28
0.0001		4	11	30
0.00001		5	12	33

($[a]$denotes integer part of a.) These values provide minimum ***guaranteed*** accuracy for all x and all ν. Much better accuracy will be obtained for small ν and small x. Guenther (1975) found, for example, that using the series (29.3)for $x = 4.60517$, $\nu = 2$, and $A = 2.2$, he obtained five decimal place accuracy.

The most extensive tables of the noncentral χ^2-distribution are those of Haynam, Govindarajulu, and Leone (1973). These tables are especially intended to facilitate calculations involving the power of various χ^2 fests. Using $\chi^2_{\nu,1-\alpha}$ to denote the upper $100\alpha\%$ point of the ***central*** χ^2-distribution with ν degrees of freedom and

$$\beta(\nu, \lambda, \alpha) = \Pr\left[\chi'^2_\nu(\lambda) > \chi^2_{\nu,1-\alpha}\right] \tag{29.55}$$

to denote power with respect to noncentrality A, the values tabulated are the following:

Table 1. β to four decimal places,

$$\alpha = 0.001, 0.005, 0.01, 0.025, 0.05, 0.1,$$
$$\lambda = 0(0.1)1.0(0.2)3.0(0.5)5(1)40(2)50(5)100,$$
$$\nu = 1(1)30(2)50(5)100.$$

Table 2. A to three decimal places for the same values of a and ν as in Table 1,

$$1 - \beta = 0.1(0.02)0.7(0.01)0.99.$$

Table 3. ν to three decimal places for the same values of a, A and β as in Tables 1 and 2.

The first tables (apart from special calculations) relating to the noncentral χ^2-distribution were compiled by Fix (1949). These tables give λ to three decimal places for

$$\alpha = 0.01, 0.05,$$
$$1 - \beta = 0.1(0.1)0.9,$$
$$\nu = 6(1)20(2)40(5)60(10)100.$$

This table is also reproduced in the Bol'shev-Smirnov tables (1965). A similar table is included in Owen's tables (1962).

Bark et al. (1964) gave tables of $\Pr[\chi_2'^2(\omega^2) \geq u^2] = Q(u, \omega)$ to six decimal places for $\omega = 0(0.02)3.00$ and $u = 0(0.02)$ until $Q(u, \omega)$ is less than 0.0000005. For cases when $\omega > 3$ and $u \leq 3$, they suggested using the formula

$$Q(u, \omega) = 1 - Q(\omega, u) + Q(\omega - u, 0)e^{-u\omega}I_0(u\omega), \qquad (29.56a)$$

and gave tables of the function $e^{-x}I_0(x)$. For u $\geq \omega > 3$ the formula suggested is

$$Q(u, \omega) = q - R(q, \varepsilon), \qquad (29.56b)$$

where $q = 1 - \Phi(u - \omega - (2\omega)^{-1})$, $\varepsilon = (1 + \omega^2)^{-1}$ and $R(q, \varepsilon)$ is also given in these tables. Using (29.56a), (29.56b) can also be used for $\omega > u > 3$.

Johnson (1968) gave tables of percentile points: values $x(\nu$, A, a) such that $\Pr[\chi_\nu'^2(\lambda) > x(\nu$, A, a)]= a − to four significant figures for $\sqrt{\lambda} = 0.2(0.2)6.0$; $\nu = 1(1)12, 15, 20$; a = 0.001, 0.0025, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995, 0.9975, 0.999. Tables of $\sqrt{x(\nu, \lambda, \alpha)}$ to four significant figures for the same values of $\sqrt{\lambda}$ and v, but only for a = 0.01, 0.025, 0.05, 0.95, 0.975, and 0.99 were given by Johnson and Pearson (1969).

A computer program for calculating $p(x; \nu, \lambda)$ and $F(x;$ v, $\lambda)$ was published by Bargmann and Ghosh (1964) [also by Robertson (1969)]. They used formulas (29.2) and (29.4) and provided parameters that are in the range 10^{-8} to 10^{+8}, obtaining accuracy to five significant figures. More detailed tables for the cases $\nu = 2, 3$, calculated in connection with "coverage" problems, are described in Section 9.

Narula and Desu (1981) have developed a rapid algorithm for computing $F(x; \nu$, A) from (29.13). The algorithm, written in FORTRAN 66, uses Lau's (1980) algorithm for the incomplete gamma function and Pike and Hill's (1966) algorithm for calculating the log gamma function.

Wiener (1975) provided a simple computer program (LAMBDA) written in FORTRAN, for use in calculating the power of a test of the hypothesis (H_0) that X has a χ_ν^2 distribution against alternatives that it has a $\chi_\nu'^2(\lambda)$ distribution ($\lambda \neq 0$). If the critical region (leading to rejection of H_0) is

$$X > \chi_{\nu, 1-\alpha}^2$$

(so that the significant level of the test is α), the power with respect to an alternative hypothesis (H_1) specifying $\lambda = \lambda_1$ is

$$\beta(\lambda_1|\alpha) = \Pr[X > \chi^2_{\nu,1-\alpha}] = \Pr[\chi'^2_\nu(\lambda) > \chi^2_{\nu,1-\alpha}] = F(\chi^2_{\nu,1-\alpha};\nu,\lambda_1), \tag{29.57}$$

which is computed by the program. Wiener (1975) also gives values of λ satisfying

$$\beta(\lambda|\alpha) = \beta \tag{29.58}$$

$$\text{for}\quad \nu = 1(1)30(2)50(5)100;$$
$$\alpha = 0.001, 0.005, 0.01, 0.025, 0.05; 0.1;$$
$$\beta = 0.01(0.01)0.30(0.20)0.90,$$

tabulated in the same form as in Haynam, Govindarajulu, and Leone (1970).

Posten (1989) provided another recursive algorithm for evaluating $F(x;\nu,\lambda)$ in terms of a single central $F(x;\nu,0)$, including a modification to deal with technical difficulties when λ is large, and using a recursive relation to evaluate $F(x;\nu+2j,0)$. For evaluation of the central $F(x;\nu+2j,0)$ values, Posten recommends using a continued fraction method [e.g., Boardman and Kopitzke (1975)].

Farebrother (1987) and Ding (1992) have presented algorithms for computing the non-central χ^2 distribution function. It is important to add here that Boomsma and Molenaar (1994) recently reviewed four packages for MS-DOS personal computers (Electronic Tables, P Calc, Sta Table, and STATPOWER) producing cumulative probabilities and quantiles for many common continuous distributions discussed in this and the previous volumes, with a particular emphasis on the non-central χ^2 distribution.

Zolnowska (1965) described a method of generating random numbers from "Rayleigh-Rice" distributions.

8 APPROXIMATIONS

Many approximations to noncentral χ^2-distributions, in particular to the value of $\Pr[\chi'^2_\nu(\lambda) \le x]$, have been suggested. In selecting an approximation, both simplicity and accuracy should be considered, although these tend to be contrary requirements.

Early approximations to noncentral χ^2 fall roughly into two groups. In the first group we have normal approximations to the distribution of some fractional power of a noncentral χ^2 variate [Patnaik (1949); Abdel-Aty (1954); Sankaran (1959, 1963)]. These approximations can sometimes be extended by the inclusion of one or two terms in terms of an Edgeworth expansion, but the improvement in accuracy is often quite small, especially in view of the usually tedious calculation involved.

The second group consists of central gamma (i.e., central χ^2) approximations of form $\alpha\chi^2 + \beta$, where α, β are suitably chosen constants, [Patnaik (1949); Pearson (1959)]. These can be converted into normal approximations, such as by the Wilson-Hilferty [Chapter 18, Eq. (18.24)] cube root approximation (and hence made to enter the first group), or they can be extended by several terms in a Laguerre series expansion [see Khamis (1965); Tiku (1965)].

Both the form of (29.4) and the inequalities in (29.31) lead one to expect that a gamma distribution should give a useful approximation. The simplest approximation consists of replacing χ'^2 by a multiple of central χ^2, $c\chi_f^2$, say, with c and f so chosen that the first two moments of the two variables $\chi_\nu'^2(\lambda)$ and $c\chi_f^2$ agree. The appropriate values of c and f are

$$c = \frac{\nu + 2\lambda}{\nu + \lambda}; \; f = \frac{(\nu + \lambda)^2}{\nu + 2\lambda} = \nu + \frac{\lambda^2}{\nu + 2\lambda}. \tag{29.59}$$

This approximation was suggested by Patnaik (1949). [Two additional corrective terms to Patnaik's approximation were derived by Roy and Mohamad (1964).] Pearson (1959) suggested an improvement of this approximation, introducing an additional constant b, and choosing b, c, and f so that the first three moments of $\chi_\nu'^2(\lambda)$ and $(c\chi_f^2 + b)$ agree. The appropriate values of b, c, and f are

$$b = -\frac{\lambda^2}{\nu + 3\lambda}, \tag{29.60}$$

$$c = \frac{\nu + 3\lambda}{\nu + 2\lambda},$$

$$f = \frac{(\nu + 2\lambda)^3}{(\nu + 3\lambda)^2} = \nu + \frac{\lambda^2(3\nu + 8\lambda)}{(\nu + 3\lambda)^2}.$$

This gives a better approximation to $F(x; \nu, \lambda)$ than does Patnaik's approximation, for x large enough. But since the Pearson approximation ascribes a nonzero value to $\Pr[-\lambda^2(\nu + 3\lambda)^{-1} < \chi'^2 \leq 0]$, it is not as good an approximation when x is small.

It can be shown that for x and ν fixed, the error of Patnaik's approximation to $F(x; \nu, \lambda)$ is $O(\lambda^2)$ as $\lambda \to 0$, $O(\lambda^{-1/2})$ as $\lambda \to \infty$; the error of

Pearson's approximation is also $O(\lambda^2)$ as $\lambda \to 0$, but $O(\lambda^{-1})$ as $\lambda \to \infty$. In both cases the error bounds are uniform in x. In both Patnaik's and Pearson's approximations f is usually fractional so that interpolation is needed if standard χ^2 tables are used.

Approximations to the central χ^2 distribution (Chapter 18, Section 5) may be applied to the approximating central χ^2's in Patnaik's and Pearson's approximations. If the Wilson-Hilferty approximation [Chapter 18, Eq. (18.24)] be applied then the approximation

$$\left(\frac{\chi'^2}{\nu+\lambda}\right)^{1/3} \quad \text{normal with expected value} \quad 1 - \frac{2(\nu+2\lambda)}{9(\nu+\lambda)} \quad \text{and}$$

$$\text{variance}\,\frac{2(\nu+2\lambda)}{\nu+\lambda} \tag{29.61a}$$

is obtained [Abdel-Aty (1954)]. Sankaran (1959, 1963) discussed a number of such further approximations, including

$$\left\{\left\{\chi'^2 - \tfrac{1}{2}(\nu-1)\right\}^{1/2} \quad \text{approximately normal with expected value}\right.$$

$$\left\{1+\tfrac{1}{2}(\nu-1)\right\}^{1/2} \quad \text{and} \quad \text{variance } 1, \tag{29.61b}$$

$$\left\{\frac{\chi'^2 - \tfrac{1}{3}(\nu-1)}{(\nu+\lambda)}\right\}^{1/2} \quad \text{approximately normal with expected}$$

$$\text{value}\left\{1 - \frac{-1}{3(\nu+\lambda)}\right\}^{1/2} \quad \text{and variance } (\nu+\lambda)^{-1}, \tag{29.61c}$$

$$\left(\frac{\chi'^2}{\nu+\lambda}\right)^{h} \quad \text{approximately normal with expected value}$$

$$1 + h(h-1)\frac{\nu+2\lambda}{(\nu+\lambda)^2} - h(h-1)(2-h)(1-3h)\frac{(\nu+2\lambda)^2}{2(\nu+\lambda)^4}$$

and variance

$$h^2\frac{2(\nu+2\lambda)}{(\nu+\lambda)^2}\left[1-(1-h)(1-3h)\frac{\nu+2\lambda}{(\nu+\lambda)^2}\right], \tag{29.61d}$$

where

$$h = 1 - \tfrac{2}{3}(\nu+\lambda)(\nu+3\lambda)(\nu+2\lambda)^{-2}.$$

Table 29.2 Errors of Johnson **(29.68)**, Patnaik **(29.59)**, **Pearson (29.60)**, Abdel-Aty **(29.61a)**, and Sankaran **(29.61b–d)** approximations for ν = 2,4 and 7

ν	λ	Exact Value	Johnson (29.68)	Patnaik (29.59)	Pearson (29.60)	Abdel-Aty (29.61a)	Sankaran (29.61b)	Sankaran (29.61c)	Sankaran (29.61d)
Upper 5% points									
2	1	8.642	0.92	−0.01	−0.04	−0.08	0.09	0.23	−0.06
	4	14.641	0.55	0.08	−0.06	0.02	0.04	0.04	−0.01
	16	33.054	0.28	0.29	−0.03	0.27	0.02	0.01	0.02
	25	45.308	0.23	0.35	−0.03	0.33	0.01	0.00	0.00
4	1	11.707	0.88	0.01	−0.02	−0.04	0.20	0.26	−0.02
	4	17.309	0.57	0.07	−0.04	0.03	0.11	0.08	−0.03
	16	35.427	0.30	0.26	−0.03	0.23	0.04	0.01	0.00
	25	47.613	0.25	0.33	−0.02	0.30	0.03	0.01	0.01
7	1	16.003	0.83	0.01	−0.01	−0.02	0.28	0.24	−0.02
	4	21.228	0.59	0.05	−0.02	0.02	0.18	0.10	0.03
	16	38.970	0.33	0.19	−0.02	0.19	0.08	0.02	−0.01
	25	51.061	0.26	0.28	−0.02	0.27	0.05	0.01	0.00
Lower 5% points									
2	1	0.17	*	0.03	−0.09	0.00	*	*	−0.05
	4	0.65	−0.43	0.29	−0.12	0.24	0.08	−0.01	0.01
	16	6.32	−0.25	0.57	−0.02	0.55	0.02	0.00	0.02
	25	12.08	−0.21	0.60	−0.01	0.59	0.01	0.00	0.03
4	1	0.91	−0.07	0.02	−0.03	0.00	*	0.14	−0.04
	4	1.77	−0.24	0.18	−0.06	0.17	0.23	0.01	−0.03
	16	7.88	−0.20	0.48	−0.02	0.47	0.06	0.00	0.02
	25	13.73	−0.17	0.53	−0.01	0.53	0.04	0.00	0.05
7	1	2.49	0.10	0.02	0.00	0.00	0.64	0.11	−0.02
	4	3.66	−0.07	0.12	−0.02	0.10	0.34	0.03	−0.01
	16	10.26	−0.15	0.38	−0.01	0.37	0.11	0.00	0.01
	25	16.23	−0.14	0.45	−0.01	0.44	0.07	0.00	0.02

Note: The tabled quantity is the approximate value less the exact value. Exact values of upper 5% points are taken from Fisher (1928), and lower 5% points from Ganvood (1934). The stability of the errors in Pearson's approximation to the upper 5% points is noteworthy. A correction of $0.16(\nu + 2)^{-1}$ would produce remarkably accurate results; this would also apply to the lower 5% points for $\nu = 4$ and $\nu = 7$.

Of these, (29.61b) is not good for small values of λ; (29.61d) is remarkably accurate for all λ but is rather complicated, and not much more accurate than Pearson's approximation. (See the comparisons in Table 29.2.)

Hayya and Ferrara (1972) found evidence from simulation that the normal approximation [based on Patnaik's (1949) approximation] that

$$\left\{\frac{2\chi'^2 - (\nu + \lambda)}{\nu + 2\lambda}\right\}^{1/2}$$

is distributed normally with expected value

$$\left\{\frac{2(\nu+\lambda)^2}{\nu+2\lambda}-1\right\}^{1/2}$$

and unit variance is "justified" at the 5% significance level. For $\lambda < 80$ the approximation is "justified" for the right tail (but not the left tail). (The authors are rather vague as the meaning of "justified," which is based on a graphic representation.) As was already mentioned above, any of these approximations might be improved by using an Edgeworth expansion, but the need for calculating higher cumulants makes this unattractive.

Rice (1968) gave an expansion (as a series in powers of ν^{-1}) for the cumulative distribution function which should give uniform accuracy over the whole range of values of the argument. Other approximations, valid for small values of λ, may be obtained from the Laguerre series expansion (29.11). Better results are obtained by expanding the distribution of an appropriate *linear function* of χ'^2 in a Laguerre series [Tang (1938); Tiku (1965)].

Bol'shev and Kuznetzov (1963) used a method in which the distribution of $\chi_\nu'^2(\lambda)$ is related to the distribution of a central χ^2 with the same number of degrees of freedom. They wished to determine a function $w(x;\nu,\lambda)$ such that

$$\Pr\left[\chi_\nu'^2(\lambda) \le x\right] \doteqdot \Pr\left[\chi_\nu^2 \le w(x;\nu,\lambda)\right].$$

This is equivalent to requiring that $w(\chi_\nu'^2(\lambda);\nu,\lambda)$ be approximately distributed as a central χ^2 with ν degrees of freedom.

For small λ,

$$w(x;\nu,\lambda) = w^*(x;\nu,\lambda) + O(\lambda^3), \tag{29.62}$$

where

$$w^*(x;\nu,\lambda) = x - x\lambda\nu^{-1} + \tfrac{1}{2}x\left\{1+(\nu+2)^{-1}x\right\}\lambda^2\nu^{-2}$$

and $O(\lambda^3)$ is uniform in any finite interval of x. Hence, as $\lambda \to 0$,

$$F(x;\nu,\lambda) = \Pr\left[\chi^2 \le w^*(x;\nu,\lambda)\right] + O(\lambda^3). \tag{29.63}$$

To estimate percentage points, namely solutions $x(\alpha,\nu,\lambda)$ of the equations

$$F(x;\nu,\alpha) = \alpha, \tag{29.64}$$

the inverse function

$$x(w^*;\nu,\lambda) = w^* + w^*\nu^{-1} + \tfrac{1}{2}w^*\left\{1-(\nu+2)^{-1}w^*\right\}\lambda^2\nu^{-2} \tag{29.65}$$

is used. If $\chi^2_{\nu,\alpha}$ is the (tabulated) $100\alpha\%$ point of the central χ^2_ν-distribution,

$$x^* = \chi^2_{\nu,\alpha} + \chi^2_{\nu,\alpha}\lambda\nu^{-1} + \tfrac{1}{2}\chi^2_{\nu,\alpha}\left\{1 - (\nu+2)^{-1}\chi^2_{\nu,\alpha}\right\}\lambda^2\nu^{-2} \quad (29.66)$$

is used as an approximation to $x(\alpha, \nu,$ A).

Next, we mention two formulas obtained by direct normal approximation. If a normal distribution is fitted to the $\chi'^2_\nu(\lambda)$-distribution, we obtain

$$F(x;\nu,\lambda) \doteqdot \Phi\left[\frac{x-\nu-\lambda}{\{2(\nu+2\lambda)\}^{1/2}}\right], \quad (29.67)$$

where

$$\Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-u^2/2}\,du.$$

Applying a normal approximation to the right-hand side of (29.6), Johnson (1959) obtained

$$F(x;\nu,\lambda) \doteqdot \Phi\left[\frac{x-\nu-\lambda+1}{\{2(\nu+2\lambda)\}^{1/2}}\right]. \quad (29.68)$$

In each case the error (as $\lambda \to \infty$) is $O(\lambda^{-1/2})$ uniformly in x.

These approximations are simple but not very accurate. The relative accuracy of a number of approximations can be judged from Table 29.2. It can be seen that only Pearson (29.60) and Sankaran (29.61d) are reliable over a wide range of values of A. Patnaik's and Abdel-Aty's formulas deteriorate as λ increases, while the other formulas improve.

Germond and Hastings (1944) gave the approximation

$$\Pr\left[\chi'^2_2(\lambda) \le R^2\right] \doteqdot \frac{R^2}{2+(R^2/2)}\exp\left[-\frac{\lambda}{2+(R^2/2)}\right] \quad (29.69)$$

which is correct to four decimal places for $R \le 0.4$. They also gave a table of (small) corrections to this formula, giving four decimal place accuracy up to $R = 1.2$, and further tables for larger values of R. For $R > 5$ the formula

$$\Pr\left[\chi'^2_2(\lambda) \le R^2\right] \doteqdot \frac{1}{\sqrt{2\pi}}\int_{\sqrt{\lambda}-\sqrt{R^2-1}}^{\infty} e^{-t^2/2}\,dt \quad (29.70)$$

gives useful results. We also note a simple empirical approximation to $x(0.95, \nu,$ A) due to Tukey (1957) which appears to be quite accurate.

The following useful approximations, for x and A both large, are due to Temme (1993):

For $x >$ A.

$$F(x;\nu,\lambda) \doteqdot 1 - \left(\frac{x}{\lambda}\right)^{(\nu-1)/4}\left\{1 - \Phi(\sqrt{2x} - \sqrt{2\lambda})\right\}. \qquad (29.71a)$$

For $x \leq$ A,

$$F(x;\nu,\lambda) \doteqdot \left(\frac{x}{\lambda}\right)^{(\nu-1)/4}\left\{1 - \Phi(\sqrt{2\lambda} - \sqrt{2x})\right\}. \qquad (29.71b)$$

9 APPLICATIONS

One use of the noncentral χ^2-distribution—that of representing the distribution of a sample variance from a normal population with unstable expected value—has been described in Section 1. A rather more generally useful application is in approximating to the power of χ^2-tests applied to contingency tables (tests of goodness of fit). In one of simplest of such tests the data consist of N observations divided among k classes $\Pi_1, \Pi_2, \cdots, \Pi_k$ with N_i observations in class Π_i $(i = 1, \ldots, k)$. If H, is the hypothesis that the probability of an observation falling into Π_i is a, $(i = 1, 2, \ldots, k)$ and the alternative hypotheses specify other values for these probabilities, then an approximation to the likelihood ratio test is one with a critical region of form

$$T = \sum_{i=1}^{k} \frac{(N_i - N\pi_{i0})^2}{N\pi_{i0}} > K_\alpha, \qquad (29.72)$$

where K_α is a suitably chosen constant. If the true values of the probabilities are a, $(i = 1, 2, \ldots, k)$ with, of course, $\sum_{i=1}^{k}\pi_i = 1$, then T is approximately distributed as $\chi'^2_{k-1}(\lambda)$ with

$$\lambda = N\sum_{i=1}^{k}\frac{(\pi_i - \pi_{i0})^2}{\pi_{i0}}.$$

If $\pi_i =$ a, for all i, that is, if H, is valid, then A $= 0$ and the approximate distribution is that of a central χ^2_{k-1}. So, to obtain a significance level approximately equal to α, we take

$$K_\alpha = \chi^2_{k-1,\alpha}.$$

The power, when the true values of the probabilities are a, $\pi_2, \ldots,$ a, is then approximately

$$\Pr\left[\chi_{k-1}'^{2}(\lambda) > \chi_{k-1,\alpha}^{2}\right].$$

There is a good discussion of more complex forms of χ^2 tests in Patnaik (1949). Noncentral χ^2 also appears in the calculation of approximate powers of certain nonparametric tests [Andrews (1954); Lehmann (1959, pp. 302–306)].

Extending an argument of Wilks (1962, p. 419), it can be shown that when the data can be represented by n independent random variables with identical distributions depending on parameters $(\theta_1, \theta_2, \cdots, \theta_k)$, then the limiting distribution (as $n \to \infty$) of

$$-2\log\text{(likelihood ratio)}$$

is, under certain sequences of alternative hypotheses converging to the null hypothesis, a noncentral χ^2-distribution. The likelihood ratio here is the ratio of two maximized values of the likelihood function, the numerator being restricted by certain of the θ's being assigned fixed values, while the denominator is not restricted.

Sugiura (1968) obtained an asymptotic expansion (up to order n^{-1}) of the non-null distribution of the logarithm of the likelihood ratio statistic for tests of multivariate linear hypotheses in the form of a linear combination of noncentral χ^2 probabilities with increasing numbers of degrees of freedom and the same noncentrality parameter.

Noncentral χ^2 also appears in a slightly disguised form in calculations of the probability that a random point $(X_1, X_2, \cdots, X_\nu)$ with the X's mutually independent normal variables, each having expected value 0 and standard deviation σ (the same for all i), falls within an offset hypersphere $\sum_{i=1}^{\nu}(X_i - \xi_i)^2 \le R^2$. This probability is evidently

$$\Pr\left[\chi_{\nu}'^{2}\left(\sigma^{-2}\sum_{i=1}^{\nu}\xi_i^2\right) \le \left(\frac{R}{\sigma}\right)^2\right].$$

For $\nu = 2$, this is the probability of hitting a circular target of radius R when the point of aim is offset (ξ_1, ξ_2) from the center of the target and variation about the point of aim is spherical normal with variance σ^2.

A number of tables of this quantity have been produced, especially for the physically interesting cases $\nu = 2$ and $\nu = 3$. An extensive summary is given by Guenther and Terragno (1964), who also give a useful bibliography. For the case $\nu = 2$, very detailed tables are available [Bell Aircraft Corp. (1956); DiDonato and Jarnagin (1962a); Marcum (1950)]. The most easily available short summary tables are in Burington and May (1970), DiDonato and Jarnagin (1962a, b), and Owen (1962, pp. 178–180). The DiDonato and

Jarnagin references contain values of R/σ to give specified values of probabilities; the other references give values of probabilities for specified values of R/σ and $(\xi_1^2 + \xi_2^2)/\sigma^2$.

For the case $\nu = 3$, there is the simple relation

$$\Pr\left[\chi_3'^2(\lambda) \le \left(\frac{R}{\sigma}\right)^2\right] = \frac{1}{\sqrt{2\pi}} \int_{\lambda^{1/2}-(R/\sigma)}^{\lambda^{1/2}+(R/\sigma)} \exp\left(-\frac{1}{2}u^2\right) du$$
$$-\frac{1}{\sqrt{2\pi\lambda}}\left[\exp\left\{-\frac{1}{2}\left(\lambda^{1/2} - \frac{R}{\sigma}\right)^2\right\}\right.$$
$$\left. - \exp\left\{-\frac{1}{2}\left(\lambda^{1/2} + \frac{R}{\sigma}\right)^2\right\}\right]. \quad (29.73)$$

There is thus less need for extensive tables in this case, but there is a short table available [Guenther (1961)].

For general conditions under which a quadratic form in normal variables is distributed as noncentral χ^2, see Chapter 29 of the first edition of this volume. Spruill (1979) has shown that the measurement of electrical power in a circuit is related to the estimation of the noncentrality parameter of a chi-square distribution.

Noncentral χ^2-distributions have been applied in various aspects of financial theory:

1. Boyle (1978, 1979) observed that the *size* of individual claims in certain classes of insurance is distributed as $K\chi_2^2$ (i.e., has an exponential distribution), while the *number* of such claims in a specified period (e.g., a year) has a Poisson distribution. It follows that the total amount of claims over the period has a $K\chi_2'^2(2\theta)$ distribution, where θ is the expected value of the Poisson distribution.
2. The distribution of interest rates under specified assumptions about technological change and assumptions about preferences and existence of a steady state distribution for the interest rate was investigated by Cox, Ingersoll, and Ross (1985). It was observed that the probability density of the interest rate at time s, conditional on its value at the current time t, following a "mean-reverting process," appears to have a noncentral χ^2-distribution with the noncentrality parameter proportional to the current spot rate.
3. A more subtle application pertains to the so-called CEV (constant elasticity of variance) model, relating volatility and stock price(s). It is assumed to be governed by a "diffusion process" of the type

$$dS = \mu S\,dt + \delta S^{\beta-2}\,dZ,$$

Jarnagin references contain values of R/σ to give specified values of probabilities; the other references give values of probabilities for specified values of R/σ and $(\xi_1^2 + \xi_2^2)/\sigma^2$.

For the case $\nu = 3$, there is the simple relation

$$\Pr\left[\chi_3'^2(\lambda) \le \left(\frac{R}{\sigma}\right)^2\right] = \frac{1}{\sqrt{2\pi}} \int_{\lambda^{1/2}-(R/\sigma)}^{\lambda^{1/2}+(R/\sigma)} \exp\left(-\frac{1}{2}u^2\right) du$$
$$- \frac{1}{} \left[\exp\left\{-\frac{1}{2}\left(\lambda^{1/2} - \frac{R}{\sigma}\right)^2\right\}\right.$$
$$\left. - \exp\left\{-\frac{1}{2}\left(\lambda^{1/2} + \frac{R}{\sigma}\right)^2\right\}\right]. \quad (29.73)$$

There is thus less need for extensive tables in this case, but there is a short table available [Guenther (1961)].

For general conditions under which a quadratic form in normal variables is distributed as noncentral χ^2, see Chapter 29 of the first edition of this volume. Spruill (1979) has shown that the measurement of electrical power in a circuit is related to the estimation of the noncentrality parameter of a chi-square distribution.

Noncentral χ^2-distributions have been applied in various aspects of financial theory:

1. Boyle (1978, 1979) observed that the size of individual claims in certain classes of insurance is distributed as $K\chi_2^2$ (i.e., has an exponential distribution), while the number of such claims in a specified period (e.g., a year) has a Poisson distribution. It follows that the total amount of claims over the period has a $K\chi_2'^2(2\theta)$ distribution, where θ is the expected value of the Poisson distribution.
2. The distribution of interest rates under specified assumptions about technological change and assumptions about preferences and existence of a steady state distribution for the interest rate was investigated by Cox, Ingersoll, and Ross (1985). It was observed that the probability density of the interest rate at time s, conditional on its value at the current time t, following a "mean-reverting process," appears to have a noncentral χ^2-distribution with the noncentrality parameter proportional to the current spot rate.
3. A more subtle application pertains to the so-called CEV (constant elasticity of variance) model, relating volatility and stock price(s). It is assumed to be governed by a "diffusion process" of the type

$$dS = \mu S\,dt + \delta S^{\beta-2}\,dZ,$$

where dZ is a Wiener process and $\beta - 2$ is the so-called elasticity of return variance with respect to price. If $\beta = 2$ (i.e., the elasticity is zero), prices are lognormally distributed and the variance of returns is constant. Utilizing Cox, Ingersoll, and Ross's (1985) results, Schroder (1989) has shown that the CEV process can be expressed as a linear combination of cdfs of two noncentral chi-squared densities with different degrees of freedom but the same noncentrality parameter.

4. Hayya and Ferrara (1972) encountered a noncentral chi-squared distribution in a risk analysis model relating costs and revenues.

10 RELATED DISTRIBUTIONS

We have noted in Section 6, that $\chi_1'(\lambda)$ is a ***folded*** normal variable (discussed in Chapter 13). Equations (29.3) and (29.4) represent a connection between the noncentral χ^2- and Poisson distributions. Other relationships, already mentioned in this chapter, are as follows:

1. If $A = 0$, the noncentral χ^2 becomes a central χ^2.
2. The limiting distribution of a standardized $\chi_\nu'^2(\lambda)$ variable is the unit normal distribution if either (a) $\nu \to \infty$, A remaining constant, or (b) $A \to \infty$, ν remaining constant.
3. The limiting distribution of a standardized (singly or doubly) noncentral F variable, as the denominator degrees of freedom tends to infinity (noncentralities remaining constant) is the distribution of multiple of a noncentral χ^2 variable (Chapter 30, Section 5).
4. Press (1966) has shown that the distribution of linear functions of independent noncentral χ^2 variates with positive coefficients can be expressed as mixtures of distributions of central χ^2's. This is part of the theory of quadratic forms in normal variables, which is the subject of a chapter in a planned volume on ***Continuous Multivariate Distributions.***
5. Mixtures of central χ^2-distributions, analogous to (29.3) but with different weights, occur as the null hypothesis distributions of certain test statistics. Bartholomew (1959a, b) encountered such a situation in constructing a test of the hypothesis that a sequence of expected values of normal distributions $\{\xi_i\}$ $(i = 1,\dots,k)$ with common variance (σ^2) is constant against alternatives specifying ordering of the ξ_i's.

The statistic, based on m mutually independent random variables $X_1,\cdots,X_m$,

$$\bar{\chi}^2 = \sum_{j=1}^{m} \frac{c_j X_j}{\sum_{j=1}^{m} c_j}, \qquad c_j > 0, \tag{29.74}$$

with X_j distributed as $\chi^2_{\nu_j}$ $(j = 1,\dots,m)$ is distributed as a mixture of a finite number of central χ^2 distributions. The name "chi-bar-squared" distribution was coined, apparently by Conoway et al. (1990) to apply to the distribution obtained by replacing the central χ^2's by noncentral χ^2's [see Chapter 18, Eq. (18.71)]. The latter can in turn be represented according to (29.4) as mixtures of central χ^2's, so that (29.74) includes all chi-bar-squared distributions. See Bartholomew (1961), Barlow et al. (1972), and Conoway et al. (1990) for further details.

Chi-bar-squared distributions can arise as compound noncentral chi-squared distributions in which the noncentrality parameter, A, of $\chi'^2_\nu(\lambda)$ is ascribed a distribution. Then the pdf of the corresponding random variable Y, say, is

$$p_Y(y) = \sum_{j=0}^{\infty} E\left[e^{-\lambda/2}\frac{\left(\frac{1}{2}\lambda\right)^j}{j!}\right]p(y;\nu+2j,0), \qquad (29.75)$$

with a corresponding formula for the cdf.

Szroeter (1992) has considered the case where

$$\text{A} = \text{T'T} + \text{c}^2,$$

where T is a r X 1 vector, with variance-covariance matrix Ω and expected value T. If ω denotes the largest eigenvalue of Ω then, as $\omega \to 0$,

$$E_\lambda[F(x;\nu,\text{A})] \to F(x;\nu,\text{A}),$$

where

$$\text{A} = \boldsymbol{\tau}'\boldsymbol{\tau} + \text{c}^2.$$

Szroeter obtained the bound

$$\begin{aligned}(1-\theta\delta^{-2})F(x;\nu,\Lambda+\delta) &\le E_\lambda[F(x;\nu,\lambda)]\\ &\le (1-\theta\delta^{-2})F(x;\nu,\max(0,\Lambda-\delta)+\theta\delta^{-2})\end{aligned} \qquad (29.76)$$

for any $\delta \ge \sqrt{\theta}$, with

$$\theta = \omega\left[2\Lambda + 2\omega\,\text{rank}(\boldsymbol{\Omega}) + \omega\{\text{rank}(\boldsymbol{\Omega})\}^2\right].$$

He also obtained the alternative upper bound

$$E_\lambda[F(x;\nu,\lambda)] \le F\left(x;\nu,(1+\omega)^{-1}\Lambda\right) \qquad (29.77a)$$

and the alternative lower bound

$$F\big((1+\omega)^{-1}x;\nu',\mathrm{A}\big) \le E_\lambda[F(x;\nu,\mathrm{A})], \tag{29.77b}$$

where

$$\nu' = \begin{cases} \nu & \text{for } \nu = r = \mathrm{rank}(\boldsymbol{\Omega}),\ c = 0, \\ \max(\nu, \mathrm{rank}(\boldsymbol{\Omega}) + 1) & \text{otherwise.} \end{cases}$$

If X_1 and X_2 are mutually independent random variables, with X_j distributed as $\chi_{\nu_j}'^2(\lambda_j)$ $(j = 1, 2)$, the distribution of the ratio $Y = X_1/X_2$ is easily obtained as that of a **mixture** of $G_{\nu_1+2j_1,\nu_2+2j_2}$ variables (as defined in Chapter 27) with products of Poisson probabilities

$$\prod_{h=1}^{2}\left\{e^{-\lambda_h/2}\frac{\left(\frac{1}{2}\lambda_h\right)^{i_h}}{i_h!}\right\} \tag{29.78}$$

as weights. Using equation (27.3) of Chapter 27, we have

$$p_Y(y) = e^{-(\lambda_1+\lambda_2)/2}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\left\{\prod_{h=1}^{2}\frac{\left(\frac{1}{2}\lambda_h\right)^{i_h}}{i_h!}\right\} \times\left[\frac{y^{(\nu_1/2)+i_1-1}}{B\left(\frac{1}{2}\nu_1+i_1,\frac{1}{2}\nu_2+i_2\right)(1+y)^{\frac{1}{2}(\nu_1+\nu_2)+i_1+i_2}}\right]. \tag{29.79a}$$

The distribution of the product $Z = X_1X_2$ is similarly that of a mixture of $\chi^2_{\nu_1+2i_1}\chi^2_{\nu_2+2i_2}$ distributions, with the same weights (29.78). The formal expression is more complicated in appearance than **(29.79a)** because the distribution of the product of $\chi^2_{\nu_1+2i_1}$ and $\chi^2_{\nu_2+2i_2}$ is more complicated in appearance than that of their ratio. Using the distribution of the product of two mutually independent gamma variables, given in Chapter 17, Section 8.4, we obtain

$$e^{-\frac{1}{2}(\lambda_1+\lambda_2)}\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\left\{\prod_{h=1}^{2}\frac{\left(\frac{1}{2}\lambda_h\right)^{i_h}}{i_h!}\right\}\frac{\left(\frac{1}{2}z\right)^{\frac{1}{4}(\nu_1+\nu_2)+\frac{1}{2}(i_1+i_2)-1}K_{\frac{1}{2}(\nu_1-\nu_2)+i_1-i_2}(\sqrt{z})}{\Gamma\left(\frac{1}{2}\nu_1+i_1\right)\Gamma\left(\frac{1}{2}\nu_2+i_2\right)}, \tag{29.79b}$$

where

$$K_g(y) = \frac{1}{2}\int_0^\infty t^{g-1}\exp\left\{-\frac{1}{2}y(t+t^{-1})\right\}dt$$

is the modified Bessel function of the second kind (see Chapter 1). [Note that $K_g(y) = K_{-g}(y)$.]

Kotz and Srinivasan (1969) have obtained (29.79a) and (29.79b), using Mellin transforms. [In their formulas, summation is with respect to i_1 and $(i_1 + i_2)$ instead of i_1 and i_2.] For two degrees of freedom, simpler expressions are available in Kotz and Srinivasan (1969).

REFERENCES

Abdel-Aty, S. H. (1954). Approximate formulae for the percentage points and the probability integral of the non-central χ^2-distribution, ***Biometrika,*** 41, 538–540.

Abramowitz, M., and Stegun, I. A. (1964). ***Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,*** National Bureau of Standards, Applied Mathematics Series, 55, Washington, DC: GPO.

Alam, K., and Rizvi, M. H. (1967). On non-central chi-squared and non-central F distributions, ***The American Statistician,*** **21**(4), 21–22.

Alam, K., and Saxena, L. (1982). Estimation of the noncentrality parameter of a chi-square distribution, ***Annals of Statistics,*** 10, 1012–1016.

Anderson, D. A. (1981a). The circular structural model, ***Journal of the Royal Statistical Society, Series B,*** 43, 131–143.

Anderson, D. A. (1981b). Maximum likelihood estimation in the non-central chi-distribution with unknown scale parameter, ***Sankhyā, Series*** B, 43, 58–67.

Andrews, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance, ***Annals of Mathematical Statistics,*** 25, 724–736.

Ashour, S. K., and Abdel-Samad, A. I. (1990). On the computation of non-central χ^2 distribution, ***Communications in Statistics—Simulation and Computation,*** 19, 1279–1291.

Bargmann, R. E., and Ghosh, S. P. (1964). ***Noncentral Statistical Distribution Programs for a Computer Language,*** IBM Research Report R.C.-1231.

Bark, L. S., Bol'shev, L. N., Kuznetzov, P. I., and Cherenkov, A. P. (1964). ***Tables of the Rayleigh-Rice Distribution,*** Computation Center, Academy of Sciences, USSR, Moscow.

Bartholomew, D. J. (1959a). A test of homogeneity for ordered alternatives, ***Biometrika,*** 46, 36–48.

Bartholomew, D. J. (1959b). A test of homogeneity for ordered alternatives II, ***Biometrika,*** 46, 328–335.

Bartholomew, D. J. (1961). A test of homogeneity of means under restricted alternatives, ***Journal of the Royal Statistical Society, Series B,*** 23, 239–272.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). ***Statistical Inference under Order Restrictions—The Theory and Application of Isotonic Regression,*** New York: Wiley.

Bell Aircraft Corporation (1956). ***Tables of Circular Normal Probabilities,*** Report No. 02-949-106, Operations Analysis Group, Dynamics Section, Bell Aircraft Corporation. Buffalo. NY.

Bennett, B. M. (1955). Note on the moments of the logarithmic noncentral χ^2 and z distributions, ***Annals of the Institute of Statistical Mathematics,*** **7,** 57–61.

Boardman, T. J., and Kopitzke, R. W. (1975). Probability and table values for statistical distributions, ***Proceedings of ASA Statistical Computing Section,*** **pp.** 81–86.

Bock, M. E., Judge, G. G., and Yancey, T. A. (1984). A simple form for the inverse moments of non-central χ^2 and F random variables and certain confluent hypergeometric functions, ***Journal of Econometrics,*** **25,** 217–234.

Bol'shev, L. N., and Kuznetzov, P. I. (1963). On computing the integral $p(x, y) = \dots$ ***Zhurnal Vychislitelnoj Matematiki i Matematicheskoi Fiziki,*** **3,** 419–430. (In Russian)

Bol'shev, L. N., and Smirnov, N. V. (1965). ***Tables of Mathematical Statistics,*** Moscow: Akademia Nauk SSSR.

Boomsma, A., and Molenaar, I. W. (1994). Four electronic tables for probability distributions, ***The American Statistician,*** **48,** 153–162.

Boyle, P. P. (1978). The Poisson-exponential model and the non-central chi-squared distribution, ***Scandinacian Actuarial Journal,*** 108–111.

Boyle, P. P. (1979). Reply to remark by Thélander, ***Scandinacian Actuarial Journal,*** 55–56.

Burington, R. S., and May, D. C. (1970). ***Handbook of Probability and Statistics with Tables,*** Second edition, Sandusky, OH: Handbook Publishers.

Chou, Y.-M., Arthur, K. H., Rosenstein, R. B., and Owen, D. B. (1984). New representations of the non-central chi-square density and cumulative, ***Communications in Statistics—Theory and Methods,*** **13,** 2673–2678.

Chow, M. S. (1987). A complete class theorem for estimating a non-centrality parameter, ***Annals of Statistics,*** **15,** 800–804.

Cohen, J. P. (1988). Noncentral chi-square: Some observations on recurrence, ***The American Statistician,*** **42,** 120–122.

Conoway, M., Pillers, C., Robertson, T., and Sconing, J. (1990). The power of the circular cone test: A noncentral chi-bar-squared distribution, ***Canadian Journal of Statistics,*** **18,** 63–70.

Cox, J., Ingersoll, E., and Ross, S. A. (1985). A theory of the term structure of interest rates, ***Econometrics,*** **53,** 385–407.

De Waal, D. J. (1974). Bayes estimate of the noncentrality parameter in multivariate analysis, ***Communications in Statistics,*** **3,** 73–79.

Di Donato, A. R., and Jarnagin, M. P. (1962a). A method for computing the generalized circular error function and circular coverage functions, NWL ***Report No. 1768,*** U.S. Naval Weapons Laboratory, Dahlgren, VA.

Di Donato, A. R., and Jarnagin, M. P. (1962b). A method for computing the circular coverage function, ***Mathematics of Computation,*** **16,** 347–355.

Ding, C. G. (1992). Algorithm AS 275: Computing the non-central χ^2 distribution function, ***Applied Statistics,*** **41,** 478–482.

Dwivedi, T. D., and Pandey, J. N. (1975). A note on Meyer's maximum likelihood estimate of the non-centrality of the non-central χ^2 variate, ***Sankhyā, Series*** B, **37.** 453–456.

Egerton, M. F., and Laycock, P. J. (1982). An explicit formula for the risk of James-Stein estimators, *Canadian Journal of Statistics*, **10**, 199–205.

Elandt, R. C. (1961). The folded normal distribution: Two methods of estimating parameters from moments, *Technometrics*, 3, 551–562.

Ennis, D. M., and Johnson, N. L. (1993). Noncentral and central chi-square, F and beta distribution functions as special cases of the distribution of an indefinite quadratic form, *Communications in Statistics—Theory and Methods*, **22**, 897–905.

Farebrother, R. W. (1987). Algorithm AS 231: The distribution of a noncentral χ^2 variable with nonnegative degrees of freedom, *Applied Statistics*, 36, 402–405.

Felsen, L. B. (1963). Radiation from a uniaxially anisotropic plasma half-space, IEEE *Transactions on Antennas and Propagation*, 11, 469–484.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient, *Proceedings of the Royal Society of London*, **121A**, 654–673.

Fix, E. (1949). Tables of non-central χ^2, *University of California Publications in Statistics*, 1, No. 2, 15–19.

Garwood, F. (1934). Cited in Abdel-Aty (1954), Unpublished Ph.D. dissertation, University of London.

Germond, H. H., and Hastings, C. (1944). *Scatter Bombing of a Circular Target*, A report submitted by the Bombing Research Group, Columbia University and the Applied Mathematics Group, Columbia University to the Applied Mathematics Panel, National Defense Research Committee, May 1944.

Ghosh, B. K. (1973). Some monotonicity theorems for χ^2, F and t distributions with applications, *Journal of the Royal Statistical Society, Series B*, *35*, 480–492.

Gideon, R. A., and Gurland, J. (1977). Some alternative expansions for the distribution function of a noncentral chi-square variable, *SIAM Journal on Mathematical Analysis*, 8, 100–110.

Graybill, F. A. (1961). *An Introduction to Linear Models*, vol. 1, New York: McGraw-Hill.

Guenther, W. C. (1961). On the probability of capturing a randomly selected point in three dimensions, *SIAM Review*, 3, 247–250.

Guenther, W. C. (1964). Another derivation of the non-central chi-square distribution, *Journal of the American Statistical Association*, *59*, 957–960.

Guenther, W. C. (1975). Evaluation of noncentral distribution integrals with a desk calculator, *Research Paper No. 80 S-1975-578*, College of Commerce and Industry, University of Wyoming, Laramie.

Guenther, W. C., and Terragno, P. J. (1964). A review of the literature on a class of coverage problems, *Annals of Mathematical Statistics*, *35*, 232–260.

Han, C. P. (1975). Some relationships between noncentral chi-squared and normal distributions, *Biometrika*, 62, 213–214.

Haynam, G. E., Govindarajulu, Z., and Leone, F. C. (1973). *Tables of the cumulative nun-central chi-square distribution*, In *Selected Tables in Mathematical Statistics*, vol. 1 (eds., H. L. Harter and D. B. Owen,) Providence, RI: American Mathematical Society, pp. 1–78.

Hayya, J. C., and Ferrara, W. L. (1972). On normal approximations to the frequency functions of standard forms where the main variables are normally distributed, *Management Science*, **19**, 173–186.

Helstrom, C. W. (1960). *Statistical Theory of Signal Detection,* Oxford: Pergamon.

Hjort, N. L. (1989). *The eccentric part of the non-central chi-square distribution,* Manuscript, Norwegian Computing Center, Oslo.

Johnson, N. L. (1959). On an extension of the connexion between Poisson and χ^2-distributions, *Biometrika,* 46, 352–363.

Johnson, N. L. (1962). The folded normal distribution: Accuracy of estimation by maximum likelihood, *Technometrics,* **4,** 249–256.

Johnson, N. L. (1968). *Tables of Percentile Points of Noncentral Chi-square Distributions,* Mimeo Series No. 568, Institute of Statistics, University of North Carolina.

Johnson, N. L., and Leone, F. C. (1964). *Statistics and Experimental Design in Engineering and the Physical Sciences,* vol. I, New York: Wiley.

Johnson, N. L., and Pearson, E. S. (1959). Tables of percentage points of non-central χ, *Biometrika,* **56,** 315–333.

Jones, M. C. (1989). Letter to the editor, *The American Statistician,* **43,** 68.

Kallenberg, W. C. M. (1990). Inequalities for noncentral chi-square distributions, *Statistics & Probability Letters, 9,* 273–278.

Kerridge, D. (1965). A probabilistic derivation of the non-central χ^2 distribution, *Australian Journal of Statistics, 7,* 37–39. (Correction: 7, 114.)

Khamis, S. H. (1965). Some basic properties of the incomplete gamma function, *Annals of Mathematical Statistics,* **36,** 926–937.

Kotz, S., and Srinivasan, R. (1969). Distribution of product and quotient of Bessel function variates, *Annals of the Institute of Statistical Mathematics,* **21,** 201–210.

Kubokawa, T., Robert, C. P., and Saleh, A. K. Md. E. (1993). Estimation of noncentrality parameters, *Canadian Journal of Statistics,* **21,** 45–57.

Lam, Y.-M. (1987). Confidence intervals for noncentrality parameters of noncentral chi-squared and F distributions, *Proceedings of ASA Statistical Computing Section,* 441–443.

Lau, C. (1980). Algorithm AS147. A simple series for the incomplete gamma integral, *Applied Statistics, 29,* 113–114.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses,* New York: Wiley (2nd ed., 1986).

Leone, F. C., Nelson, L. S., and Nottingham, R. B. (1961). The folded normal distribution, *Technometrics,* **3,** 543–550.

Lowe, J. R. (1960). A table of the integral of the bivariate normal distribution over an offset circle, *Journal of the Royal Statistical Society, Series B, 22,* 177–187.

Marcum, J. I. (1948). *A statistical theory of target detection by pulsed radar: mathematical appendix,* Research Memorandum RM-733, Rand Corporation, Santa Monica, CA. [Also published as Marcum, J. I., and Swerling, P. (1960), *IRE Transactions PGIT,* Vol. IT-4.1

Marcum, J. I. (1950). *Tables of Q-functions,* Rand Report No. RM-339, Rand Corporation, Santa Monica, CA.

Marden, J. I. (1982). Combining independent non-central chi-squared or F tests, *Annals of Statistics,* **10,** 266–267.

McNolty, F. (1962). A contour-integral derivation of the non-central chi-square distribution, *Annals of Mathematical Statistics, 33,* 796–800.

Meyer, P. L. (1967). The maximum likelihood estimate of the non-centrality parameter of a non-central χ^2 variate, ***Journal of the American Statistical Association,*** **61,** 1258–1264.

Miller, K. S., Bernstein, R. I., and Blumenson, L. E. (1958). Generalized Rayleigh processes, ***Quarterly of Applied Mathematics,*** **16,** 137–145 (and Note, ***Ibid;*** **20,** Jan. 1963).

Narula, S. C., and Desu, M. M. (1981). Algorithm AS 170. Computation of probability and non-centrality parameter of a non-central χ^2 distribution, ***Applied Statistics,*** **30,** 349–352.

Narula, S. C., and Levy, K. J. (1975). Probability density plots of the noncentral χ^2 and noncentral F distributions, ***International Statistical Reuiew,*** **43,** 79–82.

Neff, N., and Strawderman, W. E. (1976). Further remarks on estimating the parameters of a noncentral chi-square distribution, ***Communications in Statistics—Theory and Methods,*** A5, 66–76.

Oosterhoff, J., and Schreiber, B. F. (1987). A note on complete families of distributions, ***Statistica Neerlandica,*** **41,** 183–189.

Owen, D. B. (1962). ***Handbook of Statistical Tables,*** Reading, MA: Addison-Wesley.

Park, J. H. (1961). Moments of the generalized Rayleigh distribution, ***Quarterly of Applied Mathematics, 19,*** 202–232.

Patnaik, P. B. (1949). The non-central χ^2- and F-distributions and their applications, ***Biometrika,*** **36,** 202–232.

Pearson, E. S. (1959). Note on an approximation to the distribution of noncentral χ^2, ***Biometrika,*** **46,** 364.

Perlman, M. D., and Rasmussen, U. (1975). Some remarks on estimating a noncentrality parameter, ***Communications in Statistics,*** **4,** 455–468.

Pike, M. C., and Hill, I. D. (1966). Algorithm 291. Logarithm of the gamma function, ***Communications of the Association for Computing Machinery, 9,*** 684.

Posten, H. O. (1989). An effective algorithm for the noncentral chi-squared distribution function, ***The American Statistician, 43,*** 261–263.

Press, S. J. (1966). Linear combinations of noncentral chi-square variates, ***Annals of Mathematical Statistics,*** **37,** 480–487.

Quenouille, M. H. (1949). The evaluation of probabilities in a normal multivariate distribution, with special reference to the correlation ratio, ***Proceedings of Edinburgh Mathematical Society, Series 2, 8,*** 95–100.

Rainville, E. D. (1960). ***Special Functions,*** New York: Macmillan.

Rice, S. O. (1968). Uniform asymptotic expressions for saddle point integrals—Application to a probability distribution occurring in noise theory, ***Bell System Technical Journal,*** **47,** 1971–2013.

Robertson, G. H. (1969). Computation of the noncentral chi-square distribution, ***Bell System Technical Journal,*** **48,** 201–207.

Roy, J., and Mohamad, J. (1964). An approximation to the non-central chi-square distribution, ***Sankhyā, Series A,*** **26,** 81–84.

Ruben, H. (1960). Probability content of regions under spherical normal distributions, I, ***Annals of Mathematical Statistics,*** **31,** 598–618.

Ruben, H. (1974). Non-central chi-square and gamma revisited, ***Communications in Statistics*, 3, 607–633.**

Ruben, H. (1976). A new result on the probability integral of noncentral chi-square with even degrees of freedom, ***Manuscript*,** McGill University, Montreal, Canada.

Sankaran, M. (1959). On the noncentral chi-square distribution, ***Biometrika*,** 46, **235–237.**

Sankaran, M. (1963). Approximations to the noncentral chi-square distribution, ***Biometrika*, 50, 199–204.**

Schneider, H. (1989). Failure-censored variables-sampling plans for lognormal and Weibull distributions, ***Technometrics*, 31, 199–206.**

Schroder, M. **(1989).** Computing the constant elasticity of variance option pricing formula, ***Journal of Finance*, *44*,** **211–219.**

Sen, P. K. **(1989).** The mean-median-mode inequality and noncentral chi square distributions, ***Sankhyā*, *Series A*, 51, 106–114.**

Shea, B. L. (1988). Algorithm AS **239.** Chi-squared and incomplete gamma integral, ***Applied Statistics*, *37*, 466–473.**

Siegel, A. F. **(1979).** The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity, ***Biometrika*,** 66, **381–386.**

Spruill, M. C. **(1979).** Estimation of the non-centrality parameter of a chi-squared distribution, ***Unpublished Report*,** School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia.

Spruill, M. C. (1986). Computation for the maximum likelihood estimate of a non-centrality parameter, ***Journal of Multivariate Analysis*, *18*, 216–224.**

Sugiura, N. **(1968).** ***Asymptotic expansions of the power functions of the likelihood ratio tests for multivariate linear hypotheses and independence*,** Mimeo Series No. **563,** Institute of Statistics, University of North Carolina, Chapel Hill.

Szroeter, J. (1992). Bounds for non-central chi-square distributions having unobservable random non-centrality parameters, ***Statistics & Probability Letters*, 13, 73–81.**

Tang, **P.** C. **(1938).** The power function of the analysis of variance tests with tables and illustrations of their use, ***Statistical Research Memoirs*, 2, 126–149.**

Temme, N. M. **(1993).** Asymptotic and numerical aspects of the noncentral chi-square distribution, ***Computers in Mathematics and Applications*, 25(5), 55–63.**

Tiku, M. L. (1965). Laguerre series forms of non-central χ^2 and F distributions, ***Biometrika*, 52, 415–426.**

Torgerson, E. N. (1972). Supplementary notes on linear models, ***Statistical Memoirs-1*,** Institute of Mathematics, University of Oslo, Norway.

Tukey, J. W. **(1957).** Approximations to the upper **5%** points of Fisher's ***B*** distribution and non-central χ^2, ***Biometrika*,** 44, **528–530.**

Ullah, A. **(1976).** On the sampling distribution of improved estimators for coefficients in linear regression, ***Journal of Econometrics*, *2*, 143–180.**

Urkowitz, H. **(1967).** Energy detection of unknown deterministic signals, ***Proceedings of IEEE*, 55, 523–531.**

Var der Vaart, H. R. **(1967).** A note on the derivation of the non-central chi-square density function, ***Statistica Neerlandica*, 21, 99–100.**

Venables, W. N. **(1971).** ***Inference problems based on non-central distributions,*** **Ph.D.** dissertation, Department of Statistics, University of Adelaide, Australia.

Venables, W. N. **(1975).** Calculation of confidence intervals for non-central distributions, ***Journal of the Royal Statistical Society, Series B, 27,*** **406–412.**

Wiener, H. L. **(1975).** A FORTRAN program for rapid computations involving the non-central chi-square distributions, ***NRL Memorandum Report,*** Naval Research Laboratory, Washington, DC.

Wiley, J. A., Herschokoru, S. J., and Padiau, N. **(1989).** Heterogeneity in the probability of HIV transmission in sexual contact: The case of male-to-female transmission in penile-vaginal intercourse, ***Statistics in Medicine,*** **8, 93–102.**

Wilks, S. S. **(1962).** ***Mathematical Statistics,*** New York: Wiley.

Zolnowska, H. **(1965).** Generators of random numbers of Rayleigh and Rice's distributions, ***Algorytmy***, 3, **73–94.**

CHAPTER 30

Noncentral F-Distributions

1 DEFINITION AND GENESIS

In Chapter 27 the F-distribution with ν_1, ν_2 degrees of freedom was defined as the distribution of the ratio $(\chi^2_{\nu_1}/\nu_1)(\chi^2_{\nu_2}/\nu_2)^{-1}$, where the two χ^2's are mutually independent. If both χ^2's are replaced by noncentral χ^2's, we obtain the ***doubly noncentral F-distribution with ν_1, ν_2 degrees of freedom and noncentrality parameters*** λ_1, λ_2 defined as the distribution of the ratio

$$\left\{\frac{\chi'^2_{\nu_1}(\lambda_1)}{\nu_1}\right\}\left\{\frac{\chi'^2_{\nu_2}(\lambda_2)}{\nu_2}\right\}^{-1}. \tag{30.1}$$

In many applications λ_2 is equal to zero, so there is a central χ^2 in the denominator. This might be called a "singly noncentral F," but the word "singly" is usually omitted, and it is called a ***noncentral*** F ***with ν_1, ν_2 degrees of freedom and noncentrality parameter*** λ_1. The case $\lambda_1 = 0$, $\lambda_2 \neq 0$ is not usually considered separately, since this corresponds simply to the ***reciprocal*** of a noncentral F, as just defined.

We will use the notations $F''_{\nu_1,\nu_2}(\lambda_1, \lambda_2)$ for the doubly noncentral F variable defined by (30.1) and $F'_{\nu_1,\nu_2}(\lambda_1)$ for the (singly) noncentral F variable

$$\left\{\frac{\chi'^2_{\nu_1}(\lambda_1)}{\nu_1}\right\}\left(\frac{\chi^2_{\nu_2}}{\nu_2}\right)^{-1} \tag{30.2}$$

(and also for the corresponding distributions). Note that with these symbols

$$F''_{\nu_1,\nu_2}(0, \lambda_2) = \left[F'_{\nu_2,\nu_1}(\lambda_2)\right]^{-1}.$$

In this chapter we will be mainly concerned with (singly) noncentral F-distributions. Doubly noncentral F-distributions will appear again in Section 7.

The noncentral F-distribution is used in the calculation of the power functions of tests of general linear hypotheses. As pointed out in Chapter 27, these include standard tests used in the analysis of variance such as by Tang (1938), Madow (1948), Lehmann (1959), and Scheffé (1959), to mention just a few of the earlier published studies. Later references include Cohen (1977), Fleishrnan (1980), and Cohen and Nel (1987), with emphasis on fixed effects models.

2 HISTORICAL REMARKS

The noncentral beta distribution, which is related to the noncentral F-distribution (see Section 7), was derived by Fisher (1928) in connection with research on the distribution of the multiple correlation coefficient (see Chapter 32). Its properties were discussed by Wishart (1932). The noncentral F-distribution itself was derived by Tang (1938) though Patnaik (1949) seems to have been the first to call the distribution by this name. Tang (1938) also used the doubly noncentral F-distribution (though without actually using this name) in studies of the properties of analysis of variance tests under nonstandard conditions. For a general account of the distribution and its applications to linear models, see Odeh and Fox (1975).

3 PROPERTIES

Since the numerator and denominator in (30.2) are independent, it follows that

$$\mu'_r(F'_{\nu_1,\nu_2}(\lambda_1)) = \left(\frac{\nu_2}{\nu_1}\right)^r \mu'_r\left(\chi'^2_{\nu_1}(\lambda_1)\right)\mu'_{-r}\left(\chi^2_{\nu_2}\right),$$

$$= e^{-\lambda_1/2}\left(\frac{\nu_2}{\nu_1}\right)^r \frac{\Gamma(\frac{1}{2}\nu_2 - r)}{\Gamma(\frac{1}{2}\nu_2)} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda_1)^j}{j!} \frac{\Gamma(\frac{1}{2}\nu_1 + j + r)}{\Gamma(\frac{1}{2}\nu_1 + j)},$$

whence

$$E\left[F'_{\nu_1,\nu_2}(\lambda_1)\right] = \frac{\nu_2(\nu_1 + \lambda_1)}{\nu_1(\nu_2 - 2)}, \qquad \nu_2 > 2, \qquad (30.3a)$$

$$\operatorname{var}(F'_{\nu_1,\nu_2}(\lambda_1)) = 2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1 + \lambda_1)^2 + (\nu_1 + 2\lambda_1)(\nu_2 - 2)}{(\nu_2 - 2)^2(\nu_2 - 4)}, \qquad \nu_2 > 4,$$

(30.3b)

and the third central moment is

$$\mu_3(F'_{\nu_1,\nu_2}(\lambda_1)) = \frac{4}{(\nu_2-4)(\nu_2-6)} \times\left[4\left(\frac{\nu_1+\lambda_1}{\nu_2-2}\right)^3 + \frac{6(\nu_1+\lambda_1)(\nu_1+2\lambda_1)}{(\nu_2-2)^2} + \frac{2(\nu_1+3\lambda_1)}{\nu_2-2}\right], \qquad \nu_2 > 6. \quad (30.3c)$$

Pearson and Tiku (1970) provided alternative formulas, introducing the symbol $A, = \lambda_1/\nu_1$:

$$\mu'_1 = \frac{\nu_2}{(\nu_2-2)}(1+\Lambda_1), \qquad (30.3a)'$$

$$\mu_2 = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}\left\{1+2\Lambda_1+\frac{\nu_1}{(\nu_1+\nu_2-2)}\Lambda_1^2\right\}, \qquad (30.3b)'$$

$$\mu_3 = \frac{8\nu_2^3(\nu_1+\nu_2-2)(2\nu_1+\nu_2-2)}{\nu_1^2(\nu_2-2)^3(\nu_2-4)(\nu_2-6)} x\left\{1+3\Lambda_1+\frac{6\nu_1}{(2\nu_1+\nu_2-2)}\Lambda_1^2+\frac{2\nu_1^2}{(\nu_1+\nu_2-2)(2\nu_1+\nu_2-2)}\Lambda_1^3\right\} \qquad (30.3c)'$$

$$\mu_4 = \frac{12\nu_2^4(\nu_1+\nu_2-2)}{\nu_1^3(\nu_2-2)^4(\nu_2-4)(\nu_2-6)(\nu_2-8)} \times\Big[\{2(3\nu_1+\nu_2-2)(2\nu_1+\nu_2-2) + (\nu_1+\nu_2-2)(\nu_2-2)(\nu_1+2)\}(1+4\Lambda_1) + 2\nu_1(3\nu_1+2\nu_2-4)(\nu_2+10)\Lambda_1^2 + 4\nu_1^2(\nu_2+10)\Lambda_1^3 + \frac{\nu_1^3(\nu_2+10)}{(\nu_1+\nu_2-2)}\Lambda_1^4\Big]. \qquad (30.3d)$$

Reciprocal moments of $F'_{\nu_1,\nu_2}(\lambda_1)$ are as follows:

For $\nu_1 > 2r$ and even,

$$E\left[\{F'_{\nu_1,\nu_2}(\lambda_1)\}^{-r}\right] = \left(\frac{\nu_1}{\nu_2}\right)^r(-1)^{r-(\nu_1/2)}\frac{\Gamma(\frac{1}{2}\nu_2+r)}{\Gamma(r)\Gamma(\frac{1}{2}\nu_2)}$$
$$\times\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\frac{1}{2}\lambda_1\right)^{s-(\nu_1/2)+1}\Gamma\left(\frac{1}{2}\nu_1-s-1\right)$$
$$\times\left\{e^{-\lambda_1/2}-\sum_{t=0}^{(\nu_1/2)-s-2}\frac{(-\frac{1}{2}\lambda_2)^t}{t!}\right\}. \tag{30.4a}$$

For $\nu_1 > 2r$ and odd,

$$E\left[\{F'_{\nu_1,\nu_2}(\lambda_1)\}^{-r}\right] = \left(\frac{\nu_1}{\nu_2}\right)^r(-1)^{r-\frac{1}{2}(\nu_1-1)}\frac{\Gamma(\frac{1}{2}\nu_2+r)}{\Gamma(r)\Gamma(\frac{1}{2}\nu_2)}$$
$$\times\sum_{s=0}^{r-1}\binom{r-1}{s}\left(\frac{1}{2}\lambda_1\right)^{s-(\nu_1/2)+1}\Gamma\left(\frac{1}{2}\nu_1-s-1\right)$$
$$\times\left\{2\pi^{-1/2}D\left(\left(\frac{1}{2}\lambda_1\right)^{1/2}\right)-\left(\frac{1}{2}\lambda_1\right)^{1/2}\sum_{t=0}^{\frac{1}{2}(\nu_1-5)-s}\frac{(-\frac{1}{2}\lambda_1)^t}{\Gamma(t+\frac{3}{2})}\right\} \tag{30.4b}$$

where

$$D(y) = e^{-y^2}\int_0^y e^{u^2}\,du$$

(the **Dawson** integral) [Bock, Judge, and Yancey (1984)]. (For $\nu_1 \le 2r$, the reciprocal moment is infinite.)

We also note the characteristic function

$$e^{-\lambda_1/2}\sum_{j=0}^{\infty}\frac{(\frac{1}{2}\lambda_1)^j}{j!}\,{}_1F_1\left(\frac{1}{2}\nu_1+j;\,-\frac{\nu_2}{2};\,-\frac{\nu_2 it}{\nu_1}\right). \tag{30.5}$$

Recall from Chapter 29 that the distribution of $\chi'^2_{\nu_1}(\lambda_1)$ can be represented as a mixture of central $\chi^2_{\nu_1+2j}$ distributions in proportions $e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j/j!$ $(j = 0, 1, 2, \dots)$. So

$$G'_{\nu_1,\nu_2}(\lambda_1) = \frac{\chi'^2_{\nu_1}(\lambda_1)}{\chi^2_{\nu_2}} \tag{30.6}$$

is distributed as a mixture of central G_{ν_1+2j,ν_2} distributions [as defined in Chapter 27, Eq. (27.3)] in proportions $e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j/j!$ ($j = 0, 1, 2, \dots$). Hence the probability density function of $G'_{\nu_1,\nu_2}(\lambda_1)$ is (we use G' for convenience)

$$p_{G'}(g) = \sum_{j=0}^{\infty}\left(\frac{(\frac{1}{2}\lambda_1)^j}{j!}e^{-\lambda_1/2}\right)\frac{g^{(\nu_1/2)+j-1}}{B(\frac{1}{2}\nu_1+j,\frac{1}{2}\nu_2)(1+g)^{\frac{1}{2}(\nu_1+\nu_2)+j}}$$

$$= \frac{e^{-\lambda_1/2}}{B(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2)}\cdot\frac{g^{(\nu_1/2)-1}}{(1+g)^{(\nu_1+\nu_2)/2}}\sum_{j=0}^{\infty}\left[\frac{\frac{1}{2}\lambda_1 g}{1+g}\right]^j$$

$$\times\frac{(\nu_1+\nu_2)(\nu_1+\nu_2+2)\cdots(\nu_1+\nu_2+2\cdot\overline{j-1})}{j!\nu_1(\nu_1+2)\cdots(\nu_1+2\cdot\overline{j-1})},$$

$$0 < g. \quad (30.7)$$

The probability density function of $F'_{\nu_1,\nu_2}(\lambda_1) = (\nu_2/\nu_1)G'_{\nu_1,\nu_2}(\lambda_1)$ is (using now the contraction F' for convenience)

$$p_{F'}(f) = \frac{e^{-\lambda_1/2}\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{B(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2)}\cdot\frac{f^{(\nu_1/2)-1}}{(\nu_2+\nu_1 f)^{(\nu_1+\nu_2)/2}}\sum_{j=0}^{\infty}\left(\frac{\frac{1}{2}\lambda_1\nu_1 f}{\nu_2+\nu_1 f}\right)^j$$

$$\times\frac{(\nu_1+\nu_2)(\nu_1+\nu_2+2)\cdots(\nu_1+\nu_2+2\cdot\overline{j-1})}{j!\nu_1(\nu_1+2)\cdots(\nu_1+2\cdot\overline{j-1})}$$

$$= p_{F_{\nu_1,\nu_2}}(f)e^{-\lambda_1/2}\sum_{j=0}^{\infty}\left\{\left[\frac{\frac{1}{2}\lambda_1\nu_1 f}{\nu_2+\nu_1 f}\right]^j\right.$$

$$\left.\times\frac{(\nu_1+\nu_2)(\nu_1+\nu_2+2)\cdots(\nu_1+\nu_2+2\cdot\overline{j-1})}{j!\nu_1(\nu_1+2)\cdots(\nu_1+2\cdot\overline{j-1})}\right\}, \quad (30.8)$$

where $p_{F_{\nu_1,\nu_2}}(f)$ is the density function of the central F-distribution with ν_1, ν_2 degrees of freedom. Note that while

$$p_{G'}(g) = \sum_{j=0}^{\infty}\left[\frac{e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j}{j!}\right]p_{G_{\nu_1+2j,\nu_2}}(g),$$

it is not true that

$$p_{F'}(f) = \sum_{j=0}^{\infty}\left[\frac{e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j}{j!}\right]p_{F_{\nu_1+2j,\nu_2}}(f).$$

The noncentral F density for $\nu_1 = \nu_2 = 1$ is

$$p_{F'}(f) = \frac{e^{-\lambda/2}}{\pi} f^{-1/2}(1+f)^{-1} {}_1F_1\left(1; \frac{1}{2}; c\right)$$
$$= \left(e^{-\lambda/2} f^{-1/2}(1+f)^{-1}\lambda^{-1}\right)\left\{1 + 2e^c c^{1/2} D(c^{1/2})\right\}, \quad (30.9)$$

where $c = (\lambda/2)f(1+f)^{-1}$ and $D(\cdot)$ is the Dawson integral.

The cumulative distribution can be expressed in terms of an infinite series of multiples of incomplete beta function ratios:

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le f\right] = \Pr\left[G'_{\nu_1,\nu_2}(\lambda_1) \le \frac{\nu_1}{\nu_2} f\right]$$
$$= \sum_{j=0}^{\infty}\left(\frac{\left[\frac{1}{2}\lambda_1\right]^j}{j!} e^{-\lambda_1/2}\right) \cdot I_{\nu_1 f/(\nu_2+\nu_1 f)}\left(\tfrac{1}{2}\nu_1 + j, \tfrac{1}{2}\nu_2\right), \quad (30.10)$$

where $I_p(a,b)$ is the incomplete beta function ratio given by $I_p(a,b) = \int_0^p t^{a-1}(1-t)^{b-1}\,dt/B(a,b)$. Since it is possible to express the incomplete beta function ratio in several different ways (Chapter 1), there is a corresponding range of different expressions for the cumulative distribution function of noncentral F. For the special case where ν_2 is an even integer, there are some quite simple expressions in finite terms. Sibuya (1967) pointed out that these can all be obtained by using the formal identity

$$\sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\lambda_1\right)^j}{j!} e^{-\lambda_1/2} h(j) = \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\lambda_1\right)^j}{j!} \Delta^j h(0)$$

[cf. Eq. (29.12)′, Chapter 29], (30.11)

with $h(\cdot)$ an incomplete beta function ratio, together with recurrence relationships satisfied by this function.

In particular, Sibuya (1967) showed that (if ν_2 is an even integer)

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le f\right]$$
$$= e^{-(\lambda_1/2)(1-Y)} \sum_{i=0}^{(\nu_2/2)-1} \frac{\left[\frac{1}{2}\lambda_1(1-Y)\right]^i}{i!} I_Y\left(\frac{1}{2}\nu_1 + i, \frac{1}{2}\nu_2 - i\right), \quad (30.12)$$

where $Y = \nu_1 f/(\nu_2 + \nu_1 f)$.

Replacing $I_Y(\frac{1}{2}\nu_1 + i, \frac{1}{2}\nu_2 - i)$ by a polynomial, we obtain a formula given by Seber (1963):

$$Y^{\nu_1/2}\left\{\exp\left[-\frac{1}{2}\lambda_1(1-Y)\right]\right\}\sum_{i=0}^{(\nu_2/2)-1}(1-Y)^i$$
$$\times\sum_{j=0}^{i}\left\{\binom{i+(\nu_1/2)-1}{i-j}\frac{\left[\frac{1}{2}\lambda_1 Y\right]^j}{j!}\right\}$$
$$= Y^{\nu_1/2}\left\{\exp\left[-\frac{1}{2}\lambda_1(1-Y)\right]\right\}\sum_{i=0}^{(\nu_2/2)-1}T_i, \qquad (30.13)$$

where

$$T_{-1} = 0,$$
$$T_0 = 1,$$
$$T_i = i^{-1}(1-Y)\left\{\left(2i-2+\tfrac{1}{2}\nu_1+\tfrac{1}{2}\lambda_1 Y\right)T_{i-1} - \left(i+\tfrac{1}{2}\nu_1-2\right)(1-Y)T_{i-2}\right\},$$
$$i = 1, 2, \ldots, \tfrac{1}{2}\nu_2 - 1.$$

This formula was obtained (in slightly different form) by Nicholson *(1954)* and Hodges (1955), though these authors did not give the recurrence formula for T_i. An expression of similar type, given by Wishart *(1932)* and Tang (1938), is

$$Y^{\frac{1}{2}(\nu_1+\nu_2)-1}e^{-(\lambda_1/2)(1-Y)}\sum_{i=0}^{(\nu_2/2)-1}T_i', \qquad (30.14)$$

where

$$T'_{-1} = 0,$$
$$T'_0 = 1,$$
$$T'_i = i^{-1}(Y^{-1}-1)\left\{\left[\tfrac{1}{2}(\nu_1+\nu_2) - i + \tfrac{1}{2}\lambda_1 Y\right]T'_{i-1} + \tfrac{1}{2}\lambda_1(1-Y)T'_{i-2}\right\}$$
$$i = 1, 2, \ldots, \tfrac{1}{2}\nu_2 - 1.$$

Equations *(30.12)–(30.14)* apply only where ν_2 is an even integer. Price *(1964)* obtained some finite expressions which are applicable when ν_2 is an odd integer. These are rather complicated, and are not reproduced here. If formula *(30.11)* be applied directly to (30.10), the following ***infinite*** series expansion, valid for general ν_2, is obtained:

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le f\right] = I_Y\left(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2\right) - \frac{\Gamma\left(\frac{1}{2}(\nu_1+\nu_2)\right)}{\Gamma\left(\frac{1}{2}\nu_1+1\right)} Y^{\nu_1/2}(1-Y)^{\nu_2/2} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\Delta^{j-1}t_j, \tag{30.15}$$

where

$$t_1 = 1,$$

$$t_{j+1} = \left(\tfrac{1}{2}\nu_1 + j\right)^{-1}\left[\tfrac{1}{2}(\nu_1+\nu_2) + j - 1\right]Yt_j.$$

This can be expressed in terms of generalized Laguerre polynomials [see Tiku (1965a)]. Tiku also obtained a more complicated but more rapidly convergent expansion than (30.15):

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le f\right] = I_{Y'}\left(\tfrac{1}{2}a, \tfrac{1}{2}\nu_2\right) + \sum_{j=3}^{\infty} \frac{(-1)^j}{j!} b_j \frac{Y'^{a/2}(1-Y')^{\nu_2/2}}{B\left(\frac{1}{2}a, \frac{1}{2}\nu_2\right)} \Delta^{j-1}t_j, \tag{30.16}$$

where the t_j's are as defined in (30.15),

$$a = (\nu_1+\lambda_1)^2(\nu_1+2\lambda_1)^{-1},$$

$$Y' = 1 - \left[1 + \frac{\nu_1(\nu_1+\lambda_1)}{\nu_2(\nu_1+2\lambda_2)}f\right]^{-1},$$

$$b_3 = 2\lambda_1^2(\nu_1+2\lambda_1)^{-2},$$

$$b_4 = 6\lambda_1^2(\nu_1+4\lambda_1)(\nu_1+2\lambda_1)^{-3},$$

$$b_5 = 24\lambda_1^2(\nu_1+6\nu_1\lambda_1+11\lambda_1^2)(\nu_1+2\lambda_1)^{-4}, \cdots .$$

Note that the b's do not depend on ν_2.

As would be intuitively be expected, it can be shown that $\Pr[F'_{\nu_1,\nu_2}(\lambda_1) \le f]$ is a decreasing function of λ_1. The probability density function is unimodal. As ν_2 tends to infinity the distribution of $F'_{\nu_1,\nu_2}(\lambda_1)$ approaches that of $\nu_1^{-1} \times$ (noncentral χ^2 with ν_1 degrees of freedom and noncentrality parameter λ_1). Also, as λ_1 tends to zero, the distribution of course tends to the (central) F_{ν_1,ν_2}-distribution.

Alternative expressions for the cumulative distribution function and the moment generating function were given by Venables (1975).

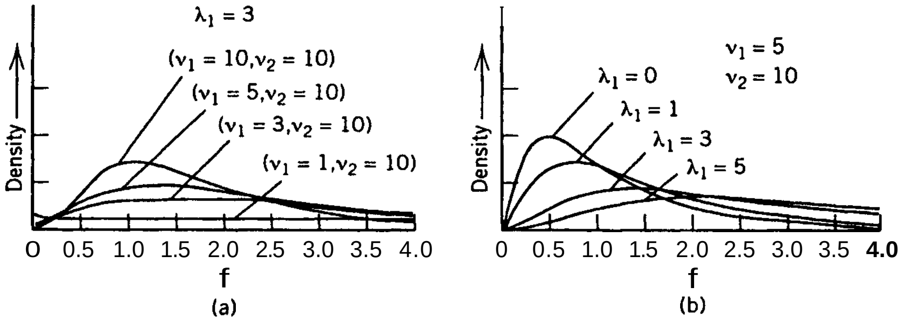


Figure 30.1a, b Noncentral F density functions

Narula and Levy *(1975)* provided plots (shown in Figure 30.1*a*) of noncentral F density for $\lambda_1 = 3$ and $(\nu_1, \nu_2) = (10, 10), (5, 10), (3, 10)$, and $(1, 10)$. As the number of degrees of freedom in the numerator decreases, the curves become flatter with the median, mean and mode shifting to the right. Another diagram (shown in Figure 30.1*b*) provided by Narula and Levy *(1975)* contains a series of noncentral F density curves for $(\nu_1, \nu_2) = (5, 10)$ as the noncentrality parameter increases from $\lambda_1 = 0$ to $\lambda_1 = 1$, $\lambda_1 = 3$, and $\lambda_1 = 5$. Here also the curves become flatter for larger values of λ_1 and similar shift of location characteristics to the right takes place.

4 TABLES AND COMPUTER PROGRAMS

4.1 Tables

The earliest tables to be published were those of Tang *(1938)*. These tables were motivated by the calculation of power functions of variance ratio tests, and they give values of $\Pr[F'_{\nu_1,\nu_2}(\lambda_1) > F_{\nu_1,\nu_2,\alpha}]$ to three decimal places for $\alpha = 0.95, 0.99$; $\nu_1 = 1(1)8$; $\nu_2 = 2(2)6(1)30, 60, \infty$, and $\sqrt{\lambda_1/(\nu_1 + 1)} = 1.0(0.5)3.0(1)8$.

The latter table is reproduced in a number of textbooks. It has been extended by Lachenbruch *(1966)* with values of the probability to four decimal places for the same values of α and $\nu_1 = 1(1)12(2)16(4)24, 30(10)50, 75$; $\nu_2 = 2(2)20(4)40(10)80$; and $\sqrt{\lambda_1/(\nu_1 + 1)} = 1.0(0.5)3.0(1)8$. Lachenbruch also gave tables of percentiles $F'_{\nu_1,\nu_2,\alpha}(\lambda_1)$ of the noncentral F-distribution for $\lambda_1 = 2(2)20$; $\alpha = 0.01, 0.025, 0.05, 0.1, 0.5, 0.9, 0.95, 0.975, 0.99$; $\nu_1 = 1(1)10, 15, 20, 30, 50, 60, 120$; and $\nu_2 = 2(2)10(10)40, 60$. Values are generally to four decimal places, except for $\nu_1 = 1$ or $\nu_2 \geq 30$, when they are only to three significant figures. Only three significant figures are also given for $\nu_1 = 2, 3, 4$, with $\nu_2 = 2$; $\nu_1 = 2, 3$, with $\nu_2 = 4$; $\nu_1 = 2$ with $\nu_2 = 6$ and $\nu_1 = 120$ with $\nu_2 = 30$.

Pearson and Hartley (1951) gave graphical representations of Tang's tables, in the form of power functions of analysis of variance tests. Patnaik (1949) has published a chart showing the relations among A,, ν_1, and ν_2 implied by the constraints

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \geq F_{\nu_1,\nu_2,\alpha}\right] = \beta \tag{30.17}$$

for a = 0.95, $\beta = 0.5$ or 0.9. Fox (1956) gave charts showing contours of $\phi = \sqrt{\lambda/(\nu_1 + 1)}$ in the (ν_1, ν_2) plane when (30.17) is satisfied, for a = 0.95, 0.99, and $\beta = 0.5(0.1)0.9$.

Lehmer (1944) gave values of ϕ to three decimal places for a = 0.95, 0.99; $\beta = 0.2$, 0.3, with $\nu_1 = 1(1)9$ and $120/\nu_1 = 1(1)6(2)12$, $\nu_2 = 2(2)18$ and $240/\nu_2 = 1(1)4(2)12$. There is also a table of this kind published by Ura (1954). It gives $\sqrt{\lambda/\nu_1}$ to two decimal places for a = 0.95, $\beta = 0.90$, with $\nu_1 = 1(1)9$ and $120/\nu_1 = 0(1)6(2)12$, $\nu_2 = 2(2)18$ and $120/\nu_2 = 0(1)6$.

Tiku (1967) gave values of $\Pr[F'_{\nu_1,\nu_2}(\lambda_1) > F_{\nu_1,\nu_2,1-\alpha}]$ to four decimal places for a = 0.005, 0.01, 0.025, 0.05; $\nu_1 = 1(1)10$, 12; $\nu_2 = 2(2)30$, 40, 60, 120, ∞; $[\lambda_1/(\nu_1 + 1)]^{1/2} = 0.5(0.5)3.0$.

4.2 Computer Programs

Bargmann and Ghosh (1964) reported a FORTRAN program that computes the probability density and cumulative distribution functions of noncentral F-distributions. Fleishman (1980) reported on a program with similar capabilities, for which he used Venables (1975) representation. Several effective and easy-to-use recursive algorithms for evaluating incomplete beta function ratios (Chapter 25) and noncentral beta distribution functions can be applied. The first published algorithms specifically for computing noncentral F (or, equivalently, beta distribution functions) are due to Norton (1983), followed by Schader and Schmid (1986) and Lenth (1987). These algorithms all involve "accurate routines" for computing incomplete beta function ratios, notably an early example constructed by Majumder and Bhattacharjee (1973).

All of the algorithms actually use the incomplete noncentral beta function ratio (see Section 7)

$$I_x(a, b; \lambda_1) = \sum_{j=0}^{\infty} \left[e^{-\lambda_1/2} \frac{(\frac{1}{2}\lambda_1)^j}{j!} \right] I_x(a + j, b), \tag{30.18}$$

where $I_x(a, b)$ is the usual incomplete beta function ratio [Chapter 1, Eq. (1.91)]. Some of the algorithms have different error bounds for the difference, E,, between the exact value and the value obtained by truncating the summation in (30.18) at j = r; some have different procedures for evaluating the incomplete beta function.

Norton (1983) used the bound

$$(0<)E_r < \left[\frac{1}{(r+1)!}\left(\frac{1}{2}\lambda_1\right)^{r+1}\right] I_x(a+r+1,b)\min\left\{1, e^{-\lambda_1/2} + \frac{\lambda_1}{2r+4}\right\} \tag{30.19a}$$

This bound was previously used by Guenther (1978). Lenth (1987) improved this bound to

$$(0<)E_r < I_x(a+r+1,b)\left[1 - \sum_{j=0}^{r}\left\{e^{-\lambda_1/2}\frac{(\frac{1}{2}\lambda_1)^j}{j!}\right\}\right] \tag{30.19b}$$

and Wang (1992) increased the lower bound to $e^{-\lambda/2}(\frac{1}{2}\lambda_1)^{r+1}/(r+1)!$

Schader and Schmid (1986) and Lenth (1987) avoided the expenditure of time necessary to calculate the quantities $I_x(a+j,b)$ in (30.18) for $j = 0, 1, 2, \ldots,$ by repeated use of the relation

$$I_x(a+1,b) = I_x(a,b) - \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)}x^a(1-x)^b$$

[cf. Chapter 25, Eq. (25.72a)] (30.20)

[as well as, of course, $\Gamma(a+1) = a\Gamma(a)$]. Frick (1990) has commented on Lenth's work, noting that good results can be expected for λ_1 small, but for λ_1 large, a large number of terms must be used in the summation in (30.18) (i.e., r must be large). Frick suggested omitting the first s terms in the summation as well as those with j greater than r. The additional error arising from omitting the first s terms is bounded by

$$I_x(a,b)\Phi\left(\sqrt{\frac{2}{\lambda_1}}\left(s - \frac{1}{2}\lambda_1\right)\right) \le \Phi\left(\sqrt{\frac{2}{\lambda_1}}\left(s - \frac{1}{2}\lambda_1\right)\right), \tag{30.21}$$

since

$$\Pr[Y \le k] \le \Phi\left(\frac{k-\theta}{\sqrt{\theta}}\right)$$

if Y has a Poisson (θ) distribution (see Chapter 4). To ensure that the introduced error is less than a specified amount δ, we have to take

$$s \le \max\left(\frac{1}{2}\lambda_1 - U_\delta\sqrt{\frac{\lambda_1}{2}}, 0\right), \tag{30.22}$$

Table 30.1 Values of r + 1 needed to achieve error bounds 10^{-8} and 10^{-10}

Error Bound	Value of A,								
	0.5	1.0	2.0	4	6	8	10	15	20
	6	8	11	14	17	20	22	27	32
10^{-10}	8	10	12	16	19	22	25	31	36

where $\Phi(U_\delta) = 6$. For example, taking $U_\delta = -5$, we have $\delta = 0.0000003$. Then $s \le \frac{1}{2}\lambda_1 - 5\sqrt{\lambda_1/2}$ is clearly adequate.

Apart from this modification, Lenth's algorithm has been somewhat elaborated by Posten (1993) who applied a similar technique to one he used with noncentral chi-square distributions, see Chapter 29. Posten suggested starting from a value for j approximately equal to $\frac{1}{2}\lambda_1$ and then working "outward" (increasing and decreasing j) until the sum of the Poisson multipliers

$$P(r,s) = \sum_{j=s}^{r} \left\{ e^{-\lambda_1/2} \frac{(\frac{1}{2}\lambda_1)^j}{j!} \right\} \tag{30.23}$$

is sufficiently close to 1. If $P(r,s) \ge 1 - \epsilon$, then the error in using $\Sigma_{j=s}^{r}$ in place of $\Sigma_{j=0}^{\infty}$ in (30.18) is not greater than ϵ.

Lee (1992) noted that the error in taking $\Sigma_{j=0}^{r}$ in place of $\Sigma_{j=0}^{\infty}$ in (30.18) is bounded by (30.19b) and he provided a table (Table 30.1) that gives the number of terms $(r + 1)$ needed to achieve various error bounds.

The method for evaluating $I_x(a, b;$ A,) used in IMSL (1987) for 0.5 < A, < 20 when $\max(a, b) < 200$ is less than half as efficient (in CPU time) as Lee's. However, Lee's approach requires evaluation of $r + 1$ incomplete beta function ratios, whereas Posten's and Lenth's approaches need only one such computation. Singh and Relyea (1992) followed similar lines to Lenth (1987) and Posten (1993), and they used Guenther's (1978) error bound (30.19a). They differed in using explicit expressions for the incomplete beta function ratios.

As the preceding discussion illustrates—even without technical details—clear examples of multiplication and overlap of results are prevalent in the statistical literature, especially in the area of statistical algorithms. Much of it is due to lack of coordination, almost identical publication in different journals, and unjustified publication of results providing "epsilon" improvement.

5 APPROXIMATIONS

From formula (30.2) we can see that approximations to noncentral χ^2-distributions can be used to give approximations to noncentral F-distributions. Thus the simple approximation to the distribution of $\chi'^2_{\nu_1}(\lambda_1)$ by that of $c\chi^2_{\nu}$, with $c = (\nu_1 + 2\lambda_1)(\nu_1 + \lambda_1)^{-1}$; $\nu = (\nu_1 + \lambda_1)^2(\nu_1 + 2\lambda_1)^{-1}$ gives an approximation to the distribution of $F'_{\nu_1,\nu_2}(\lambda_1)$ by that of $(c\nu/\nu_1)F_{\nu,\nu_2} = (1 + \lambda_1\nu_1^{-1})F_{\nu,\nu_2}$. (Note the need for the factor ν/ν_1.) The accuracy of this approximation has been studied by Patnaik (1949). Of course the distribution of F_{ν,ν_2} itself may also be approximated by one of the methods described in Chapter 27, leading to a composite approximation to the distribution of $F'_{\nu_1,\nu_2}(\lambda_1)$. Thus, using Paulson's approximation, Severo and Zelen (1960) were led to suggest

$$\frac{\left(1 - \dfrac{2}{9\nu_2}\right)[\nu_1 F'/(\nu_1 + \lambda_1)]^{1/3} - \left[1 - [2(\nu_1 + 2\lambda_1)/9](\nu_1 + \lambda_1)^{-2}\right]}{\left\{[2(\nu_1 + 2\lambda_1)/9](\nu_1 + \lambda_1)^{-2} + \left(2\nu_2^{-1}/9\right)\left[\dfrac{\nu_1 F'}{(\nu_1 + \lambda_1)}\right]^{2/3}\right\}^{1/2}} \tag{30.24}$$

as approximately having a unit normal distribution. Laubscher (1960) also independently derived this result and compared it with that obtained by using Fisher's square root approximations to the distributions of the χ^2's and F_{ν_1,ν_2}, namely

$$(2\nu_2 - 1)^{1/2}\left(\frac{\nu_1 F'}{\nu_2}\right)^{1/2} - \left\{\frac{2(\nu_1 + \lambda_1) - (\nu_1 + 2\lambda_1)}{\nu_1 + \lambda_1}\right\}^{1/2}$$
$$\times\left\{\frac{\nu_1 F'}{\nu_2} + \frac{\nu_1 + 2\lambda_1}{\nu_1 + \lambda_1}\right\}^{-1/2} \tag{30.25}$$

approximately unit normal. Laubscher (1960) compared values of $\Pr[F'_{\nu_1,\nu_2}(\lambda_1) < f]$ as approximated by (30.24) and (30.25) with the exact values for the choices of parameters

ν_1	ν_2	λ_1
3	10, 20	4, 16
5	10, 20	6, 24
8	10, 30	9, 36

with $f = F_{\nu_1,\nu_2,\alpha}$; $\alpha = 0.95, 0.99$.

Despite the fact that the Wilson-Hilferty transformation, on which (30.24) is based, is generally more accurate than Fisher's transformation, on which (30.25) is based, the latter approximation is slightly more accurate than the former. The position was reversed for the larger values of f and λ_1.

However, a similar comparison, carried out by Fowler (1984) for $\nu_1 = 1(1)6$, 8, 12, 24, and $\nu_2 = 6(2)30, 40, 60, 120$, and 240, with $\lambda_1^2/\nu_1 = 0.1$, 0.25, 0.4, and he found that (30.24) was generally superior, at least in the lower tail

$[F(f;\nu_1,\nu_2,\lambda_1) \le 0.5]$ for $\nu_1 \le 6$. [Cohen (1977) unfortunately relied more often on (30.25) in his power analyses because it appeared to be somewhat better than (30.24) except for small ν_1, ν_2, and λ_1. See also Cohen and Nel (1987).]

Laubscher (1960) considered

$$\left(\frac{1}{2}\nu_2 - 2\right)^{1/2} \cosh^{-1}\left[\frac{\nu_1(\nu_2-2)^{1/2}(F' + (\nu_2/\nu_1))}{\nu_2(\nu_1+\nu_2-2)^{1/2}}\right] \qquad (30.26)$$

as a possible transformation to unit normal—based on approximate variance-equalizing property—but the result was accurate only for very large λ_1. Tiku (1966) obtained quite good results by fitting the distribution of $F'_{\nu_1,\nu_2}(\lambda_1)$ by that of $(b + cF'_{\nu,\nu_2})$, choosing b, c, and ν so as to make the first three moments agree. The values that do this are

$$\nu = \frac{1}{2}(\nu_2-2)\left[\sqrt{\frac{H^2}{H^2-4K^3}} - 1\right],$$

$$c = \left(\frac{\nu'_1}{\nu_1}\right)(2\nu'_1 + \nu_2 - 2)^{-1}\left(\frac{H}{K}\right),$$

$$b = -\nu_2(\nu_2-2)^{-1}(c - 1 - \lambda_1\nu_1^{-1}), \qquad (30.27)$$

where

$$H = 2(\nu_1+\lambda_1)^3 + 3(\nu_1+\lambda_1)(\nu_1+2\lambda_1)(\nu_2-2) + (\nu_1+3\lambda_1)(\nu_2-2)^2,$$

$$K = (\nu_1+\lambda_1)^2 + (\nu_2-2)(\nu_1+2\lambda_1).$$

On the other hand, Mudholkar, Chaubey, and Lin (1976) first approximated the equivalent degrees of freedom (ν) for $\chi'^2_{\nu_1}(\lambda_1)$ by

$$\nu = \frac{(\nu_1+2\lambda_1)^3}{(\nu_1+3\lambda_1)^2} \qquad \text{[cf. Chapter 29, Eq. (29.60)]}, \qquad (30.28a)$$

and then they chose c and b to give the correct first two moments of $F'_{\nu_1,\nu_2}(\lambda_1)$, leading to

$$\begin{cases} c = \left(\dfrac{\nu}{\nu_1}\right)\{\nu^2 + (\nu_2-2)\}^{-1/2} \left\{(\nu_2-2)(\nu_1+2\lambda_1) + (\nu_1+\lambda_1)^2\right\}^{1/2}, \\ b = -\nu_2(\nu_2-2)^{-1}(c-1-\lambda_1\nu_1^{-1}). \end{cases}$$

(30.28b)

This appeared to be better for the right tail (larger values of f) than Tiku's approximation (30.271, but the position was reversed in the left tail (lower values). Generally the two approximations were of comparable accuracy.

Tiku (1966) found that his approximation [with b, c, and v given by (30.27)] was better than Severo and Zelen's (30.24) or Patnaik's. Of the latter two, Severo and Zelen's seems to be the easier one to compute, although slightly less accurate than Patnaik's for large ν_2.

Pearson and Tiku (1970) analyzed the relation between noncentral and central F-distributions by plotting the (β_1, β_2) for the two classes of distributions. They discovered that "(a) for a given ν_2, the (β_1, β_2) points for the distribution of central F appeared to lie very close to a straight line; (b) again, for the same ν_2, the beta points for $F'_{\nu_1,\nu_2}(\lambda_1)$ lay very nearly on the line (a), converging on to it as either ν_1 or A, were increased."

They noted that Tiku's three-moment approximation (30.27) involves the use of the correct β_1 but an incorrect β_2. The differences in the values of β_2 become very small as either $\phi = \sqrt{\lambda/(\nu_2 + 1)}$ or ν_1 increases. Errors of 0.15 or less in β_2 will rarely affect the position of upper percentage points by more than 1/100 of the standard deviation, but they do affect the lower percentage points rather substantially, by as much as 3/100 or 4/100 of the standard deviation at the 0.5 and 1.0% points. (A better approximation in the lower tail could probably be obtained by using a distribution having the correct first *four* moments when value of β_1 is greater than 4.)

By analogy with the central F-distribution, one might expect to obtain useful approximations by considering the distribution of

$$Z'_{\nu_1,\nu_2}(\lambda_1) = \tfrac{1}{2}\log F'_{\nu_1,\nu_2}(\lambda_1) \quad \text{(i.e., the } \textit{noncentral Z distribution.)}$$

Since $Z' = Z'_{\nu_1,\nu_2}(\lambda_1) = \frac{1}{2}\log(\nu_2/\nu_1) + \frac{1}{2}\log\chi'^2_{\nu_1}(\lambda_1) - \frac{1}{2}\log\chi^2_{\nu_2}$, the cumulants of Z' are

$$\kappa_1(Z') = \frac{1}{2}\left[\log\left(\frac{\nu_2}{\nu_1}\right) + \kappa_1\left(\chi'^2_{\nu_1}(\lambda_1)\right) - \kappa_1\left(\chi^2_{\nu_2}\right)\right],$$

$$\kappa_r(Z') = 2^{-r}\left[\kappa_r\left(\chi'^2_{\nu_1}(\lambda_1)\right) + (-1)^r\kappa_r\left(\chi^2_{\nu_2}\right)\right], \qquad r \ge 2. \quad (30.29)$$

Barton, David, and O'Neill (1960) gave formulas to aid in the computation of $\kappa_r(\chi'^2_{\nu_1}(\lambda_1))$ and utilized them in calculating the power function of the F-test, by fitting Edgeworth series to the distribution of Z'.

Pearson (1960) obtained good results by fitting the distributions of Z' by S, distributions (see Chapter 12, Section 4.3). [It should be noted that Tiku (1965a) stated incorrectly that Pearson fitted F'-distributions by S, distributions.] However, the computations of the cumulants of Z' is rather laborious. In Barton, David, and O'Neill (1960) where details are given, the cumulants are expressed in terms of polygamma functions [Chapter 1, Eq. (1.39)] and special 9-functions, tabulated in the article.

Noting that

$$\Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le f\right] = \Pr\left[\left\{\nu_1^{-1}\chi'^2_{\nu_1}(\lambda_1)\right\}^{1/3} - f^{1/3}\left(\nu_2^{-1}\chi^2_{\nu_2}\right)^{1/3} \le 0\right], \quad (30.30)$$

and that $\{\chi_{\nu_1}'^2(\lambda_1)\}^{1/3}$ and $(\chi_{\nu_2}^2)^{1/3}$ may be well approximated by normal distributions, Mudholkar, Chaubey, and Lin (1976) suggested using an Edgeworth expansion for the distribution of

$$V = \left\{\nu_1^{-1}\chi_{\nu_1}'^2(\lambda_1)\right\}^{1/3} - f^{1/3}\left(\nu_2^{-1}\chi_{\nu_2}^2\right)^{1/3} \tag{30.31}$$

making use of Aty's (1954) expressions for cumulants of cube roots of noncentral chi-square variables. (They actually used the first three terms of the expressions in their calculations.)

6 ESTIMATION OF THE NONCENTRALITY PARAMETER λ_1

Considerable attention has been devoted to estimation of the noncentrality parameter λ_1 from a single observed value F' of a $F'_{\nu_1,\nu_2}(\lambda_1)$ variable (ν_1 and ν_2 being known). The major part of this section will describe results of these efforts. At the conclusion of the section, there is a short discussion of maximum likelihood estimation based on n independent observed values F'_1, $F'_2, \cdots, F'_n$.

The uniformly minimum variance unbiased estimator of λ_1 is

$$\lambda_1^* = \nu_1\nu_2^{-1}(\nu_2 - 2)F' - \nu_1 \tag{30.32}$$

[Perlman and Rasmussen (1975)]. Unfortunately, it is not always positive, and so is inadmissible. Chow (1987) showed that

$$(\lambda_1^*)_+ = \begin{cases} \nu_1\nu_2^{-1}(\nu_2 - 2)F' - \nu_1 & \text{if } F' > \nu_2(\nu_2 - 2)^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{30.33}$$

is also inadmissible. With expected square error as criterion, any estimator of form $a\{\nu_2^{-1}(\nu_2 - 2)F' - 1\}_+$ is inadmissible. The estimator $a\{\nu_2^{-1}(\nu_2 - 1)F' - 1\}$ dominates λ_1^* for all λ_1 provided that

$$\max\left(0, \frac{\nu_2 - 6}{\nu_2 - 2}\right) \le a \le 1.$$

Rukhin (1993) has investigated linear functions of F' as estimators of λ_1. For analytical purposes it is convenient to replace F' by $G' = \nu_1 F'/\nu_2$ distributed as $G'_{\nu_1,\nu_2}(\lambda_1)$. The expected quadratic error of $aG' + b$ is

$$E\left[(aG' + b - \lambda_1)^2\right] = 2(\nu_2 - 2)^{-1}(\nu_2 - 4)^{-1}\left\{\nu_1 + 2\lambda_1 + \frac{1}{2}(\nu_1 + \lambda_1)^2\right\}a^2 + 2(\nu_2 - 2)^{-1}(\nu_1 + \lambda_1)a(b - \lambda_1) + (b - \lambda_1)^2,$$
$$\nu_2 > 4. \tag{30.34}$$

Calculations show that any estimator with $a > \nu_2 - 4$ is improved by taking $a = \nu_2 - 4$. If $\nu_2 \leq 4$, the estimator $aG' + b$ has infinite expected quadratic error and is therefore inadmissible. [See also Rasmussen (1973).]

Bayesian estimator of λ_1 was discussed by Perlman and Rasmussen (1975) and DeWaal *(1974)*. If λ_1 has, as prior distribution, that of $\gamma\chi_n^2 (\gamma > 0)$, then under quadratic loss the Bayesian estimator is

$$\frac{\gamma}{1+\gamma} \cdot \frac{\gamma(\nu_1 + \nu_2)\nu_1 F'/\nu_2}{1 + \gamma + (\nu_1 F'/\nu_2)}. \tag{30.35}$$

As $\gamma \to \infty$, we obtain the improper Bayesian estimator

$$\nu_1(\nu_1 + \nu_2)\nu_2^{-1}F' + \nu_1. \tag{30.36}$$

Both *(30.35)* and *(30.36)* have larger mean square errors than

$$\nu_1(\nu_2 - 4)\{\nu_2^{-1}F' - (\nu_2 - 2)^{-1}\}. \tag{30.37}$$

Perlman and Rasmussen *(1975)* noted that any ***proper*** prior distribution (no matter how diffuse) yields a Bayes estimator that is closer to (30.37) than to either *(30.35)* or *(30.36)*. Indeed these two estimators cannot be less than ν_1, which seems very strange.

For $(\nu_1, \nu_2) \geq 5$ the estimator

$$a\nu_1\{\nu_2^{-1}(\nu_2 - 2)F' - 1\} + b\nu_1^{-1}\nu_2 F'^{-1} \tag{30.38}$$

has lower mean square error than

$$a\nu_1\{\nu_2^{-1}(\nu_2 - 2)F' - 1\} \tag{30.39}$$

if $0 < b < 4\nu_2^{-1}(\nu_2 + 2)^{-1}(\nu_1 - 4)(\nu_1 + \nu_2 - 2)a$ for all $a > 0$. Perlman and Rasmussen (1975) recommended using the values

$$a = (\nu_2 - 2)^{-1}(\nu_2 - 4),$$

$$b = 2\nu_2^{-1}(\nu_2 + 2)^{-1}(\nu_2 - 2)^{-1}(\nu_1 - 4)(\nu_2 - 4)(\nu_1 + \nu_2 - 2). \tag{30.40}$$

They also remarked that the introduction of improper prior distributions may be the reason that the estimators *(30.35)* and *(30.36)* appear to be so strange. [See also Efron *(1970,* 1973).] Gelfand *(1983)* has studied methods of searching for appropriate prior distributions.

We now consider construction of a confidence interval for λ_1, based on a single observed value F'. Venables *(1975)* suggested a method for constructing a confidence interval for λ_1 that is similar to the one he proposed for estimating the noncentrality parameter of a noncentral chi-square distribution [Chapter 29, Eqs. (29.47) et seq.]. Analogously to Eq. (29.47) of Chapter 29, he constructed a confidence distribution for λ_1, given an observed value F' of a $F'_{\nu_1,\nu_2}(\lambda_1)$ variable, with cdf

$$\Pr[F_{\nu_1,\nu_2} > F'] + p(F';\nu_1,\nu_2,0)\sum_{j=1}^{\infty}\frac{\{\frac{1}{2}(\nu_1+\nu_2)\}^{[j-1]}}{(\frac{1}{2}\nu_1)^{[j]}}\cdot\left(\frac{2}{\nu_2}\right)^{j-1} \times\frac{(\frac{1}{2}F')^j}{(1+\nu_2^{-1}F')^{j-1}}\Pr[\chi^2_{2j}\le\lambda_1]. \tag{30.41}$$

The moment-generating function corresponding to (30.41) is

$$\Pr[F_{\nu_1,\nu_2} > F'] + \left\{1-\frac{2F't}{\nu_2(1-2t)}\right\}^{-\nu_2+(1/2)}(1-2t)^{(\nu_1/2)-1} \times F\left(\frac{F'}{1-2(1+\nu_2^{-1}F')};\nu_1,\nu_2,0\right), \tag{30.42}$$

which Venables approximated by

$$\left\{1-\frac{2F't}{\nu_2(1-2t)}\right\}^{-\nu_2/2}(1-2t)^{(\nu_1/2)-1}, \tag{30.43}$$

giving approximate cumulants

$$\begin{aligned}\kappa_r^* &= 2^{r-1}(r-1)!\{\nu_2(1+\nu_2^{-1}F')^r-\nu_1-\nu_2+2\}\\ &= 2^{r-1}(r-1)!\{rF'-\nu_1+2+O(\nu_2^{-1})\}\end{aligned}$$

[cf. Eq. (29.50) of Chapter 29]. *(30.44)*

However, he did not proceed to develop a Cornish-Fisher-type expansion for quantiles (confidence limits) for A, (as he did for A in Chapter 29), but he suggested instead fitting a distribution using the first few approximate moments.

Guirguis (1990), wishing to solve the equation

$$\Pr[F'_{\nu_1,\nu_2}(\lambda_1)\le F'] = \alpha \tag{30.45}$$

for λ_1, uses an iterative method based on the formula

$$\frac{\partial \Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \le F'\right]}{\partial\lambda_1} = \frac{1}{2}\left\{\Pr\left[F'_{\nu_1+2,\nu_2}(\lambda_1) \ge \frac{\nu_1 F'}{\nu_1+2}\right] - \Pr\left[F'_{\nu_1,\nu_2}(\lambda_1) \ge F'\right]\right\}. \quad (30.46)$$

For $F' > 0$, $\partial \Pr[F'_{\nu_1,\nu_2}(\lambda_1) \le F']/\partial\lambda_1$ is negative, and there is no solution to (30.45) if $F' < F'_{\nu_1,\nu_2,\alpha}$. If this is not so, there is a unique solution of (30.45) since $\Pr[F'_{\nu_1,\nu_2}(\lambda_1) \le F'] \to 0$, as $\lambda_1 \to \infty$.

Guirguis uses a modification of linear-Newton (L-Newton) iteration, which he terms exponential-Newton (E-Newton). In seeking a solution of the equation

$$g(x) = \alpha,$$

the $(n+1)$-th iterate x^E_{n+1} is calculated as

$$x^E_{n+1} = x^E_n + \frac{g(x^E_n)}{g'(x^E_n)}\log\left(\frac{\alpha}{g(x^E_n)}\right). \quad (30.47)$$

The E-Newton method is better than the L-Newton method when a good initial guess is not available.

Lam (1987) proposes an iterative procedure for computing confidence intervals for the non-centrality parameter λ. A computer program written in Fortran 77 is available from the author.

Guirguis (1990) compared the L- and E-Newton methods with a quadratic-Newton (Q-Newton) method employed by Narula and Weistroffer (1986) with $\nu_1 = 8$, $\nu_2 = 2$, $\alpha = 0.01$, $F' = 0.5(1)9.5$, and he found that the E-Newton converges faster than L-Newton for $F' = 0.5$ and 1; both are about the same for $F' > 1$. Q-Newton is much slower than either L- or E-Newton.

If n independent random variables $G'_1,\dots,G'_n$ have a common distribution $\chi'^2_{\nu_1}(\lambda_1)/\chi^2_{\nu_2}$, with ν_1 and ν_2 known, the maximum likelihood estimator $\hat{\lambda}_1$ of λ_1 is the solution of the equation

$$n = \sum_{i=1}^{n} \frac{\nu_1 G'_i}{2 + \nu_1 G'_i} \cdot \frac{{}_2F'_0\left(\tfrac{1}{2}(\nu_1+\nu_2), \tfrac{1}{2}\nu_1; \hat{\lambda}_1\nu_1 G'_i(\nu_2+\nu_1 G'_i)^{-1}\right)}{{}_2F_0\left(\tfrac{1}{2}(\nu_1+\nu_2), \tfrac{1}{2}\nu_1; \hat{\lambda}_1\nu_1 G'_i(\nu_2+\nu_1 G'_i)^{-1}\right)}, \quad (30.48)$$

where

$${}_2F_0(a,b;x) = \sum_{j=0}^{\infty} \frac{a^{[j]}}{b^{[j]}}\frac{x^j}{j!}$$

is the confluent hypergeometric function (Chapter 1, Section A7). Pandey and Rahman (1971) proved the uniqueness of a positive solution of (30.48) provided

$$\sum_{i=1}^{n} \frac{\nu_1 F_i'}{\nu_2 + \nu_1 F_i'} > \frac{n\nu_1}{\nu_1 + \nu_2}.$$

7 RELATED DISTRIBUTIONS

7.1 Doubly Noncentral F-Distributions

The ***doubly noncentral F-distribution*** defined by (30.1) has already been noted. Using the representation of each of the noncentral χ^2's as mixtures of central χ^2 distributions, we see that

$$G''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) = \frac{\chi'^2_{\nu_1}(\lambda_1)}{\chi'^2_{\nu_2}(\lambda_2)}$$

is distributed as a mixture of G_{ν_1+2j,ν_2+2k}-distributions in proportions $\{e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j/j!\}$. $\{e^{-\lambda_2/2}(\frac{1}{2}\lambda_2)^k/k!\}$. Hence (using the contracted forms F'', G'' for the variables) the probability density function of G'' is

$$\begin{aligned} p(g;\nu_1,\nu_2;\lambda_1,\lambda_2) &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left[\frac{e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j}{j!}\right]\left[\frac{e^{-\lambda_2/2}(\frac{1}{2}\lambda_2)^k}{k!}\right]\left[B\left(\tfrac{1}{2}\nu_1+j,\frac{1}{2}\nu_2+k\right)\right]^{-1} \\ &\quad \times g^{(\nu_1/2)+j-1}(1+g)^{-\frac{1}{2}(\nu_1+\nu_2)-j-k}, \end{aligned} \tag{30.49}$$

and that of F'' $(=\nu_2 G''/\nu_1)$ is

$$\begin{aligned} p(f;\nu_1,\nu_2;\lambda_1,\lambda_2) &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left[\frac{e^{-\lambda_1/2}(\frac{1}{2}\lambda_1)^j}{j!}\right]\left[\frac{e^{-\lambda_2/2}(\frac{1}{2}\lambda_2)^k}{k!}\right]\nu_1^{(\nu_1/2)+j}\nu_2^{(\nu_2/2)+k}f^{(\nu_1/2)+j-1} \\ &\quad \times(\nu_2+\nu_1 f)^{-\frac{1}{2}(\nu_1+\nu_2)-j-k}\left[B(\tfrac{1}{2}\nu_1+j,\tfrac{1}{2}\nu_2+k)\right]^{-1} \\ &= p_{F_{\nu_1,\nu_2}}(f)\cdot\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left[\frac{e^{-\lambda_1/2}[\frac{1}{2}\lambda_1\nu_1 f/(\nu_2+\nu_1 f)]^j}{j!}\right] \\ &\quad \times\left[\frac{e^{-\lambda_2/2}[\frac{1}{2}\lambda_2\nu_2/(\nu_2+\nu_1 f)]^k}{k!}\right]\frac{B(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2)}{B(\frac{1}{2}\nu_1+j,\frac{1}{2}\nu_2+k)}. \end{aligned} \tag{30.50}$$

This result was also obtained by Malik (1970) using a Mellin transform and independently in a slightly different though equivalent form, by Bulgren (1971).

Pe and Drygas (1994) have obtained the representations

$$p(f;\nu_1,\nu_2;\lambda_1,\lambda_2) = e^{-(\lambda_1+\lambda_2)/2}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\lambda_1\right)^j}{j!}\,{}_2F_1\left(1-\tfrac{1}{2}\nu_1-j,-j;\tfrac{1}{2}\nu_2;\frac{\lambda_1\nu_1 f}{\lambda_2\nu_2}\right)$$

$$\times\left(\frac{\nu_1}{\nu_2}\right)^{(\nu_1/2)+j}\frac{f^{(\nu_1/2)+j-1}}{B\left(\frac{1}{2}\nu_1+j,\frac{1}{2}\nu_2\right)\left(1+\nu_1\nu_2^{-1}f\right)^{\frac{1}{2}(\nu_1+\nu_2)+j}}$$

$$= e^{-(\lambda_1+\lambda_2)/2}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\lambda_1\right)^j}{j!}\sum_{k=0}^{j}\left(\frac{\lambda_1}{\lambda_2}\right)^k\binom{j}{k}\left(\frac{\nu_1}{\nu_2}\right)^{(\nu_1/2)+j+k}$$

$$\times\frac{f^{(\nu_1/2)+j+k-1}}{B\left(\frac{1}{2}\nu_1+j-k,\frac{1}{2}\nu_2+k\right)\left(1+\nu_1\nu_2^{-1}f\right)^{\frac{1}{2}(\nu_1+\nu_2)+j}},$$

$$0<f. \qquad (30.50)'$$

The cdf of G" is of course

$$\Pr\left[G''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)\le g\right]$$

$$= e^{-(\lambda_1+\lambda_2)/2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\lambda_1\right)^j\left(\frac{1}{2}\lambda_2\right)^k}{j!k!}I_{g/(1+g)}\left(\frac{1}{2}\nu_1+j,\frac{1}{2}\nu_2+k\right). \qquad (30.51)$$

The cdf of F'', $\Pr[F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)\le f]$ is obtained by replacing g by $f\nu_1\nu_2^{-1}$ on the right-hand side of (30.51).

The rth moment about zero of F'' is

$$\mu'_r(F'') = E[F''^r]$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r e^{-(\lambda_1+\lambda_2)/2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\lambda_1\right)^j\left(\frac{1}{2}\lambda_2\right)^k}{j!k!}E\left[\left(\chi^2_{\nu_1+2j}\right)^r\right]E\left[\left(\chi^2_{\nu_2+2k}\right)^{-r}\right]$$

$$= \left(\frac{\nu_2}{\nu_1}\right)^r e^{-(\lambda_1+\lambda_2)/2}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}\lambda_1\right)^j\left(\frac{1}{2}\lambda_2\right)^k}{j!k!}$$

$$\times\frac{\Gamma\left(\frac{1}{2}\nu_1+j+r\right)\Gamma\left(\frac{1}{2}\nu_2+k-r\right)}{\Gamma\left(\frac{1}{2}\nu_1\right)\Gamma\left(\frac{1}{2}\nu_2\right)}, \qquad \nu_2>2r. \qquad (30.52)$$

[For $\nu_2 \leq 2r$, $\mu'_r(F'')$ is infinite.] Tiku (1972) expressed (30.52) in the form

$$\mu'_r(F'') = \mu'_r(F'_{\nu_1,\nu_2}(\lambda_1))M(r, \tfrac{1}{2}\nu_2; -\tfrac{1}{2}\lambda_2) \tag{30.52$'$}$$

where $M(a,b;x) = {}_2F_0(a,b;x) = \sum_{j=0}^{\infty} a^{[j]}b^{[j]}x^j/j!$ is a confluent hypergeometric function (see Chapter 1, Section A7).

Bulgren (1971) gave tables of percentile points of F'', namely values $f = F''_{\nu_1,\nu_2,\alpha}(\lambda_1,\lambda_2)$ such that

$$\Pr\left[F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) \leq f\right] = \alpha \tag{30.53}$$

for $\nu_1 = 2, 4, 8$; $\nu_2 = 4, 15, 30, 60$; $\alpha = 0.95, 0.99$; $\lambda_1, \lambda_2 = 0.5, 1.5, 2(1)6, 9, 10, 24$. Winer's (1971) textbook contains tables of doubly noncentral F-distributions.

More extensive tables were provided by Tiku (1974). These include values of f for $\nu_1 = 1(1)8, 10, 12$; $\nu_2 = 2(2)12, 16, 20, 24, 30, 40, 60$; $\alpha = 0.95, 0.99$; $\phi_1 = \{\lambda_1/(2\nu_1+1)\}^{1/2} = 0(0.5)3.0$; $\phi_2 = \{\lambda_2/(2\nu_2+1)\}^{1/2} = 0(1)8$. Another table gives values of $\Pr[F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) > f]$ for the above values of ϕ_1 and ϕ_2; $\nu_1 = \nu_2 = 4(2)12$ and $(1+\nu_1 f/\nu_2)^{-1} = 0.02(0.08)0.50, 0.60, 0.75, 0.95$.

Tiku (1972) had investigated an approximation to the distribution of $F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)$ by that of a linear function of a central F variable. His analysis consisted of the following steps:

1. If $(\lambda_2/\nu_2) < \frac{1}{2}$, $M(r, \frac{1}{2}\nu_2; -\frac{1}{2}\lambda_2)$ converges rapidly and is approximately equal to $(1+\nu_2^{-1}\lambda_2)^{-r}$ and hence the rth moment of F'' [see (30.52)'] is approximately equal to $\mu'_r(F'_{\nu_1,\nu_2}(\lambda_1))(1+\nu_2^{-1}\lambda_2)^{-r}$.
2. In turn the distribution of $(F''+a)/h$ is approximately distributed as F_{ν,ν_2} with

$$\nu = \tfrac{1}{2}(\nu_2 - 2) + \left[\left\{1 - \frac{32(\nu_2-4)}{(\nu_2-6)^2\beta_1}\right\}^{-1/2} - 1\right]^{-1},$$

$$h = \tfrac{1}{2}\nu(\nu_2-2)(\nu_2-6)\mu_3\{\nu_2\mu_2(2\nu+\nu_2-2)\}^{-1},$$

and

$$a = h\nu_2(\nu_2-2)^{-1} - \mu'_1,$$

where μ'_1, μ_2, μ_3, and β_1 refer to the distributions of F''.

Table 30.2 is an extract from a more extensive table in Tiku (1972), showing the errors in approximate values of $\Pr[F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) > F_{\nu_1,\nu_2,0.95}]$ for various values of ν_1, ν_2, and ϕ_i. The errors become large for larger values of ϕ_2. Tiku (1972) recommends using the exact formula if $\lambda_2/\nu_2 > \frac{1}{2}$.

Table 30.2 **True values (1) and errors $\times$ 10^4 (2) in approximate values of** $\Pr[F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) > F_{\nu_1,\nu_2,0.95}]$, with $\nu_1 = 4$

	4 2	0.0		0.5		1.0		2.0		3.0	
ν_2	ϕ_1	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
8	0	0.0500	0	0.0328	−7	0.0215	−17	0.0092	−30	0.0039	−31
	1	0.2398	−1	0.2788	−18	0.1326	−52	0.0717	−119	0.0381	−15
	2	0.7714	−1	0.6886	−4	0.6070	−22	0.4562	−125	0.3301	−283
	3	0.9868	1	0.9729	−7	0.9536	28	0.8980	91	0.8222	128
24	0	0.0500	0	0.0358	−2	0.0256	−4	0.0130	−9	0.0065	−11
	1	0.3302	−1	0.2764	−1	0.2300	−6	0.1566	−28	0.1045	−52
	2	0.9192	0	0.8915	−1	0.8601	1	0.7879	7	0.7068	7
	3	0.9995	0	0.9991	0	0.9985	1	0.9963	10	0.9920	12

Approximations to the distributions of the noncentral χ^2-distributions can be used to derive approximations to the doubly noncentral F-distributions. It is of course possible to approximate the distribution of just one of the two noncentral χ^2's. Thus, if the distribution of $\chi'^2_{\nu_2}(\lambda_2)$ is approximated by that of $c'\chi^2_{\nu'}$, with $c' = (\nu_2 + 2\lambda_2)(\nu_2 + \lambda_2)^{-1}$ and $\nu' = (\nu_2 + \lambda_2)^2(\nu_2 + 2\lambda_2)^{-1}$, then the corresponding approximation to the distribution of $F''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)$ is that of $(\nu_2/(c'\nu'))F'_{\nu_1,\nu'}(\lambda_1) = (1 + \lambda_2\nu_2^{-1})^{-1}F'_{\nu_1,\nu'}(\lambda_1)$. If both numerator and denominator are approximated, the approximating distribution is that of

$$\frac{1 + \lambda_1\nu_1^{-1}}{1 + \lambda_2\nu_2^{-1}}F_{\nu,\nu'}, \tag{30.54}$$

with $\nu = (\nu_1 + \lambda_1)^2(\nu_1 + 2\lambda_1)^{-1}$; $\nu' = (\nu_2 + \lambda_2)^2(\nu_2 + 2\lambda_2)^{-1}$.

The doubly noncentral F-distribution is used when estimating the effect, on the power function of analysis of variance tests, of nonrandom effects in the residual variation. For example, in a standard one-way classification, if each individual in a group has a departure from the group mean, depending on the order of observation, then the residual (within group) sum of squares is distributed as a multiple of a noncentral rather than a central χ^2 variable [Scheffé (1959, pp. 134–135)].

Application of doubly noncentral F-distributions to two-way cross-classification analysis of variance is illustrated in Tiku (1972), among numerous other sources. Engineering applications are described in Wishner (1962) and Price (1964).

7.2 Noncentral Beta Distributions

If $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ are mutually independent, then it is known (Chapter 27) that $\chi^2_{\nu_1}(\chi^2_{\nu_1} + \chi^2_{\nu_2})^{-1}$ has a standard beta distribution with parameters $\frac{1}{2}\nu_1, \frac{1}{2}\nu_2$. If $\chi^2_{\nu_1}$ is replaced by the noncentral $\chi'^2_{\nu_1}(\lambda_1)$, the resultant distribution is called a noncentral beta distribution with (shape) parameters $\frac{1}{2}\nu_1, \frac{1}{2}\nu_2$ and noncentrality parameter A,. If both χ^2's are replaced by noncentral χ^2's, giving

$$\beta''_{\nu_1,\nu_2}(\lambda_1,\lambda_2) = \chi'^2_{\nu_1}(\lambda_1)\{\chi'_{\nu_1}(\lambda_1) + \chi'_{\nu_2}(\lambda_2)\}^{-1}$$

$$= G''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)\{1 + G''_{\nu_1,\nu_2}(\lambda_1,\lambda_2)\}^{-1}, \qquad (30.55)$$

the corresponding distribution is a doubly noncentral beta distribution with (shape) parameters $\frac{1}{2}\nu_1, \frac{1}{2}\nu_2$ and noncentrality parameters λ_1, A,. Noncentral beta distributions can be represented as mixtures of central beta distributions in the same way as noncentral F's can be represented as mixtures of central F's.

Each noncentral χ^2 may be approximated by Patnaik's approximation (Chapter 29). This leads to approximating the distribution by that of

$$\frac{(\nu_1 + 2\lambda_1)(\nu_2 + \lambda_2)}{(\nu_1 + \lambda_1)(\nu_2 + 2\lambda_2)} \times \text{(Beta variable with parameters } f, .f_2) \qquad (30.56)$$

where f, $= (\nu_j + \lambda_j)^2(\nu_j + 2\lambda_j)^{-1}$ $(j = 1,2)$.

DasGupta (1968) compared this approximation with (1) an expansion using Jacobi's polynomials (see Chapter 1, Sections A6, A11) with initial beta distribution having correct first and second moments, and (2) Laguerre series expansions for each of the noncentral χ^2 distributions. He found that the Patnaik approximation was in general sufficiently accurate for practical purposes. Although approximations (1) and (2) are rather more accurate, they are more troublesome to compute.

It will be recalled that the distribution of $\chi'^2_{\nu}(\lambda)$ is related to that of the difference between two independent Poisson variables (see Chapter 29). By a similar argument [Johnson (1959)] it can be shown that if ν_1 is even,

$$\Pr[F'_{\nu_1,\nu_2}(\lambda_1) < f] = \Pr[Y - Z \geq \tfrac{1}{2}\nu_1], \qquad (30.57)$$

where Y and Z are mutually independent, Y has a negative binomial distribution (Chapter 5, Section 1) with parameters $\frac{1}{2}\nu_2, \nu_1 f/\nu_2$, and Z has a Poisson distribution with parameter $\frac{1}{2}\lambda_1$. By a direct extension of this argument, we obtain the relation for the doubly noncentral F-distribution,

$$\Pr[F''_{\nu_1,\nu_2}(\lambda_1, \text{A,}) < f] = \sum_{j=0}^{m} e^{-\lambda_2/2}\frac{(\frac{1}{2}\lambda_2)^j}{j!}\Pr\left[Y_j - Z \geq \frac{1}{2}\nu_1\right], \qquad (30.58)$$

where Y_j and Z are independent, Y_j has a negative binomial distribution with parameters $\frac{1}{2}\nu_2 + $ j, $\nu_1 f/\nu_2$, and Z is distributed as in (30.57).

Gupta and Onukogu (1983) derived an expression for the density of a product of two independent non-central beta variables with shape parameters $(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)$ and $(\frac{1}{2}\delta_1, \frac{1}{2}\delta_2)$, and noncentrality parameters A, and A,, respectively. Their representation is in terms of a Poisson weighted sum of mixtures of the corresponding central beta distributions.

REFERENCES

Aty, A. S. H. (1954). Approximate formulae for the percentage points and the probability integral of the noncentral χ^2 distribution, ***Biometrika, 44,*** 538–540.

Bargmann, R. E., and Ghosh, S. P. (1964). ***Noncentral statistical distribution programs for a computer language,*** Report No. RC-1231, IBM Watson Research Center, Yorktown Heights, NY.

Barton, D. E., David, F. N., and O'Neill, A. F. (1960). Some properties of the distribution of the logarithm of noncentral F, ***Biometrika,*** **47,** 417–429.

Bennett, B. M. (1955). Note on the moments of the logarithmic noncentral χ^2 and z distributions, ***Annals of the Institute of Statistical Mathematics,*** **7,** 57–61.

Bock, M. E., Judge, G. G., and Yancey, T. A. (1984). A simple form for the inverse moments of noncentral χ^2 and F random variables and certain confluent hypergeometric functions, ***Journal of Econometrics,*** **25,** 217–234.

Bulgren, W. G. (1971). On representations of the doubly non-central F distribution, ***Journal of the American Statistical Association, 66,*** 184–186.

Chow, M. S. (1987). A complete class theorem for estimating a noncentrality parameter, ***Annals of Statistics,*** **15,** 800–804.

Cohen, J. (1977). ***Statistical Power Analysis for the Behavioral Sciences,*** rev. ed., New York: Academic Press.

Cohen, J., and Nel, J. C. M. (1987). A comparison of two noncentral F approximations with applications to power analysis in set correlation, ***Multicariate Behavioral Research,*** 22, 483–490.

DasGupta, P. (1968). Two approximations for the distribution of double noncentral beta, ***Sankhyā, Series B,*** **30,** 83–88.

DeWaal, D. J. (1974), Bayes estimate of the noncentrality parameter in multivariate analysis, ***Communications in Statistics,*** 3, 73–79.

Dixon, W. J. (1962). Rejection of observations, In ***Contributions to Order Statistics,*** (eds., A. E. Sarhan and B. G. Greenberg), New York: Wiley, pp. 299–342.

Efron, B. (1970). Comments on Blyth's paper, ***Annals of Mathematical Statistics,*** **41,** 1049–1054.

Efron, B. (1973). Discussion on the paper by David, Stone and Zidek, ***Journal of the Royal Statistical Society, Series B,*** **35,** 219.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient, ***Proceedings of the Royal Society of London, Series A,*** **121,** 654–673.

Fleishman, A. I. (1980). Confidence intervals for correlation ratios, ***Educational and Psychological Measurement,*** **40,** 659–670.

Fowler, R. L. (1984). Approximating probability levels for testing null hypotheses with noncentral F distributions, ***Educational and Psychological Measurement, 44,*** 275–281.

Fox, M. (1956). Charts of the power of the F-test, ***Annals of Mathematical Statistics,*** **27,** 484–497.

Frick, H. (1990). AS R-84. A remark on Algorithm AS 226, Computing noncentral beta probabilities, ***Applied Statistics,*** **39,** 311–312.

Gelfand, A. E. (1983). Estimation in noncentral distributions, ***Communications in Statistics—Theory and Methods, 12,*** 463–475.

Guenther, W. C. (1978). Evaluation of probabilities for noncentral distributions and the difference of two T-variables with a desk calculator, ***Journal of Statistical Computation and Simulation,*** **6,** 199–206.

Guirguis, G. H. (1990). A note on computing the noncentrality parameter of the noncentral F distribution, ***Communications in Statistics—Simulation and Computation, 19,*** 1497–1511.

Gupta, D., and Onukogu, I. B. (1983). The distribution of the product of non-central beta variates, ***Biometrical Journal,*** **25,** 621–624.

Helstrom, C. W., and Ritcey, J. A. (1985). Evaluation of the noncentral F distribution by numerical contour integration, ***SIAM Journal on Scientific and Statistical Computing,*** 505–515.

Hodges, J. L. (1955). On the noncentral beta-distribution, ***Annals of Mathematical Statistics,*** **26,** 648–653.

Institute of Mathematical and Statistical Libraries. (1987). Packages, Version 1.0, vol. 3, Houston, Texas.

Johnson, N. L. (1959). On an extension of the connexion between the Poisson and χ^2 distributions, ***Biometrika,*** **46,** 352–363.

Lachenbruch, P. A. (1966). The noncentral F distribution-extensions of Tang's tables, ***Annals of Mathematical Statistics,*** **37,** 774. (Abstract). [Tables in University of North Carolina Mimeo Series No. 531 (1967).]

Lam, Y.-M. (1987). Confidence limits for non-centrality parameters of noncentral chi-squared and F distributions, ***ASA Proceedings of the Statistical Computing Section,*** 441–443.

Laubscher, N. H. (1960). Normalizing the noncentral t and F distributions, ***Annals of Mathematical Statistics, 31,*** 1105–1112.

Lee, C. M.-S. (1992). On the computation of central and noncentral Beta probabilities, ***Journal of Statistical Computation and Simulation,*** **43,** 1–10.

Lehmann, E. L. (1959). ***Testing Statistical Hypotheses,*** New York: Wiley.

Lehmer, E. (1944). Inverse tables of probabilities of errors of the second kind, ***Annals of Mathematical Statistics,*** **15,** 388–398.

Lenth, R. V. (1987). Computing noncentral beta probabilities, ***Applied Statistics,*** **36,** 241–244.

Madansky, A., and Olkin, I. (1969). Approximate confidence regions for constraint parameter. In ***Multivariate Analysis-II*** (ed., P. R. Krishnaiah), New York: Academic Press.

Madow, W. G. (1948). On a source of downward bias in the analysis of variance and covariance, ***Annals of Mathematical Statistics, 19,*** 351–359.

Majumder, K. L., and Bhattacharjee, G. P. (1973). Algorithm AS63. The incomplete beta integral, ***Applied Statistics,*** **22,** 409–411.

Malik, H. J. (1970). An alternative derivation of doubly noncentral F distribution, *Metron*, **28**, 180–184.

Marakathavalli, N. (1955). Unbiased test for a specified value of the parameter in the noncentral F distribution, *Sankhyā*, 15, 321–330.

Mudholkar, G. S., Chaubey, Y. P., and Lin, C. C. (1976). Some approximations for the noncentral F distribution, *Technometrics*, **18**, 351–358.

Narula, S. C., and Levy, K. J. (1975). Probability density plots of the noncentral χ^2 and noncentral F distributions, *International Statistical Review*, **43**, 79–82.

Narula, S. C., and Weistroffer, H. R. (1986). Computation of probability and noncentrality parameter of noncentral F distribution, *Communications in Statistics—Simulation and Computation*, **15**, 871–878.

Nicholson, W. L. (1954). A computing formula for the power of the analysis of variance test, *Annals of Mathematical Statistics*, **25**, 607–610.

Norton, V. (1983). A simple algorithm for computing the noncentral F distribution, *Applted Statisttcs*, *32*, 84–85.

Odeh, R. E., and Fox, M. (1975). *Sample Size Choice; Charts for Experiments with Linear Models*, New York: Dekker.

Pandey, J. N., and Rahman, M. (1971). The maximum likelihood estimate of the noncentrality parameter of a noncentral F variate, *SIAM Journal on Mathematical Analysis*, 2, 269–276.

Park, J. H. (1964). Variations of the noncentral t and beta distributions, *Annals of Mathematical Statistics*, **35**, 1583–1593.

Patnaik, P. B. (1949). The non-central χ^2 and F-distributions and their applications, *Biometrtka*, **36**, 202–232.

Pe, T., and Drygas, H. (1994). Alternative representations of some doubly noncentral distributions: Univariate case, *Statistical Papers*, **35** *(to appear)*.

Pearson, E. S. (1960). Editorial note, *Biometrika*, **47**, 430–431.

Pearson, E. S., and Hartley, H. O. (1951). Charts of the power function for analysis of variance tests, derived from the noncentral F distribution, *Biometrika*, **38**, 112–130.

Pearson, E. S., and Tiku, M. L. (1970). Some notes on the relationship between the distributions of central and noncentral F, *Biometrika*, **57**, 175–179.

Perlman, M. D., and Rasmussen, U. A. (1975). Some remarks on estimating a noncentrality parameter, *Communications in Statisttcs*, *4*, 455–468.

Posten, H. O. (1993). An effective algorithm for thc noncentral beta distribution function, *The American Statistician*, **47**, 129–131.

Price, R. (1964). Some noncentral F distributions expressed in closed form, *Biometrika*, **51**, 107–122.

Rasmussen, U. (1973). Testing and estimation problems concerning noncentrality parameters, *Ph.D. dissertation*, University of Minnesota, Minneapolis.

Rukhin, A. L. (1993). Estimation of the noncentrality parameter of an F distribution, *Journal of Stattstical Planning and Inference*, **35**, 201–211.

Schader, M. and Schmid, F. (1986). Distribution function and percentage points for the central and noncentral F distribution, *Staristtsche Hefte*, 27, 67–74.

Scheffé, H. (1959). *The Analysis of Vartance*, New York: Wiley.

Seber, G. A. F. (1963). The noncentral chi-squared and beta distributions, ***Biometrika,*** 50, 542–544.

Severo, N., and Zelen, M. (1960). Normal approximation to the chi-square and noncentral F probability functions, ***Biometrika,* 47,** 411–416.

Sibuya, M. (1967). On the noncentral beta distribution function, ***Unpublished manuscript.***

Singh, K. P., and Relyea, G. E. (1992). Computation of noncentral F probabilities. A computer program, ***Computational Statistics & Data Analysis,* 13,** 95–102.

Steffens, F. E. (1968). ***Probability integrals of doubly noncentral*** F- ***and t-distributions with regression applications,*** Research Report No. 267, Council for Scientific and Industrial Research, Pretoria, South Africa.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use, ***Statistical Research Memoirs,*** 2, 126–150.

Tiku, M. L. (1965a). Laguerre series forms of noncentral χ^2 and F distributions, ***Biometrika,*** 52, 414–427.

Tiku, M. L. (1965b). Series expansions for the doubly noncentral F distribution, ***Australian Journal of Statistics,* 7,** 78–89.

Tiku, M. L. (1966). A note on approximating to the noncentral F distribution, ***Biometrika,*** 53, 606–610.

Tiku, M. L. (1967). Tables of the power of the F test, ***Journal of the American Statistical Association, 62,*** 525–539.

Tiku, M. L. (1972). A note on the distribution of the doubly noncentral F distribution, ***Australian Journal of Statistics,* 14,** 37–40.

Tiku, M. L. (1974). Doubly Noncentral F Distribution Tables and Applications, ***Selected Tables in Mathematical Statistics,*** vol. 2 (eds., H. L. Harter and D. B. Owen), Providence, RI: American Mathematical Society, pp. 139–178.

Ura, S. (1954). A table of the power function of the analysis of variance test, ***Reports of Statistical Application Research,*** JUSE, **3,** 23–28.

Venables, W. (1975). Calculation of confidence intervals for non-centrality parameters, ***Journal of the Royal Statistical Society, Series B,* 37,** 406–412.

Wang, M. C. (1992). Self-validating computation of non-central F probabilities, ***Proceedings of the ASA Statistical Computing Section,*** pp. 51–55.

Weibull, C. (1953). The distribution of *t*- and F-statistics and of correlation and regression coefficients in stratified samples from normal populations with different means, ***Skandinavisk Aktuarietidskrift,*** 1–2, Supplement.

Winer, B. J. (1971). ***Statistical Principles in Experimental Design,*** New York: McGraw-Hill.

Wishart, J. (1932). A note on the distribution of the correlation ratio, ***Biometrika,* 24,** 441–456.

Wishner, R. P. (1962). Distribution of the normalized periodogram detector, IRE ***Transactions on Information Theory,*** IT-8, 342–349.

CHAPTER 31

Noncentral t-Distributions

1 DEFINITION

The distribution of the ratio

$$t'_\nu(\delta) = \frac{U + \delta}{\chi_\nu \nu^{-1/2}}, \tag{31.1}$$

where U and χ_ν are independent random variables distributed as standard normal $[N(0, 1)]$ and chi with ν degrees of freedom, respectively, and δ is a constant, is called the noncentral t-distribution with ν degrees of freedom and noncentrality parameter δ. Sometimes δ^2 (or even $\frac{1}{2}\delta^2$), rather than δ, is termed the noncentrality parameter. If δ is equal to zero, the distribution is that of (central) t with ν degrees of freedom, as was discussed in Chapter 28.

When there is no fear of confusion, δ may be omitted and t'_ν used instead of $t'_\nu(\delta)$. Occasionally even the ν may be omitted, and t' used. However, whenever there is possibility of confusion—such as when two or more values of the noncentrality parameter are under discussion—the full symbol $t'_\nu(\delta)$ should be used.

2 HISTORICAL REMARKS

The noncentral t-distribution was derived (though not under that name) by Fisher (1931), who showed how tables of repeated partial integrals of the standard normal distribution could be used in connection with this distribution. Tables given by Neyman, Iwaszkiewicz, and Kolodziejczyk (1935) and Neyman and Tokarska (1936) were based on evaluation of probability integrals of certain noncentral t-distributions.

Tables from which percentage points of noncentral t-distributions could be obtained were given by Johnson and Welch (1940). Later tables [Resnikoff and Lieberman (1957); Locks, Alexander, and Byars (1963); Bagui (1993)] are fuller and require less calculation. Charts based on the probability integral are given by **Pearson** and **Hartley (1954)** (see also Section 7).

Formal expressions for the distribution function are rather more complicated than those for noncentral χ^2 and F. A number of different formulas can be found in Amos **(1964)**. A number of approximations for probability integrals, and for percentage points, of noncentral t-distributions have been proposed. Comparisons of various approximations for percentage points have been made by van Eeden **(1961)** and by Owen **(1963)**. See Section 6 for more details. Computer programs for calculating percentage points have been described by Owen and Amos **(1963)** and by Bargmann and Ghosh (1964). Amos **(1964)** reported on comparisons of two such computer programs. See Section 7 for more details.

3 APPLICATIONS AND ESTIMATION

The statistic $\sqrt{n}(\bar{X} - \xi_0)/S$ is used in testing the hypothesis that the mean of a normal population is equal to ξ_0. If $\bar{X}$ $(= n^{-1}\sum_{i=1}^n X_i)$ and

$$S\left(= \sqrt{(n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}\right)$$

are calculated from a random sample of size n, and the population mean is equal to ξ_0, then $\sqrt{n}(\bar{X} - \xi_0)/S$ should be distributed as (central) t with n − 1 degrees of freedom. If, however, the population mean ξ is not equal to ξ_0, then $\sqrt{n}(\bar{X} - \xi_0)/S$ is distributed as $t'_{n-1}(\sqrt{n}(\xi - \xi_0)/\sigma)$, where σ is the population standard deviation. The power of the test is calculated as a partial integral of the probability density function of this noncentral t-distribution.

Similarly a statistic used in testing equality of means of two normal populations, (Π_1) and (Π_2) (with common, though unknown, variance σ^2), using random samples of sizes n_1, n_2, respectively, is

$$\frac{\sqrt{n_1 n_2 (n_1 + n_2)^{-1}}\,(\bar{X}_1 - \bar{X}_2)}{\sqrt{(n_1 + n_2 - 2)^{-1}[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]}}. \tag{31.2}$$

If the two population means are indeed equal, this statistic should be distributed as (central) t with $(n_1 + n_2 - 2)$ degrees of freedom. But, if (mean of population Π_1) − (mean of population Π_2) $= \theta$, then the statistic is distributed as $t'_{n_1+n_2-2}\left(\theta\sigma^{-1}\sqrt{n_1 n_2 (n_1 + n_2)^{-1}}\right)$. Here, again, the power of the test can be calculated as a partial integral of the appropriate noncentral t-distribution.

Charts giving powers of t-tests have been published in Pearson and Hartley (1954) and Croarkin (1962). Tables of the power function are also available in Neyman, Iwaszkiewicz, and Kolodziejczyk (1935) and Davies (1954). Noncentral t-distributions also occur as distributions of certain test statistics in multivariate analysis [see, e.g., Gupta and Kabe (1992)].

It is sometimes desired to calculate confidence intervals for the ratio of population mean to standard deviation (reciprocal of the coefficient of variation). Such intervals may be computed in suitable cases by noting that if $\bar{X}$ and S^2 be calculated from a random sample of values $X_1, X_2, \cdots, X_n$ from a normal population with expected value ξ and standard deviation σ, then $\sqrt{n}\,\bar{X}/S$ is distributed as $t'_{n-1}(\sqrt{n}\,\xi/\sigma)$. Symmetrical $100(1-\alpha)\%$ confidence limits for ξ/σ are obtained as solutions for θ of the equations

$$t'_{n-1,\frac{\alpha}{2}}\left(\sqrt{n}\,\theta\right) = \frac{\sqrt{n}\,\bar{X}}{S} \tag{31.3a}$$

and

$$t'_{n-1,1-\frac{\alpha}{2}}\left(\sqrt{n}\,\theta\right) = \frac{\sqrt{n}\,\bar{X}}{S}. \tag{31.3b}$$

[Approximations to the distribution of the sample coefficient of variation in relation to the noncentral t-distribution were studied by McKay (1932) and an accurate (to four decimal places) approximation to the percentage points in terms of the χ^2 percentage points was derived by Iglewicz, Myers, and Howe (1968).]

Belobragina and Eliseyev (1967) have constructed a nomogram that indirectly gives the lower $100(1-\alpha)\%$ bounds for ξ/σ. Actually their charts show upper bounds for $\Phi(-\xi/\sigma)$, given $\bar{X}/S$. Sample sizes included are 5, 10, 20, 50, and 100; and confidence coefficients 90%, 95%, 97%, 99%, and 99.9%.

Rukhin (1992) considered point estimation of $(\xi/\sigma)^2$. The maximum likelihood estimator is

$$\frac{n}{n-1}\left(\frac{\bar{X}}{S}\right)^2,$$

and the minimum variance unbiased estimator (for $n > 3$) is

$$\frac{n-3}{n-1}\left(\frac{\bar{X}}{S}\right)^2 - \frac{1}{n}. \tag{31.4}$$

Rukhin pointed out that this estimator can be negative and so is inadmissible, and he proceeded to search for good estimators of form

$$\frac{a}{n-1}\left(\frac{\overline{X}}{S}\right)^2 + \frac{b}{n}, \qquad a, b \geq 0. \tag{31.5}$$

Using mean square error as a measure of inaccuracy, Rukhin has recommended estimators of form

$$\frac{n-5}{n-1}\left(\frac{\overline{X}}{S}\right)^2 + \frac{b}{n}, \qquad n > 5, b > 0, \tag{31.6}$$

for cases where $|\xi/\sigma|$ is expected to be large. These estimators are admissible within the class of estimators (31.5), but not generally.

We now consider the problem of constructing a 100(1 − a)% confidence interval for the 100 $P\%$ quantile of the distribution of X, which is $5 + u_P\sigma$ [with ξ and a defined as above, and $\Phi(u_P) = P$]. This problem was considered by Stedinger (1983a, b) and Chowdhury and Stedinger (1991) in connection with estimating P^{-1} year flood flows in hydraulic engineering (see also below). Since $\sqrt{n}(\overline{X} - \xi - u_P\sigma)/S$ is distributed as $t'_{n-1}(-u_P\sqrt{n})$, we have

$$\begin{aligned}\Pr\Bigg[\overline{X} - t'_{n-1,1-\frac{\alpha}{2}}\left(-u_P\sqrt{n}\right)\frac{S}{\sqrt{n}} &< \xi + u_P\sigma \\ &< \overline{X} - t'_{n-1,\frac{\alpha}{2}}\left(-u_P\sqrt{n}\right)\Big|\xi,\sigma\Bigg] \\ = 1 - \alpha.\end{aligned}$$

Hence

$$\left[\overline{X} - \left\{\frac{t'_{n-1,1-\frac{\alpha}{2}}\left(-u_P\sqrt{n}\right)}{\sqrt{n}}\right\}S,\ \overline{X} - \left\{\frac{t'_{n-1,\frac{\alpha}{2}}\left(-u_P\sqrt{n}\right)}{\sqrt{n}}\right\}S\right] \tag{31.7}$$

is a 100(1 − a)% confidence interval for $\xi + u_P\sigma$. An equivalent formula [since $t'_{\nu,\epsilon}(-\delta) = -t'_{\nu,1-\epsilon}(\delta)$] is [Wolfowitz (1946)]

$$\left[\overline{X} + \left\{\frac{t'_{n-1,\frac{\alpha}{2}}\left(u_P\sqrt{n}\right)}{\sqrt{n}}\right\}S,\ \overline{X} + \left\{\frac{t'_{n-1,1-\frac{\alpha}{2}}\left(u_P\sqrt{n}\right)}{\sqrt{n}}\right\}S\right]. \tag{31.7$'$}$$

Tables (to three decimal places) of the multipliers of S are presented in Stedinger (1983a) for

$$1-\alpha = 0.50, 0.90, 0.99; \quad P = 0.90, 0.98, 0.99;$$

$$n = 4(1)20, 22, 25, 27, 30(5)60(10)100.$$

[See also Chowdhury and Stedinger (1991).] Comparisons with some approximate formulas are also included in this paper, which is particularly concerned with applications in distributions of flood flows. This is one among many of the fields of application developed in recent years, broadening the already extensive bibliography (of over 100 references) in Owen (1968).

Kühlmeyer's (1970) monograph also contains many examples. Here we note a few later examples. Hall and Sampson (1973) utilized (31.7) in constructing approximate tolerance limits for the distribution of the product of two independent normal variables in connection with pharmaceutical tablet manufacture. Malcolm (1984) used noncentral t-distributions in setting microbiological specifications for foods. Lahiri and Teigland (1987) found that noncentral t-distributions gave good fits to distributions of forecasters' estimates of Gross National Product and Inflation Price Deflator. Miller (1989) used noncentral t-distribution in calculating parametric empirical Bayes factors for calculating normal tolerance bounds. Dasgupta and Lahiri (1992) utilized noncentral t-distribution in one of several models they consider for interpretation of survey data.

Phillips (1993) applies noncentral t-distributions in constructing tests for hypotheses that the probability that the ratio between 'bioavailability' for a new and standard drug formulations falls between specified limits.

4 MOMENTS

The rth moment of $t'_\nu(\delta)$ about zero is

$$\begin{aligned} E[t'^r_\nu(\delta)] &= \nu^{r/2} E[\chi_\nu^{-r}] E[(U+\delta)^r] \\ &= \left(\frac{1}{2}\nu\right)^{r/2} \frac{\Gamma(\frac{1}{2}(\nu-r))}{\Gamma(\frac{1}{2}\nu)} \sum_{0\le j\le r/2} \binom{r}{2j} \frac{(2j)!}{2^j j!} \delta^{r-2j}. \end{aligned} \tag{31.8}$$

Hogben, Pinkham, and Wilk (1961) gave an alternative form, in which the sum is replaced by the expression $e^{-\delta^2/2}(d/d\delta)^r(e^{\delta^2/2})$. Merrington and

Pearson (1958) gave the following expressions:

$$\mu_1' = \left(\frac{1}{2}\nu\right)^{1/2} \frac{\Gamma(\frac{1}{2}(\nu-1))}{\Gamma(\frac{1}{2}\nu)}\delta, \tag{31.9a}$$

$$\text{var}(t_\nu') = \mu_2 = \frac{\nu}{\nu-2}(1+\delta^2) - \mu_1'^2, \tag{31.9b}$$

$$\mu_3 = \mu_1'\left\{\frac{\nu(2\nu-3+\delta^2)}{(\nu-2)(\nu-3)} - 2\mu_2\right\}, \tag{31.9c}$$

$$\mu_4 = \frac{\nu^2}{(\nu-2)(\nu-4)}(3+6\delta^2+\delta^4)$$

$$-\mu_1'^2\left\{\frac{\nu[(\nu+1)\delta^2+3(3\nu-5)]}{(\nu-2)(\nu-3)} - 3\mu_2\right\}. \tag{31.9d}$$

The rth central moment can be expressed as a polynomial in δ^2:

$$\mu_r = \sum_{j=0}^{[r/2]} c_{r,r-2j}(\nu)\delta^{r-2j}.$$

Hogben, Pinkham, and Wilk (1961) provided tables of the coefficients $c_{r,i}(\nu)$ to six significant figures for r = 2, 3, 4, and

$$\nu = 1(1)25(5)80(10)100(50)200(100)1000.$$

They also give values of μ_1'/δ [see (31.9a)] to six significant figures.

For ν large (with δ fixed)

$$\mu_1' \doteqdot \delta, \quad \text{var}(t_\nu') \doteqdot 1 + \tfrac{1}{2}\delta^2\nu^{-1}, \quad \mu_3 \doteqdot \nu^{-1}\delta\left[3 + \tfrac{5}{4}\delta^2\nu^{-1}\right],$$

and the skewness moment ratio $\sqrt{\beta_1}$ is approximately $\nu^{-1}\delta(3-\delta^2\nu^{-1})$. Note that the skewness has the same sign as 6; this is also true of the expected value. Further the distribution of $t_\nu'(-\delta)$ is a mirror image (reflected at $t_\nu' = 0$) of that of $t_\nu'(\delta)$.

The (β_1, β_2) values of the t_ν' distribution fall in the Type IV region of the Pearson system of frequency distributions (see Chapter 12). Merrington and Pearson (1958) found the interesting approximate relation:

$$\beta_2 \doteqdot \frac{1.406(\nu-3.2)}{\nu-4}\beta_1 + \frac{3(\nu-2)}{\nu-4}.$$

5 DISTRIBUTION FUNCTION

Since the events $t'_\nu(\delta) < 0$ and $U + 6 < 0$ are identical,

$$\Pr[t'_\nu(\delta) < 0] = \Pr[U < -\delta] = \Phi(-\delta). \qquad (31.10)$$

Also,

$$\Pr[t'_\nu(\delta) \le t] = 1 - \Pr[t'_\nu(-\delta) \ \mathrm{I} \ -t]. \qquad (31.10)'$$

The relationship

$$\Pr[t'_\nu \le t] = \Pr\left[U + \delta \le \frac{t\chi_\nu}{\sqrt{\nu}}\right]$$

leads to

$$\Pr[t'_\nu \le t] = \frac{1}{2^{(\nu/2)-1}\Gamma(\frac{1}{2}\nu)} \int_0^\infty x^{\nu-1} e^{-x^2/2} \times \frac{1}{\sqrt{2\pi}} \int_0^{tx/\sqrt{\nu}} \exp\left[-\frac{1}{2}(u-\delta)^2\right] du\, dx. \qquad (31.11)$$

The right-hand side of (31.11) can be written

$$\frac{1}{2^{(\nu/2)-1}\Gamma(\frac{1}{2}\nu)} \int_0^\infty \left\{ x^{\nu-1} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(tx/\sqrt{\nu})-\delta} e^{-u^2/2} \right\} du\, dx. \qquad (31.11')$$

For computational purposes the following formulas, due to Kiihlmeyer (1970), may be useful:

For ν odd,

$$\Pr[t'_\nu(\delta) \ \mathrm{I} \ t] = \Phi(-\delta\sqrt{B}) + 27\left(-\delta\sqrt{B}, \frac{t}{\sqrt{\nu}}\right) + 2(M_1 + M_3 + \cdots + M_{\nu-2})$$

For ν even,

$$\Pr[t'_\nu(\delta) \le t] = \Phi(-\delta) + 2\pi(M_0 + M_2 + \cdots + M_{\nu-2}),$$

where

$$B = \nu(\nu + t^2)^{-1},$$

$$\tau(h, a) = (\sqrt{2\pi})^{-1} \int_0^a \exp\{-\tfrac{1}{2}h^2(1 + x^2)\}\, dx,$$

$$M_k = (1 - k^{-1})B\left\{a_k \delta\left(\frac{t}{\sqrt{\nu}}\right) M_{k-1} + M_{k-2}\right\},$$

with

$$a_k = \{(k-2)a_{k-1}\}^{-1} \text{ for } k \ge 3,\ a_2 = 1,$$

$$M_{-1} = 0,$$

$$M_0 = \left(\frac{t}{\sqrt{\nu}}\right)\sqrt{B}\,(\delta\sqrt{B})\varphi(\delta\sqrt{B})\Phi\left(\delta t\sqrt{\frac{B}{\nu}}\right),$$

$$M_1 = B\left\{\delta M_0 + \frac{1}{\sqrt{2\pi}}\frac{t}{\sqrt{\nu}}\varphi(\delta)\right\},$$

$$M_2 = \frac{1}{2}B\left\{\delta\frac{t}{\sqrt{\nu}}M_1 + M_0\right\},$$

$$\varphi(x) = \left(\sqrt{2\pi}\right)^{-1}e^{-x^2/2} = \Phi'(x).$$

Differentiating *(31.11)* with respect to t gives the probability density function

$$p_{t'_\nu}(t) = \frac{e^{\delta^2/2}}{2^{(\nu-1)/2}\sqrt{\pi\nu}\,\Gamma(\frac{1}{2}\nu)}\int_0^\infty x^\nu \exp\left[-\tfrac{1}{2}\left\{(1+t^2\nu^{-1})x^2 - 2(t\nu^{1/2})x\right\}\right]dx$$

$$= \frac{\nu!}{2^{(\nu-1)/2}\sqrt{\pi\nu}\,\Gamma(\frac{1}{2}\nu)}\left\{\exp\left[-\frac{\nu\delta^2}{\nu+t^2}\right]\right\}$$

$$\times\left(\frac{\nu}{\nu+t^2}\right)^{(\nu-1)/2} Hh_\nu\left[-\frac{\delta t}{\sqrt{\nu+t^2}}\right], \qquad (31.12)$$

where $Hh_\nu(x) = (\nu!)^{-1}\int_0^\infty u^\nu \exp[-\frac{1}{2}(u+x)^2]\,du$. See also *Eq.* (31.19) below. This form was given by Fisher (1931), in the introduction to Airey's (1931) tables of the *Hh* functions. Note that $(\sqrt{2\pi})^{-1}Hh_\nu(x)$ is the *vth* repeated partial integral of the standard normal probability density function

$$\left(\sqrt{2\pi}\right)^{-1}Hh_\nu(x) = \left(\sqrt{2\pi}\right)^{-1}\int_x^\infty\int_{y_\nu}^\infty \cdots \int_{y_3}^\infty\int_{y_2}^\infty e^{-y_1^2/2}\,dy_1\,dy_2\,\cdots\,dy_\nu. \qquad (31.13)$$

Voit and Rust *(1990)* point out that the Hh_ν function can be expressed in terms of the Whittaker *U* function as

$$Hh_\nu(-\theta) = U(\nu + \tfrac{1}{2}, -\theta)e^{-\theta^2/4}, \qquad (31.14)$$

where

$$U(\nu + \tfrac{1}{2}, -\theta) = \frac{\sqrt{\pi}}{2^{\nu/2}\Gamma(\frac{1}{2}\nu + 1)}(2y_1 + y_2)$$

and y_1 and y_2 are the odd and even solutions of the differential equation

$$\frac{d^2y}{d\theta^2} - \left(\frac{1}{4}\theta^2 + \nu + \frac{1}{2}\right)y = 0.$$

They exploit this observation to construct another, S-system, representation of noncentral t-distributions. (See the remarks on page 519.)

There are still other formulas for the distribution of $t'_\nu(\delta)$. First, we have

$$p_{t'_\nu}(t) = \frac{e^{-\delta^2/2}\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{1}{2}\nu\right)}\left(\frac{\nu}{\nu+t^2}\right)^{(\nu+1)/2} \times \sum_{j=0}^{\infty}\frac{\Gamma\left(\frac{1}{2}(\nu+j+1)\right)}{j!\,\Gamma\left(\frac{1}{2}(\nu+1)\right)}\left[\frac{t\delta\sqrt{2}}{\sqrt{\nu+t^2}}\right]^j. \tag{31.15}$$

This is of form $p_{t'_\nu}(t) = \sum_{j=0}^{\infty} c_j[\theta(t)]^j$, where the c_j's are constants and $\theta(t) = t\delta\sqrt{2}(\nu + t^2)^{-1/2}$. The expression $[\theta(t)]^0$ is to be interpreted as 1 for all values (including 0) of $\theta(t)$. If $\delta = 0$, the expression for $p_{t'_\nu}(t)$ reduces to that of a (central) t_ν density. Note that if 6 and t are of opposite signs, the series alternates in sign. The series in (31.15) can be evaluated term by term to give values for probabilities in terms of incomplete beta function ratios. If the range of integration be taken from 0 to t (> 0), each term is positive, and

$$\Pr[0 < t'_\nu \le t] = \frac{1}{2}e^{-\delta^2/2}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\delta^2\right)^{j/2}}{\Gamma\left(\frac{1}{2}j+1\right)}I_{t^2/(\nu+t^2)}\left(\frac{1}{2}(j+1), \frac{1}{2}\nu\right). \tag{31.16}$$

[See also Hawkins (1975).] Guenther (1975) provided an expansion in terms of incomplete beta function ratios (see Chapter 1, Section A6)

$$\Pr[0 < t'_\nu < t] = \sum_{j=0}^{\infty}\left[p_j I_x\left(j+\frac{1}{2}, \frac{\nu}{2}\right) + q_j I_x\left(j+1, \frac{\nu}{2}\right)\right], \tag{31.17}$$

where

$$x = \frac{t^2}{t^2+\nu},$$

$$p_j = \frac{\frac{1}{2}e^{-\delta^2/2}(\delta^2/2)^j}{j!},$$

$$q_j = \frac{\frac{1}{2}\delta e^{-\delta^2/2}(\delta^2/2)^j}{\sqrt{2}\,\Gamma\left(j+\frac{3}{2}\right)}.$$

This formula is suitable for pocket calculators, and the formula was used by Lenth (1989) for his computer algorithm (see Section 7).

It is possible to obtain an expansion similar to (31.16) for $\Pr[-t < t'_\nu \le 0]$ with $-t < 0$, but the terms now alternate in sign. However, probabilities of the latter kind can be evaluated from (31.16) as well as the value of $\Pr[-t < t'_\nu \le t]$, which is equal to $\Pr[t'^2_\nu \le t^2]$. In this case we note that t'^2_ν

has the noncentral F-distribution $F'_{1,\nu}(\delta)$ (Chapter 30, Section 4). Hence

$$\Pr[-t \le t'_\nu \le t] = e^{-\delta^2/2} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\delta^2\right)^j}{j!} \Pr\left[F_{1+2j,\nu} < (1+2j)^{-1}t^2\right]$$

$$= e^{-\delta^2/2} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\delta^2\right)^j}{j!} I_{t^2/(\nu+t^2)}\left(j + \frac{1}{2}, \frac{1}{2}\nu\right). \quad (31.18)$$

If ν is even, and $t > 0$, $\Pr[0 < t'_\nu < t]$ can be expressed as a finite sum in terms of Hh functions:

$$\Pr[0 < t'_\nu < t] = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\delta^2\nu}{2(\nu+t^2)}\right] \times \sum_{j=0}^{(\nu-2)/2} \frac{(2j)!}{2^j j!} \left[\frac{\nu}{2(\nu+t^2)}\right]^j Hh_{2j}\left[-\frac{\delta t}{\sqrt{\nu+t^2}}\right]. \quad (31.19)$$

Among the formulas given by Amos (1964), one to which he gave special attention, in regard to its computing use, expresses the cumulative distribution function of t'_ν in terms of confluent hypergeometric functions (see Chapter 1, Section A7):

$$M(a,b;x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b)}{\Gamma(a)\Gamma(b+j)} \frac{x^j}{j!}$$

(b cannot be zero, or a negative integer). The formula (for $\nu > 2$) is

$$\begin{aligned}\Pr[t'_\nu \le t] = 1 - \left(\sqrt{2\pi}\right)^{-1} \int_{-\infty}^{\delta\nu^{1/2}/(\nu+t^2)^{1/2}} \exp\left(-\frac{1}{2}u^2\right) du \\ + \left(\sqrt{\pi}\right)^{-1}\left[\frac{t}{\sqrt{\nu+t^2}} \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\nu\right)} \right. \\ \times \sum_{j=0}^{\infty} \frac{\Gamma\left(1-\frac{1}{2}\nu+j\right)}{j!(2j+1)\Gamma\left(1-\frac{1}{2}\nu\right)} \left(\frac{t^2}{\nu+t^2}\right)^j \\ \times M\left[j+\frac{1}{2}, \frac{1}{2}; \frac{-\delta^2\nu}{2(\nu+t^2)}\right] \\ - \frac{\delta\sqrt{\nu}}{\sqrt{2(\nu+t^2)}} \sum_{j=1}^{\infty} \frac{\Gamma\left(\frac{1}{2}(1-\nu)+j\right)}{j!\,\Gamma\left(\frac{1}{2}(1-\nu)\right)} \left(\frac{t^2}{\nu+t^2}\right)^j \\ \left. \times M\left(j+\frac{1}{2}, \frac{3}{2}; \frac{-\delta^2\nu}{2(\nu+t^2)}\right)\right].\end{aligned} \quad (31.20)$$

If ν is even, the first of the two summations contains only a finite number $(\frac{1}{2}\nu + 1)$ of terms; if ν is odd, the second summation contains only a finite number $[\frac{1}{2}(\nu + 3)]$ of terms.

Hodges and Lehmann (1965) derived an asymptotic (with $\nu \to \infty$) series for the power of the t-test (see Section 3) with ν degrees of freedom in terms of the central moments of $\chi_\nu/\sqrt{\nu}$. They found that using this series for ν "not too small" (the case $\nu = 40$ was investigated in detail), determination of the power with sufficient precision is possible in many cases. Moreover the series is useful as an indicator of the proper interpolation procedures with respect to δ in the noncentral t tables (see Section 6).

If $t > 0$, Guenther (1975) recommended separating the infinite sum in equation (31.20) into two sums, one for j odd and one for j even. In summing the even terms, using the identity

$$j! = \frac{2^j\Gamma[(j+1)/2]\Gamma[(j+2)/2]}{\sqrt{\pi}}$$

and putting $t^2 = u$, followed by $u = (j+1)/f$, we have

$$\Pr[t'_\nu > t] = \frac{1}{2}\sum_{j=0}^{\infty}\frac{e^{-\delta^2/2}(\delta^2/2)^j}{j!}\int_{t^2/(2j+1)}^{\infty}p(f;2j+1,\nu)\,df$$
$$+\sum_{j=0}^{\infty}\frac{\delta e^{-\delta^2/2}}{\sqrt{2\pi}}\frac{j!\,(2\delta^2)^j}{(2j+1)!}\int_{t^2/(2j+2)}^{\infty}p(f;2j+2,\nu)\,df, \quad (31.21)$$

where $p(f;\nu_1,\nu_2)$ is the (central) F density with ν_1, ν_2 degrees of freedom [Chapter 27, Eq. (27.4)]. This expression is especially suitable for pocket calculator use.

Ifram (1970) noted that the pdf of a normal random variable X with expected value δ and variance 1 can be expressed as

$$\begin{aligned} p_X(x) &= \phi(x-\delta) \\ &= \phi(x)\exp\left(-\tfrac{1}{2}\delta^2 + \delta x\right) \\ &= \tfrac{1}{2}E_K\left[p_{\chi^2_{2K+1}}(|x|)\right] + \tfrac{1}{2}\delta x^{-1}E_K\left[p_{\chi^2_{2K+2}}(|x|)\right], \end{aligned} \quad (31.22)$$

where

$$p_{\chi^2_\nu}(x) = \left\{2^{\nu/2}\Gamma\left(\tfrac{1}{2}\nu\right)\right\}^{-1}x^{(\nu/2)-1}e^{-x/2} \qquad \text{[cf. Eq. (18.1), Chapter 18]}$$

and K has a Poisson distribution with expected value $\frac{1}{2}\delta^2$. From this he deduced that

$$p_{t'_\nu(\delta)}(t) = \frac{1}{2}E_K\left[p_{\sqrt{G_{2K+1,\nu}}}(|t|)\right] + \frac{1}{2}\frac{\delta}{t}\frac{\Gamma((\nu-1)/2)}{\sqrt{2}\,\Gamma\left(\frac{1}{2}\nu\right)}E_K\left[p_{\sqrt{G_{2K+3,\nu-1}}}(|t|)\right], \quad (31.23)$$

where G_{ν_1,ν_2} is distributed as the ratio $\chi^2_{\nu_1}/\chi^2_{\nu_2}$ of independent χ^2's. If U^2 has a G_{ν_1,ν_2} distribution, then U is said to be a $\sqrt{G_{\nu_1,\nu_2}}$ random variable. Voit and Rust (1990) suggest evaluating noncentral t-distributions using a canonical S-system form [see Chapter 12, Eq. (12.90)]. They note that the noncentral t density, given by (31.15) can be written

$$\frac{\exp(-\delta^2/2)}{\sqrt{\pi\nu}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\left[\frac{\nu}{\nu+t^2}\right]^{(\nu+1)/2}S(t), \qquad (31.15)'$$

where

$$S(t)=\sum_{j=0}^{\infty}\frac{\Gamma((\nu+j+1)/2)}{\Gamma((\nu+1)/2)}\left[\frac{t\delta\sqrt{2}}{\sqrt{\nu+t^2}}\right]^j,$$

and that

$$S(t) \quad \text{and} \quad Z(t)=\frac{(\nu+t^2)^{3/2}}{\nu\delta}S'(t)$$

satisfy the differential equation

$$\frac{dZ(t)}{dt}=\frac{t\nu\delta^2}{(\nu+t^2)^2}Z(t)+\frac{\nu(\nu+1)\delta}{(\nu+t^2)^{3/2}}S(t). \qquad (31.24)$$

This permits construction of an S-system representation [Chapter 12, Eq. (12.90)] for central and noncentral t-distributions. They also describe an S-system approach for computing quantiles and moments of noncentral t-distributions.

6 APPROXIMATIONS

This rather lengthy Section mirrors the extensive literature on approximations to non-central distributions which attracted a substantial number of researchers in the last fifty years or so. In spite of the advances in computer technology, this topic of research is far from extinct or exhaustion, as indicated by the ingenious work of Deutler (1984) to be described below.

If δ is kept at a fixed value, and ν is increased without limit, the t'_ν distribution tends to the normal distribution $N(\delta, 1)$. If ν (> 2) is kept fixed, and δ increased without limit, the standardized t'_ν-distribution tends to the standardized χ_ν^{-1} distribution.

Earlier approximations were based on an indirect approach. Jennett and Welch (1939) used the approximate normality of $(U - t'_0\chi_\nu\nu^{-1/2})$ in the equation

$$\Pr[t'_\nu \le t] = \Pr\left[U - t\chi_\nu \nu^{-1/2} \le -\delta\right] \tag{31.25}$$

to obtain

$$\Pr[t'_\nu \le t] = \left(\sqrt{2\pi}\right)^{-1} \int_{-\infty}^{x} e^{-u^2/2}\, du,$$

with

$$x = \left[1 + t^2\nu^{-1}\operatorname{var}(\chi_\nu)\right]^{-1/2}\left\{-\delta + t\nu^{-1/2}E[\chi_\nu]\right\}.$$

An approximation to the percentage point $t'_{\nu,\alpha}(\delta)$, defined by

$$\Pr\left[t'_\nu(\delta) \le t'_{\nu,\alpha}(\delta)\right] = \alpha,$$

is found by putting $x = u_\alpha$ and solving for t. The resulting approximation is

$$t'_{\nu,\alpha} = \frac{\delta b_\nu + u_\alpha\sqrt{b_\nu^2 + (1 - b_\nu^2)(\delta^2 - u_\alpha^2)}}{b_\nu^2 - u_\alpha^2(1 - b_\nu^2)}, \tag{31.26a}$$

where

$$b_\nu = \nu^{-1/2}E[\chi_\nu] = \frac{(2/\nu)^{1/2}\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\nu\right)},$$

$$\operatorname{var}(\chi_\nu) = \nu\left(1 - b_\nu^2\right).$$

Values of b_ν are given in Table 35 of Pearson and Hartley (1954) for $\nu = 1(1)20(5)50(10)100$, and also in van Eeden (1958).

Johnson and Welch (1940) introduced the further approximations $\operatorname{var}(\chi_\nu) = \frac{1}{2}$, and $E[\chi_\nu] \doteqdot \sqrt{\nu}$, giving the value $(1 + \frac{1}{2}t^2\nu^{-1})^{-1/2}(t - \delta)$ for x, and leading to the approximation

$$t'_{\nu,\alpha} \doteqdot \left[\delta + u_\alpha\sqrt{1 + \tfrac{1}{2}(\delta^2 - u_\alpha^2)\nu^{-1}}\right]\left[1 - \tfrac{1}{2}u_\alpha^2\nu^{-1}\right]^{-1}. \tag{31.26b}$$

[Masuyama (1951) showed how rough values of this approximation may be obtained using a special type of "improved binomial probability paper."] An approximation intermediate between (31.26a) and (31.26b) is obtained by using the correct value of $E[\chi_\nu]$ but replacing $\operatorname{var}(\chi_\nu)$ by $\frac{1}{2}$. This was given by van Eeden (1958) as

$$t'_{\nu,\alpha} = \frac{\delta b_\nu + u_\alpha\sqrt{b_\nu^2 + \frac{1}{2}(\delta^2 - u_\alpha^2)\nu^{-1}}}{b_\nu^2 - \frac{1}{2}u_\alpha^2\nu^{-1}} \tag{31.26c}$$

The preceding approximations give real values for $t'_{\nu,\alpha}$ only for limited ranges of δ and u,: For (31.26a) we must have

$$b_\nu^2 + (1 - b_\nu^2)(\delta^2 - u_\alpha^2) > 0;$$

that is, $$u_\alpha^2 < b_\nu^2(1 - b_\nu^2)^{-1} + \mathrm{S}^2.$$

For (31.26b) we must have

$$1 + \tfrac{1}{2}(\delta^2 - u_\alpha^2)\nu^{-1} > 0;$$

that is, $$u_\alpha^2 < 2\nu + \mathrm{S}^2.$$

For (31.26c) we must have

$$b_\nu^2 + \tfrac{1}{2}(\delta^2 - u_\alpha^2)\nu^{-1} > 0;$$

that is, $$u_\alpha^2 < 2\nu b_\nu^2 + \delta^2.$$

Since $2\nu b_\nu^2 < b_\nu^2(1 - b_\nu^2)^{-1} < 2\nu$, it follows that (31.26b) can be used over a wider range of values of a (for given δ) than (31.26a), and (31.26a) over a wider range than (31.26c). However, it should be borne in mind that when approaching limiting allowable values of a, the formulas may become less reliable. Also the wider range of (31.26b) is offset by its lower accuracy [van Eeden (1958)].

The problem of solving the inverse equation,

$$\Pr[t'_\nu(\delta) \le t] = \Pr\left[\frac{(U + \delta)\sqrt{\nu}}{\chi_\nu} \le t\right] = \alpha, \tag{31.27}$$

for δ, given t, ν, and a, needed to establish confidence intervals for δ was tackled by Deutler (1984), and he improved on Johnson and Welch's (1940) earlier result. Denoting the solution of (31.27) by $\delta(t;\nu,$a), the crude approximation [Johnson and Welch (1940)]

$$\delta(t;\nu,\alpha) \doteqdot b_\nu t - u_\alpha\sqrt{1 + (2\nu)^{-1}(t^2 - a_\nu^2)}, \tag{31.28}$$

where $\Phi(u_\alpha) =$ a, $b_\nu = \sqrt{(2/\nu)}\,\Gamma(\frac{1}{2}(\nu + 1))/\Gamma(\frac{1}{2}\nu)$ as in (31.26a), and $a_\nu^2 = 2\nu(1 - b_\nu^2)$, was refined by Deutler (1984) by utilizing a numerically stable

Cornish-Fisher expansion for $\chi_\nu/\sqrt{\nu}$, so that u_α is replaced by

$$z_{\nu,\alpha} = u_\alpha + A_1 + A_2 + A_3 + \cdots, \tag{31.29}$$

with

$$\begin{aligned}
A_1 &= \tfrac{1}{6}(u_\alpha^2 - 1)\zeta_3,\\
A_2 &= \tfrac{1}{24}(u_\alpha^3 - 3u_\alpha)\zeta_4 - \tfrac{1}{36}(2u_\alpha^3 - 5u_\alpha)\zeta_3^2,\\
A_3 &= \tfrac{1}{120}(u_\alpha^4 - 6u_\alpha^2 + 3)\zeta_5 + \tfrac{1}{24}(u_\alpha^4 - 5u_\alpha^2 + 2)\zeta_3\zeta_4\\
&\quad + \tfrac{1}{324}(12u_\alpha^4 - 53u_\alpha^2 + 17)\zeta_3^3,
\end{aligned}$$

where

$$\zeta_r = \left\{\frac{t^2}{1 + t^2 b_\nu^2/(2\nu)}\right\}^{r/2} (-1)^r \kappa_r\left(\frac{\chi_\nu}{\sqrt{\nu}}\right), \qquad r \geq 3, \tag{31.30}$$

is the rth cumulant of

$$\left\{U - \left(\frac{\chi_\nu}{\sqrt{\nu}}\right)t + a_\nu t\right\}\left\{1 + \frac{1}{2}b_\nu^2\nu^{-1}t^2\right\}^{-1/2}$$

[a_ν and b_ν are as defined in (31.28).] The relevant expansions in powers of ν^{-1} are

$$\begin{aligned}
\kappa_3(\chi_\nu/\sqrt{\nu}) &= \frac{1}{2^2\nu^2} + \frac{1}{2^4\nu^3} - \frac{13}{2^7\nu^4} - \frac{75}{2^9\nu^5} + \frac{1215}{2^{13}\nu^6}\\
&\quad + \frac{17403}{2^{15}\nu^7} - \frac{122101}{2^{18}\nu^8} - \frac{3371095}{2^{20}\nu^9} + \frac{88464187}{2^{25}\nu^{10}},\\
\kappa_4(\chi_\nu/\sqrt{\nu}) &= \frac{3}{2^4\nu^4} + \frac{3}{2^4\nu^5} - \frac{45}{2^7\nu^6} - \frac{57}{2^6\nu^7} + \frac{4875}{2^{12}\nu^8}\\
&\quad + \frac{24129}{2^{12}\nu^9} - \frac{226155}{2^{15}\nu^{10}},\\
\kappa_5(\chi_\nu/\sqrt{\nu}) &= -\frac{3}{2^4\nu^4} - \frac{9}{2^6\nu^5} + \frac{\text{[illegible]}}{2^9\nu^6} + \frac{2625}{2^{11}\nu^7} - \frac{19188}{2^{15}\nu^8}\\
&\quad - \frac{1321815}{2^{17}\nu^9} + \frac{17285517}{2^{20}\nu^{10}}.
\end{aligned}$$

This approximation is very good for ν greater than 5, and becomes more and more precise as ν increases.

Cornish-Fisher expansion for $\chi_\nu/\sqrt{\nu}$, so that u_α is replaced by

$$z_{\nu,\alpha} = u_\alpha + A_1 + A_2 + A_3 + \cdots, \tag{31.29}$$

with

$$A_1 = \tfrac{1}{6}(u_\alpha^2 - 1)\zeta_3,$$

$$A_2 = \tfrac{1}{24}(u_\alpha^3 - 3u_\alpha)\zeta_4 - \tfrac{1}{36}(2u_\alpha^3 - 5u_\alpha)\zeta_3^2,$$

$$A_3 = \tfrac{1}{120}(u_\alpha^4 - 6u_\alpha^2 + 3)\zeta_5 + \tfrac{1}{24}(u_\alpha^4 - 5u_\alpha^2 + 2)\zeta_3\zeta_4 + \tfrac{1}{324}(12u_\alpha^4 - 53u_\alpha^2 + 17)\zeta_3^3,$$

where

$$\zeta_r = \left\{\frac{t^2}{1 + t^2 b_\nu^2/(2\nu)}\right\}^{r/2} (-1)^r \kappa_r\left(\frac{\chi_\nu}{\sqrt{\nu}}\right), \qquad r \geq 3, \tag{31.30}$$

is the rth cumulant of

$$\left\{U - \left(\frac{\chi_\nu}{\sqrt{\nu}}\right)t + a_\nu t\right\}\left\{1 + \frac{1}{2}b_\nu^2 \nu^{-1} t^2\right\}^{-1/2}$$

[a_ν and b_ν are as defined in (31.28).] The relevant expansions in powers of ν^{-1} are

$$\kappa_3(\chi_\nu/\sqrt{\nu}) = \frac{1}{2^2\nu^2} + \frac{1}{2^4\nu^3} - \frac{13}{2^7\nu^4} - \frac{51}{2^9\nu^5} + \frac{1215}{2^{13}\nu^6} + \frac{17403}{2^{15}\nu^7} - \frac{122101}{2^{18}\nu^8} - \frac{3371095}{2^{20}\nu^9} + \frac{88464187}{2^{25}\nu^{10}},$$

$$\kappa_4(\chi_\nu/\sqrt{\nu}) = \frac{3}{2^4\nu^4} + \frac{3}{2^4\nu^5} - \frac{45}{2^7\nu^6} - \frac{57}{2^6\nu^7} + \frac{4875}{2^{12}\nu^8} + \frac{24129}{2^{12}\nu^9} - \frac{226155}{2^{15}\nu^{10}},$$

$$\kappa_5(\chi_\nu/\sqrt{\nu}) = -\frac{3}{2^4\nu^4} - \frac{9}{2^6\nu^5} + \frac{[illegible]}{2^9\nu^6} + \frac{2625}{2^{11}\nu^7} - \frac{19188}{2^{15}\nu^8} - \frac{1321815}{2^{17}\nu^9} + \frac{17285517}{2^{20}\nu^{10}}.$$

This approximation is very good for ν greater than 5, and becomes more and more precise as ν increases.

Johnson and Welch *(1940)* also provided tables from which $\delta(t;\nu,\alpha)$ can be calculated directly. The tables provide values of $\lambda(t;\nu,\alpha)$, and $\delta(t;\nu,\alpha)$ is calculated from

$$\delta(t;\nu,\alpha) = t - \lambda(t;\nu,\alpha)\sqrt{1+\frac{t^2}{2\nu}}. \tag{31.31}$$

Values of $\lambda(t;\nu,\alpha)$ are tabulated for

$$\alpha = 0.5(0.1)0.9, 0.95, 0.975, 0.99, 0.995,$$

$$\nu = 4(1)9, 16, 36, 144, \infty.$$

The higher values of ν are selected to facilitate interpolation with respect to $12/\sqrt{\nu}$.

In Kühlmeyer's *(1970)* monograph, $\lambda(t;\nu,\alpha)$ is tabulated against $y = \{1 + t^2(2\nu)^{-1}\}^{-1/2}$ for $|t|/\sqrt{2\nu} \le 0.75$ and against

$$y' = t(2\nu + t^2)^{-1/2}\ \left[= \sqrt{(1-y^2)}\right]$$

for $|t|/\sqrt{2\nu} < 0.75$ for $\alpha = 0.99, 0.95, 0.90,$ and the same values of ν. [For $\alpha = 0.01, 0.05, 0.10$ the identity $\delta(t;\nu,\alpha) = -\delta(-t;\nu,1-\alpha)$ can be used.]

Direct approximations to the distribution of noncentral t are of later date. For small values of δ and large (> 20) values of ν, the simple approximation of the standardized t'_ν variable by a standard normal variable gives fair results. This is equivalent to using the formula

$$t'_{\nu,\alpha}(\delta) \doteqdot \frac{\nu}{\nu-1}\delta b_\nu + u_\alpha\sqrt{\frac{\nu}{\nu-2}(1+\delta^2) - \left(\frac{\nu}{\nu-1}\right)^2\delta^2 b_\nu^2}. \tag{31.32}$$

Since $\operatorname{var}(t'_\nu)$ is not finite if $\nu \le 2$, this formula cannot be used for $\nu \le 2$. As we have implied above, it is in fact unlikely to be useful unless ν is fairly large and δ is fairly small. When $\delta = 0$, the approximation becomes

$$t_{\nu,\alpha} \doteqdot u_\alpha\sqrt{\frac{\nu}{\nu-2}},$$

which has already been noted, as a fair approximation, in Section 4 of Chapter 28.

A better approximation is obtained, as is to be expected, if a Pearson Type IV distribution is fitted, making the first four moments agree with those of the noncentral t-distribution. Merrington and Pearson (1958) found that this gives upper and lower 5%, 1%, and 0.5% points with an error no greater than 0.01, for a considerable range of values of δ and ν (including ν as small as 8). Further investigations, by **Pearson (1963),** confirmed the closeness of the two systems of distributions.

If a Cornish-Fisher expansion be applied to the distribution of $t'_\nu(\delta)$, the following approximate expansion (up to and including terms in ν^{-2}) is obtained:

$$t'_{\nu,\alpha}(\delta) \doteqdot u_\alpha + \delta + \tfrac{1}{4}\left[u_\alpha^3 + u_\alpha + \left(2u_\alpha^2 + 1\right)\delta + u_\alpha\delta^2\right]\nu^{-1}$$
$$+ \tfrac{1}{96}\left[5u_\alpha^5 + 16u_\alpha^3 + 3u_\alpha + 3\left(4u_\alpha^4 + 12u_\alpha^2 + 1\right)\delta\right.$$
$$\left. + 6\left(u_\alpha^3 + 4u_\alpha\right)\delta^2 - 4\left(u_\alpha^2 - 1\right)\delta^3 - 3u_\alpha\delta^4\right]\nu^{-2}. \quad (31.33\text{a})$$

Setting $\delta = 0$ in (31.33a), we obtain the approximation [see Chapter 28, Eq. (28.16)] for central t_ν percentage points:

$$t_{\nu,\alpha} \doteqdot u_\alpha + \tfrac{1}{4}\left(u_\alpha^3 + u_\alpha\right)\nu^{-1} + \tfrac{1}{96}\left(5u_\alpha^5 + 16u_\alpha^3 + 3u_\alpha\right)\nu^{-2}.$$

If these terms in (31.33a) be replaced by the exact value of $t_{\nu,\alpha}$, then the approximation becomes

$$t'_{\nu,\alpha}(\delta) \doteqdot t_{\nu,\alpha} + \text{S} + \tfrac{1}{4}\delta\left(1 + 2u_\alpha^2 + u_\alpha\delta\right)\nu^{-1}$$
$$+ \tfrac{1}{96}\delta\left[3\left(4u_\alpha^4 + 12u_\alpha^2 + 1\right) + 6\left(u_\alpha^3 + 4u_\alpha\right)\delta\right.$$
$$\left. - 4\left(u_\alpha^2 - 1\right)\delta^2 - 3u_\alpha\delta^3\right]\nu^{-2}. \quad (31.33\text{b})$$

Extensive numerical comparisons for v = 2(1)9, given by van Eeden (1958) indicate that, for 6 > 0, formula (31.33a) gives the better results for lower percentage points ($\alpha < \tfrac{1}{2}$), while (31.33b) is better for a > $\tfrac{1}{2}$.

Azorín (1953) obtained another type of approximation by constructing an approximate variance equalizing transformation. Starting from the relationship

$$\operatorname{var}(t'_\nu) = a^2 + b^2\{E[t'_\nu]\}^2,$$

with

$$a = \sqrt{\frac{\nu}{\nu - 2}},$$

$$b = \Gamma\left(\tfrac{1}{2}\nu\right)\sqrt{2\left\{(\nu - 2)\left[\Gamma\left(\tfrac{1}{2}(\nu - 1)\right)\right]^2 - 1\right\}^{-1}},$$

we obtain the transformation

$$\frac{1}{b}\sinh^{-1}\frac{bt'_\nu}{a} - \sinh^{-1}\frac{bE[t'_\nu]}{a} \quad (31.34\text{a})$$

which is to be approximated as a standard normal variable. This transformation was studied by Laubscher (1960). Azorín suggested two similar transformations of simpler form (each to be approximated as a standard normal variable):

$$\sqrt{\nu}\,\sinh^{-1}\left(\frac{t'_\nu}{\sqrt{\nu}}\right) - \delta, \tag{31.34b}$$

$$\sqrt{\tfrac{2}{3}\nu}\,\sinh^{-1}\left(\frac{t'_\nu}{\sqrt{(2/3)\nu}}\right) - \delta. \tag{31.34c}$$

These transformations approximate noncentral t by S_U distributions (see Chapter 12, Section 4.3). Transformations of type (31.34) would be expected to give accuracy comparable to the Type IV approximation, in view of the close similarity between Type IV and S_U-distributions.

A remarkably accurate transformation was suggested by Harley (1957). She suggested that the distribution of $t'_\nu(\delta)$ be approximated by that of a function of the sample correlation coefficient R (Chapter 32) in random samples of size $\nu + 2$ from bivariate normal population with the population correlation coefficient

$$\rho = \delta\sqrt{\frac{2}{2\nu + 1 + \delta^2}}.$$

The proposed function is

$$\frac{R}{\sqrt{1 - R^2}}\sqrt{\frac{\nu(2\nu + 1)}{2\nu + 1 + \delta^2}}. \tag{31.35}$$

(see Chapter 32, Section 2).

While percentage points of t'_ν can be approximated from those of R, using (31.35), it is also possible to approximate percentage points of R, using those of t'_ν. It is this latter use that appears to be the more valuable, in the opinion of van Eeden (1958) and Owen (1963).

Hogben, Pinkham, and Wilk (1964) approximate the distribution of $Q = t'_\nu(\nu + t'^2_\nu)^{-1/2}$ and hence the distribution of t'_ν. Fitting a Type I (beta) distribution (with range $-1, 1$) to Q is of course equivalent to approximating the distribution of t'_ν. This approximation is claimed to be especially useful for small values of 6.

Mention may be made of approximations proposed by Halperin (1963). These are not of great accuracy but appear to provide a bound for percentage points. They are also simple to compute, using tables of percentage points of central t_ν ($t_{\nu,\alpha}$) and χ_ν^2 ($\chi_{\nu,\alpha}^2$). The suggested inequalities are (assuming that 6 > 0)

$$t'_{\nu,\alpha}(\delta) \le \frac{\delta\sqrt{\nu}}{\chi_{\nu,1-\alpha}} + t_{\nu,\alpha}, \qquad \alpha \ge \tfrac{1}{2}, \tag{31.36a}$$

$$t'_{\nu,\alpha}(\delta) \ge \frac{\delta\sqrt{\nu}}{\chi_{\nu,1-\alpha}} + t_{\nu,\alpha} \qquad \alpha \le 0.43. \tag{31.36b}$$

Kraemer (1978) approximates t'_ν cdfs by means of (central) t-distribution cdfs. The approximation is based on the following result, proved by Kraemer (1978): A function $\theta \equiv \theta(\delta,\nu)$ exists such that

$$\lim_{\nu\to\infty}\left\{\Pr[t'_\nu(\delta) \le t] - \Pr\left[t_\nu \le (g-\theta)\{(1-g^2)(1-\theta^2\nu^{-1})\}^{-1/2}\right]\right\} = 0, \tag{31.37}$$

where g = $t(t^2+\nu)^{-1/2}$. Kraemer (1978) found empirically that good results are obtained with $\theta(\delta,\nu) = \delta(\delta^2+\nu)^{-1/2}$. Using this value, together with Johnson and Welch's (1940) second normal approximation, she obtained

$$\Pr[t'_\nu(\delta) \le t] \doteqdot \Phi\left[(t-\delta)\left(1+\frac{t^2}{2\nu}\right)^{-1/2}\right], \tag{31.38}$$

which yields

$$\Phi\left[(t-\delta)\left(1+\frac{t^2}{2\nu}\right)^{-1/2}\right] \doteqdot \Pr\left[t_\nu < t\left(1+\frac{\delta^2}{\nu}\right)^{1/2} - \delta\left(1+\frac{t^2}{\nu}\right)^{1/2}\right]. \tag{31.39}$$

For moderate sized ν and $\delta > 0$, neither approximation (31.38) nor (31.39) is uniformly better. Kraemer (1978) suggested that in estimating 95th percentiles, the t-approximation would be better for $\delta < 2$, but the normal approximation would be better for 6 > 2. "One-tailed tests, and confidence intervals based on the t approximation will tend to be on the conservative side; the normal approximation will tend to err on the liberal side." Kraemer's approximation complements the existing approximations being most accurate for parameter values for which tables of the exact distributions are rather sparse.

7 TABLES, CHARTS, AND COMPUTER ALGORITHMS

The earliest tables of the noncentral t-distribution were produced by Jerzy Neyman and his co-workers in the mid-thirties of this century. Substantial contributions to the tabulation of this distribution were carried out in the early sixties by several authors to be indicated below. The most comprehensive volume of tables appeared in 1993. For historical interest and perspective, we present a rather detailed description of the earlier contributions.

The tables of Neyman, Iwaszkiewicz, and Kolodziejczyk (1935) and Neyman and Tokarska (1936) were calculated to give the power of the t-test. The first paper (Table III) gives the operating characteristic (= 1 − Power) of a 5% significance limit t-test at values 6 = 1(1)9 for $\nu = 1(1)30$ (i.e., values of $\Pr[t'_\nu(\delta) \le t_{\nu,0.95}]$) and also values of 6 satisfying the equation $\Pr[t'_\nu(\delta) \le t_{\nu,0.95}] = 0.05$. The second paper gives more extensive tables of the same kind. Owen (1965) gave tables, to five decimal places, of values of 6 such that $\Pr[t'_\nu(\delta) \le t_{\nu,1-\alpha}] = \beta$ for

$$\nu = 1(1)30(5)100(10)200, \infty,$$

$$\alpha = 0.005, 0.01, 0.025, 0.05,$$

$$\beta = 0.01, 0.05, 0.1(0.1)0.9.$$

The tables of Johnson and Welch (1940) give values of a quantity A such that

$$t'_{\nu,\alpha}(\delta) = \frac{\delta + \lambda\left[1 + \frac{1}{2}(\delta^2 - \lambda^2)\nu^{-1}\right]^{1/2}}{1 - \frac{1}{2}\lambda^2\nu^{-1}}. \tag{31.40}$$

Comparing with approximation (31.26b), it is seen that one might expect $A = u_\alpha$, so values of A should not vary too much as 6 and ν vary and interpolation becomes simpler. Values of A are given for $\nu = 4(1)9, 16, 36, 144, \infty$ (i.e. u_α); above $\nu = 9$, interpolation with respect to $12\nu^{-1/2}$ is suggested. The argument used is $y = (1 + \frac{1}{2}t'^2_\nu\nu^{-1})^{-1/2}$ for $0.6 \le |y| \le 1$, and $y' = yt'_\nu/\sqrt{2\nu}$ for $|y| \le 0.6$. Values of a [$(1-\epsilon)$ in the original] are 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.975, 0.99, and 0.995. Since the argument is a function of t'_ν, direct entry into the tables leads to a value

$$\delta' = \delta(\nu, t', \alpha) = t' - \frac{\lambda}{y} \tag{31.41}$$

such that $t'_{\nu,\alpha}(\delta') = t'$. To obtain $t'_{\nu,\alpha}(\delta)$ for a given δ, an iterative (or inverse interpolation) procedure is necessary. A table giving the results of such procedures, for the case $\alpha = 0.95$ only, is also provided. This table gives A to three or four decimal places as a function of $\eta = [S/\sqrt{2\nu}](1 + \frac{1}{2}\delta^2\nu^{-1})^{-1/2}$. The later tables of Owen (1963) include a substantial extension of this last table. The argument η is tabulated at intervals of 0.01, instead of 0.1; A is given to five decimal places, and values $\nu = 1, 2, 3$ are included, in addition to those in the Johnson and Welch tables. Values of α, however, are now 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995. Owen also gives tables of A (to five decimal places) as a function of γ and γ' (in the same form as Johnson and Welch) for these new sets of values of ν and α. There are also extensive tables of quantity k (to three decimal places) such that

$$t'_{\nu,\alpha}\left(u_p\sqrt{\nu + 1}\right) = k\sqrt{\nu + 1} \tag{31.42}$$

for $p = 0.75$, 0.9, 0.95, 0.975, 0.99, 0.999, 0.9999, 0.99999, and $\nu + 1 =$ 2(1)200(5)400(25)1000(500)2000, 3000, 5000, 10000, ∞; α has the same values as in Owen's other tables. The choice of $u_p\sqrt{\nu + 1}$ as noncentrality parameter makes it convenient to use the tables in calculating confidence intervals for percentage points of normal distributions. Thus, since the inequality $\bar{X} - kS < \xi - u_p\sigma$ is equivalent to

$$\frac{\sqrt{n}\,(\bar{X} - \xi)\sigma^{-1} + \sqrt{n}u_p}{S\sigma^{-1}} < k\sqrt{n}\,,$$

it follows that

$$\Pr\left[\bar{X} - kS < \xi - u_p\sigma\right] = \Pr\left[t'_{n-1}\left(u_p\sqrt{n}\right) < k\sqrt{n}\right] \qquad [\text{cf. } (31.7)]. \tag{31.43}$$

This probability is equal to α if

$$t'_{n-1,\alpha}\left(u_p\sqrt{n}\right) = k\sqrt{n}\,.$$

Setting $n = \nu + 1$, we obtain Eq. (31.42).

Owen also gives values of k satisfying

$$t'_{\nu,\alpha}\left(u_p\sqrt{n}\right) = k\sqrt{n} \tag{31.44}$$

for $n = 1$, $\nu = 1, 2$ (and ∞), $\alpha = 0.90$, 0.95, 0.99, and for $p =$ 0.50(0.01)0.90(0.005)0.990(0.001)0.999(0.0001)0.9999 and some even higher values of p. Further tables give values of k satisfying (31.44) for

$$n = 1(1)10, 17, 37, 145, 500, \infty,$$
$$\alpha = 0.90, 0.95, 0.99,$$
$$p = 0.75, 0.9, 0.95, 0.975, 0.99, 0.999, 0.9999, 0.99999,$$
$$\nu = 1(1)75(5)100(10)150, 200, 300, 500, 1000.$$

Some extracts from Owen's tables are included in the survey by Owen (1968).

The tables of Resnikoff and Lieberman (1957) are also based on Eq. (31.42). They give $t'_{\nu,\alpha}(u_p\sqrt{\nu+1})/\sqrt{\nu}$ for

$$p = 0.001, 0.0025, 0.004, 0.01, 0.025, 0.04, 0.065, 0.1, 0.15, 0.25,$$
$$\alpha = 0.005, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99, 0.995,$$
$$\nu = 2(1)24(5)49.$$

(For $\nu = 2$, 3, and 4, values are not given corresponding to $\alpha = 0.99, 0.995$.)

These tables also give probability integrals of the noncentral t-distribution (to four decimal places)

$$\Pr\left[t'_\nu\left(u_p\sqrt{\nu+1}\right) \le x\sqrt{\nu}\right] \tag{31.45}$$

for the same values of p and ν, and for x at intervals of 0.05. There is also a table of values of the probability density function, for the same values of p, ν, and x.

Locks, Alexander, and Byars (1963) have produced a similar, but more extensive set of tables of the probability density function. To facilitate use in connection with multiple regression, values of the noncentrality parameter equal to $u_p\sqrt{\nu+2}$ and $u_p\sqrt{\nu+1}$ are used. However, the tables do not include specified values of p; they rather take $u_p = 0.0(0.025)3.0$.

Hodges and Lehmann (1968) noted that if normal equivalent deviates are used, problems of interpolation are much less troublesome. Thus, if

$$\Pr[t'_\nu(\delta) \ge t_{\nu,1-\alpha}] = \beta,$$

then they table a quantity A satisfying the equation

$$u_{1-\alpha} + u_\beta = \delta\left(1 - \tfrac{1}{4}u_{1-\alpha}^2\nu^{-1}\right) - Au_{1-\alpha}^2(u_{1-\alpha} + u_\beta)\nu^{-2}. \tag{31.46}$$

Values of A are given to four decimal places for

$$\alpha = 0.005, 0.01, 0.025, 0.05, 0.1,$$
$$\beta = 0.5(0.1)0.9, 0.95, 0.99, \quad \text{and} \quad \nu = 3(1)6, 8, 12, 24, \infty.$$

For given α and β, A is a rather smooth function of ν. For $\nu > 6$, $\alpha \geq 0.01$, and $\beta \leq 0.09$, linear harmonic interpolation give good results, and practically useful values can be obtained well beyond these limits.

Among shorter useful tables, mention must be made of those of van Eeden (1961) and Scheuer and Spurgeon (1963). Van Eeden gives $t'_{\nu,\alpha}(u_p\sqrt{\nu+1})$ directly [to three decimal places (two for $\alpha = 0.99, 0.01$)] for $\nu = 4(1)9$, $\alpha = 0.01, 0.025, 0.05, 0.95, 0.99$, and $p = 0.125, 0.15(0.05)0.45$. Scheuer and Spurgeon give values of the same function (to three decimal places) for the values of p and ν used by Resnikoff and Lieberman (1957), but only for $\alpha = 0.025, 0.975$.

Bruscantini (1968) made a detailed study of the distribution of $Y = U + \theta\chi_2$. He referred to, and gave a short extract from, unpublished tables giving values, to five decimal places of the cumulative distribution function of the standardized variable

$$Y' = \left(Y - \theta\sqrt{\frac{\pi}{2}}\right)\left[1 + \theta^2\left(2 - \tfrac{1}{2}\pi\right)\right]^{-1/2} \tag{31.47}$$

for argument y' at intervals of 0.5 and $\theta = 2.00(0.05)7.20$. These are in fact values of $\Pr[t'_2(y'') > \theta]$ with

$$y'' = \theta\sqrt{\frac{\pi}{2}} + y'\sqrt{1 + \theta^2\left(2 - \tfrac{1}{2}\pi\right)}\,. \tag{31.48}$$

Recently [Bagui (1993)] extensive tables of values of $t'_{\nu,\alpha}(\delta)$ have appeared. Included are values to five decimal places for

$$\alpha = 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 0.95, 0.975, 0.99,$$

$$\delta = 0.1(0.1)8.0,$$

$$\nu = 1(1)60.$$

Computer algorithms for evaluation of cumulative distribution function of the noncentral t-distribution were provided by Cooper (1968) (algorithm S5) [reprinted and updated in Griffiths and Hill (1985)]—one of the first algorithms in the newly established algorithm section of Applied Statistics, Lenth (1989), Narula and Weistroffer (1986), Guirguis (1990), and Posten (1993), among others.

Cooper's algorithm, written in standard FORTRAN, uses the numerical method given by Owen (1965b). The integration is theoretically exact provided that the auxiliary functions can be evaluated exactly. The essential part involves the auxiliary function

$$T(h,a) = \frac{\tan^{-1} a}{2\pi} - \frac{1}{2\pi}\int_0^h \int_0^{ax} \exp\left(-\frac{x^2+y^2}{2}\right) dx\,dy$$

$$= \frac{1}{2\pi}\int_0^a \frac{\exp\{(-h^2/2)(1+x^2)\}}{1+x^2}\,dx. \qquad (31.49)$$

For $h \geq a$ Cooper (1968) used the approximation

$$T(h,a) = \frac{\tan^{-1} a}{2\pi} - \frac{1}{2}\frac{1}{\sqrt{(2\pi)}}\int_0^h \exp\left(\frac{-u^2}{2}\right) du$$

$$+ \frac{\tan^{-1}(1/a)}{2\pi} \qquad \text{(Cooper's algorithm AS 4).}$$

Another subroutine used is the normal integral (Cooper's algorithm S2). For large degrees of freedom Cooper used normal approximation, and fractional degrees of freedom were not allowed. The accuracy claimed by the author is to more than six decimal places. However, for ν near 100 the error of the normal approximation is in the vicinity of 5×10^{-4}.

Lenth's (1989) algorithm is based on the infinite series expansion (31.21) due to Guenther. The error E_n incurred by using a finite sum terminating at $j = n$ is bounded by

$$|E_n| < 2\left(1 - \sum_{j=0}^{n} p_j\right) I_x\left(n + \frac{3}{2}, \frac{\nu}{2}\right),$$

where, as above,

$$p_j = \frac{\frac{1}{2}e^{-\delta^2/2}(\delta^2/2)^j}{j!} \qquad \text{[See also Singh, Relyea, and Bartolucci (1992)].}$$

Accuracy of Lenth's algorithm to within $\pm 10^{-6}$ is guaranteed within 100 terms for $-11.0 \leq \delta \leq 11.0$. For evaluation of the incomplete beta function, an updated version of Majumder and Bhattacharjee's algorithm (1973) is used. In the majority of cases Cooper's algorithm is faster than Lenth's—the price paid for increased generality. Experience shows that on an IBM PC (Microsoft Fortran 77, standard library) no errors larger than 10^{-6} were observed using Lenth's (1989) algorithm.

Singh, Relyea, and Bartolucci *(1992)* and Posten (1993) give algorithms also based on Guenther's formula *(31.21).* Here we describe Posten's algorithm in some detail. It uses the expansion [cf. (31.21)]

$$\Pr[t'_\nu \le t] = 1 - \frac{1}{2}e^{-\delta^2/2} \sum_{j=0}^{\infty} \frac{(\delta/\sqrt{2})^j}{\Gamma\left(\frac{j}{2}+1\right)} I_x\left(\frac{\nu}{2}, \frac{j+1}{2}\right), \tag{31.50}$$

with $x = \nu/(\nu + t^2)$. The basic problem is to evaluate the truncated sum

$$\sum_{j=0}^{2n} T_j B_j,$$

where

$$T_j = \frac{(\delta/\sqrt{2})^j}{\Gamma\left(\frac{j}{2}+1\right)},$$

and

$$B_j = I_x\left(\frac{\nu}{2}, \frac{j+1}{2}\right).$$

The recursive formulas for evaluation of this sum are as follows: Set

$$D_i = T_{2i} \quad \text{and} \quad E_i = T_{2i+1},$$

then

$$\sum_{j=0}^{2n} T_j B_j = \sum_{i=0}^{n} (D_i B(i) + E_i BB(i)),$$

with $D_0 = 1$, $E_0 = \delta\sqrt{2/\pi}$, $D_i = (\lambda/i)D_{i-1}$, $E_i = \{\lambda/(i+\frac{1}{2})\}E_{i-1}$, and

$$B(0) = I_x\left(\tfrac{1}{2}\nu, \tfrac{1}{2}\right),$$

$$BB(0) = I_x\left(\tfrac{1}{2}\nu, 1\right),$$

$$B(i) = B(i-1) + S(i-1),$$

$$BB(i) = BB(i-1) + SS(i-1),$$

$$S(0) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{1}{2}\nu\right)} x^{\nu/2}(1-x)^{1/2},$$

$$SS(0) = \frac{\Gamma\left(\frac{1}{2}\nu + 1\right)}{\Gamma\left(\frac{1}{2}\nu\right)} x^{\nu/2}(1-x),$$

$$S(i) = (1-x)\frac{\nu + 2i - 1}{1 + 2i} S(i-1),$$

$$SS(i) = (1-x)\frac{\nu + 2i}{2 + 2i} SS(i-1).$$

For evaluation of $S(0)$ and $SS(0)$, Posten's (1986) algorithm might be used. Singh, Relyea, and Bartolucci (1992) have used a similar method.

Posten (1993) conducted a preliminary study of the algorithm over a range of values of t and 6, with degrees of freedom 4, 19, and 39. The calculated results (in double precision on an IBM 3000 series mainframe computer) were compared with results using the Cooper-Owen (1965) algorithm used in IMSL (1987). In most cases it was easy to obtain 12 or more place accuracy with fewer than N = 200 terms and often with N less than 100.

Chattamvelli and Shanmugam (1994) recently presented an algorithm for the noncentral t-distribution which does not require the evaluation of the incomplete beta function. Consequently, this algorithm is suitable for the evaluation of the noncentral t distribution function even on calculators. These authors also have presented a step-by-step algorithm which may easily be programmed.

8 RELATED DISTRIBUTIONS

8.1 Noncentral Beta Distribution

The noncentral beta distribution is defined as the distribution of the ratio

$$b'_{\nu_1,\nu_2}(\lambda) = \frac{\chi'^2_{\nu_1}(\lambda)}{\chi^2_{\nu_2} + \chi'^2_{\nu_1}(\lambda)}. \tag{31.51}$$

See Chapter 30 (Section 7). It can be seen that $[t'_\nu(\delta)]^2/[\nu + \{t'_\nu(\delta)\}^2]$ is distributed as $b'_{1,\nu}(\delta^2)$. This is also the distribution of Q^2, where Q is the variable (mentioned in Section 6) studied by Hogben, Pinkham, and Wilk (1964). [See also David and Paulson (1965, p. 434).]

8.2 Doubly Noncentral t-Distribution

If the χ_ν in the denominator of $t'_\nu(\delta)$ is replaced by a noncentral χ_ν (noncentrality parameter A), the distribution of the modified variable is called a doubly *noncentral* t-distribution with noncentrality parameters (δ, λ) and ν degrees of freedom. Symbolically

$$t''_\nu(\delta, \lambda) = \frac{U + \delta}{\chi'_\nu(\lambda)/\sqrt{\nu}}. \tag{31.52}$$

Since $\chi'_\nu(\lambda)$ is distributed as a mixture of $\chi_{\nu+2j}$-distributions in proportions $e^{-\lambda/2}(\frac{1}{2}\lambda)^j/j!$ $(j = 0, 1, 2, \ldots)$, the distribution of $t''_\nu(\delta, \lambda)$ is a mixture of

$$\sqrt{\nu(\nu + 2j)^{-1}}\, t'_{\nu+2j}(\delta)$$

distributions in these same proportions. Hence all formulas, approximations, tables, and so forth, for the noncentral t-distribution can be applied to the doubly noncentral t-distribution. For example, the rth moment of t''_ν about zero is

$$E[t''^r_\nu] = \left(\frac{1}{2}\nu\right)^{r/2} e^{-\lambda/2}\left[\sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j}{j!} \frac{\Gamma(\frac{1}{2}\nu)\Gamma(\frac{1}{2}(\nu-r)+j)}{\Gamma(\frac{1}{2}(\nu-r))\Gamma(\frac{1}{2}\nu+j)}\right]$$

$$\times E\left[(U+\delta)^r\right]\frac{\Gamma(\frac{1}{2}(\nu-r))}{\Gamma(\frac{1}{2}\nu)}$$

$$= \left(\frac{\nu}{2}\right)^{r/2} E\left[(U+\delta)^r\right]\sum_{\alpha=0}^{\infty}\tau_\alpha\left(\frac{\lambda}{2}\right)\frac{\Gamma((\nu+2\alpha-r)/2)}{\Gamma((\nu+2\alpha)/2)} \tag{31.53}$$

for $r < \nu$, where $\tau_j(\theta) = e^{-\theta}\theta^j/j!$. Krishnan (1967) pointed out that the summation in this formula can be expressed as the confluent hypergeometric function $M(\frac{1}{2}(\nu-r); \frac{1}{2}\nu; \frac{1}{2}\lambda)$ and that an even simpler form can be obtained using **Kummer's** formula $e^{-\lambda/2}M(\frac{1}{2}(\nu-r); \frac{1}{2}\nu; \frac{1}{2}\lambda) = M(\frac{1}{2}r; \frac{1}{2}\nu; -\frac{1}{2}\lambda)$.

Kocherlakota and Kocherlakota (1991) have provided the explicit formulas

$$E[t''^r_\nu] = \left(\frac{\nu}{2}\right)^{r/2} E\left[(U+\delta)^r\right]\frac{\Gamma((\nu-r)/2)}{\Gamma(\nu/2)}M\left[\frac{r}{2};\frac{\nu}{2};-\frac{\lambda}{2}\right], \tag{31.54}$$

and, in particular,

$$E[t''_\nu] = \delta\left(\frac{\nu}{2}\right)^{1/2}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}M\left[\frac{1}{2};\frac{\nu}{2};-\frac{\lambda}{2}\right], \tag{31.55a}$$

$$E\left[t''^2_\nu\right] = \frac{\delta^2}{\nu-2}M\left[1;\frac{\nu}{2};-\frac{\lambda}{2}\right], \tag{31.55b}$$

$$E\left[t''^3_\nu\right] = \delta(\delta^2+3)\left(\frac{\nu}{2}\right)^{3/2}\frac{\Gamma((\nu-3)/2)}{\Gamma(\nu/2)}M\left[\frac{3}{2};\frac{\nu}{2};-\frac{\lambda}{2}\right], \tag{31.55c}$$

$$E\left[t''^4_\nu\right] = (\delta^4+6\delta^2+3)\frac{\nu^2}{(\nu-2)(\nu-4)}M\left[2;\frac{\nu}{2};-\frac{\lambda}{2}\right]. \tag{31.55d}$$

Krishnan (1967) obtained some recurrence relations between moments about zero of $t''_\nu(\delta, \lambda)$, $t''_{\nu-2}(\delta, \lambda)$, and $t''_{\nu-4}(\delta, \lambda)$. These formulas are most conveniently given in terms of the quantities

$$\mu'_{r,\nu} = \frac{\mu'_r(t''_\nu)}{\nu^{r/2}}.$$

They are

$$\mu'_{1,\nu} = \left(1 - \frac{\nu - 4}{\lambda}\right)\mu'_{1,\nu-2} + \frac{\nu - 5}{\lambda}\mu'_{1,\nu-4}, \qquad (\nu > 5), \quad (31.56a)$$

$$\mu'_{2,\nu} = \lambda^{-1}(\delta^2 + 1 - \mu'_{2,\nu-2}), \qquad (\nu > 4), \quad (31.56b)$$

$$\mu'_{3,\nu} = (\delta^2 + 3)(\mu'_{1,\nu-2} - \mu'_{1,\nu}), \qquad (\nu > 3), \quad (31.56c)$$

$$\mu'_{4,\nu} = \tfrac{1}{2}(\delta^4 + 6\delta^2 + 3)(\delta^2 + 1)^{-1}(\mu'_{2,\nu-2} - \mu'_{2,\nu}). \qquad (\nu > 4). \quad (31.56d)$$

For ν large, δ and A remaining fixed,

$$\mu'_1(t''_\nu) = \delta\left[1 + \left(\tfrac{3}{4} - \tfrac{1}{2}\lambda\right)\nu^{-1} + O(\nu^{-2})\right], \quad (31.57a)$$

$$\mu'_2(t''_\nu) = (\delta^2 + 1)\left[1 + (2 - \lambda)\nu^{-1} + O(\nu^{-2})\right], \quad (31.57b)$$

$$\mu'_3(t''_\nu) = \delta(\delta^2 + 3)\left[1 + 3\left(\tfrac{5}{4} - \tfrac{1}{2}\lambda\right)\nu^{-1} + O(\nu^{-2})\right], \quad (31.57c)$$

$$\mu'_4(t''_\nu) = (\delta^4 + 6\delta^2 + 3)\left[1 + (6 - 2\lambda)\nu^{-1} + O(\nu^{-2})\right]. \quad (31.57d)$$

Since the available tables (at the time of writing) of hypergeometric and gamma functions were inadequate to be used for computing moments of $t''_\nu(\delta,$ A), Krishnan (1967) also presented tables, to six decimal places, for A = 2(2)8(4)20 of

$$c_1 = \frac{\mu'_1(t''_\nu)}{\delta} \qquad \text{for } \nu = 2(1)20,$$

$$c_2 = \frac{\mu'_2(t''_\nu)}{\delta^2 + 1} \qquad \text{for } \nu = 3(1)20,$$

$$c_3 = \frac{\mu'_3(t''_\nu)}{\delta(\delta^2 + 3)} \qquad \text{for } \nu = 4(1)20,$$

$$c_4 = \frac{\mu'_4(t''_\nu)}{\delta^4 + 6\delta^2 + 3} \qquad \text{for } \nu = 5(1)20.$$

(Note that the c's are independent of δ.)

In the same paper Krishnan considered two approximations to the distribution of $t''_\nu(\delta,$ A). In three special cases good results were obtained using a method suggested by Patnaik (1955) in which the distribution is approximated by that of $ct'_f(\delta)$, c and f being chosen to give correct values for the first two moments. The other method, an extension of Harley's (1957) method (see Section 6), also gave useful, though not quite as accurate, results. For this approximation one calculates $L = [(\nu - 3)\mu'_{3,\nu}][\nu\mu'_{1,\nu}]^{-1}$, $K = (1 - 2\nu^{-1})\mu'_{2,\nu}$, and

$$\rho = \left[(3K - L)\{\nu L - (\nu - 1)K\}^{-1}\right]^{1/2}.$$

Then the distribution of $t''_\nu(\delta, \lambda)$ is approximately that of

$$\left[\nu K\left\{1 + (\nu + 1)\frac{\rho^2}{1 - \rho^2}\right\}^{-1}\right]^{1/2} \frac{R}{\sqrt{1 - R^2}}, \qquad (31.58)$$

where R is distributed as the product moment sample correlation in a sample of size $\nu + 2$ from a bivariate normal population with correlation ρ (see Chapter 32, Section 2).

Krishnan (1968) gave tables of $\Pr[t''_\nu(\delta, \lambda) \le 1]$ to four decimal places for $\nu = 2(1)20$, $\delta = -5(1)5$, $\lambda = 0(2)8$. (Note that for $t_0 < 0$, $\Pr[t''_\nu(\delta, \lambda) \le t_0]$ can be evaluated as $\Pr[t''_\nu(-\delta, \lambda) \ge -t_0]$.)

Krishnan, and also Bulgren and Amos (1968), gave the following formula involving double summation:

$$\Pr[t''_\nu \le t_0] = 1 - \Phi(\sqrt{\beta}) + \varphi(\sqrt{\beta}) \sum_{j=0}^{\infty} \left\{\frac{(\frac{1}{2}\lambda)^j e^{-\lambda/2}}{j!}\right\}$$
$$\times \left[\frac{\sqrt{a}\,\Gamma(\frac{1}{2}(\nu + 1) + j)}{\Gamma(\frac{1}{2}\nu + j)} \sum_{i=0}^{\infty} \frac{(1 - \frac{1}{2}\nu - j)^{[i]}}{i!\,(2i + 1)} a^i{}_1F_1\left(-i; \frac{1}{2}; \frac{1}{2}\beta\right)\right.$$
$$\left. - \sqrt{\frac{1}{2}\beta} \sum_{i=0}^{\infty} \frac{(1 - \frac{1}{2}\nu - j)^{[i]}}{i!} a^i{}_1F_1(1 - i; \tfrac{3}{2}; \tfrac{1}{2}\beta)\right], \qquad (31.59)$$

where ${}_1F_1 \equiv M$, and

$$a = \frac{t_0^2}{\nu + t_0^2}, \quad \beta = \delta^2(1 - a), \qquad \sqrt{\beta} > 0.$$

Bulgren and Amos (1968) gave some other series representations and also a table of values of $\Pr[t''_\nu(\delta, \lambda) \le t_0]$ to six decimal places for $t_0 = 1, 2$, and $\nu = 2, 5(5)20$, $\delta = -4(2)4$, $\lambda = 0(4)8$.

The cumulative distribution function of $t''_\nu(\delta, \lambda)$ is given by

$$F(t; \nu; \delta, \lambda) = \Pr[t''_\nu(\delta, \lambda) \le t] = \Pr\left[U + \delta \le t\left(\frac{W}{\nu}\right)^{1/2}\right], \qquad (31.60)$$

where U and W are independent unit normal and $\chi_{\nu}'^{2}(\lambda)$ variables, respectively. Hence

$$F(t;\nu;\delta,\lambda) = \sum_{j=0}^{\infty}\left[e^{-\lambda/2}\frac{(\frac{1}{2}\lambda)^j}{j!}\right]F\left(t\left(1 + 2\nu^{-1}j\right)^{1/2};\nu + 2j,\delta\right), \quad (31.61)$$

where $F(t;\nu,\delta)$ is the cdf of a $t_{\nu}'(\delta)$ variable. [Kocherlakota and Kocherlakota (1991).]

Availability of computational algorithms for $t_{\nu}'(\delta)$ makes it simpler to determine $F(t;\nu;$6, A). [The routine DTNDF was used in IMSL (1987).] Krishnan (1968) and Kocherlakota and Kocherlakota (1991) observed that since

$$1 - F(t;\nu;\delta,\lambda) = F(-t;\nu;-\delta,\lambda), \quad (31.62)$$

it follows that $t''_{\nu,\alpha}(\delta,$ A) = $-t''_{\nu,1-\alpha}(-\delta,$ A). As a result tables for negative 6 are not needed. Note also that the median of $t_{\nu}''(-$6, A) is the same as *minus* of $t_{\nu}''(\delta,$ A).

Kocherlakota and Kocherlakota (1991) also provided tables of values of $t''_{\nu,\alpha}(\delta, A)$ for $\alpha = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9$, and 0.95 with $\nu = 5(5)20$, 6 = $0, 2, 4$, A = $0, 4, 8$. These authors used (31.61) combined with (31.60) for their calculations, obtaining results agreeing with Krishnan (1968) and Bulgren and Amos (1968). [Carey (1983) described another algorithm for the evaluation of $F(t;\nu;$6, A).]

8.3 Modified Noncentral t-Distribution

The most common modified t-statistic is obtained if we replace the S in the denominator (see Section 3) by the sample range W or by the mean of a number of independent sample ranges [see Lord (1947,1950) and Chapter 13]. Thus in place of the χ_{ν} in the denominator of $t_{\nu}'(\delta)$ (see Section 1) there is a variable having some other distribution but still independent of the U variable in the numerator.

The noncentrality of the distribution is associated with the 6 in the numerator. The denominator (in both original and modified forms) has the same distribution for both the central and noncentral cases. So approximations to this latter distribution that have been used for the central case can also be applied to the noncentral case, with reasonable hope of obtaining useful results. For example, if the distribution of W is approximated by that of $\chi_{\nu'}(c'\nu'^{1/2})^{-1}$, then $(U+\delta)/W$ may be approximated by $c't_{\nu'}'(\delta)$. Discussion of approximations to the distribution of noncentral modified t are found in Lord (1950) and Zaludová (1960).

8.4 Distribution of Noncentral t-Statistic When the Population is Nonnormal

Noncentral t-statistic distributions have been studied as early as in the thirties and forties in order to assess the effect of nonnormality on the power of the t-test. We note, in particular, the work of Ghurye (1949) and Srivastava (1958). The first of these authors extended some results of Geary (1936), 1947) who supposed that the population density function could be adequately represented as

$$p_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\left\{\exp\left[-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2\right]\right\}\left[1 + \frac{\sqrt{\beta_1}}{6}\left\{\left(\frac{x-\xi}{\sigma}\right)^3 - 3\left(\frac{x-\xi}{\sigma}\right)\right\}\right] \tag{31.63}$$

Srivastava, utilizing later results of Gayen (1949), obtained formulas for the case when the population density function is as in (31.63), with additional terms

$$\frac{1}{\sqrt{2\pi}\,\sigma}\left\{\exp\left[-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2\right]\right\}$$
$$\times\left[\frac{\beta-3}{24}\left\{\left(\frac{x-\xi}{\sigma}\right)^4 - 6\left(\frac{x-\xi}{\sigma}\right)^2 + 3\right\}\right.$$
$$\left. + \frac{\beta_1}{72}\left\{\left(\frac{x-\xi}{\sigma}\right)^6 - 15\left(\frac{x-\xi}{\sigma}\right)^4 + 45\left(\frac{x-\xi}{\sigma}\right)^2 - 15\right\}\right]$$

(i.e., the next terms in the Edgeworth expansion). The correction to the normal theory power is of the form

$$-\sqrt{\beta_1}\,P_{\sqrt{\beta_1}} - (\beta_2 - 3)P_{\beta_2} - \beta_1 P_{\beta_1}, \tag{31.64}$$

where the P's do not depend on the β's but on the noncentrality, the degrees of freedom, and the significance level of the test.

Bowman, Lam, and Shenton (1986) have studied the even moments and approximations of

$$T = \sqrt{n}\,\bar{X}\left\{(n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right\}^{-1/2}, \tag{31.65}$$

where the X_i's are independent random variables, each having the standard exponential distribution, with the pdf

$$p_X(x) = \exp(-x).$$

[Since $E[X_i] = 1 \neq 0$, (31.65) would have a noncentral r-distribution, $t'_{n-1}(\sqrt{n})$, if the common distribution of the X's were normal with expected value 1.]

Mulholland (1977) suggested studying the distribution of

$$W_n = \sum_{i=1}^{n} X_i^2 \Bigg/ \left(\sum_{i=1}^{n} X_i \right)^2 = n^{-1} + (n-1)n^{-1}T^{-2}, \qquad (31.66)$$

utilizing the recurrence relation

$$p_{W_n}(w) = (n-1)\int_1^{\infty} y^{-(n-2)} p_{W_{n-1}}\left(wy^2 - (y-1)^2\right) dy. \qquad (31.67)$$

Bowman, Lam, and Shenton obtained the following formulas for the pdfs of W_2, W_3, and W_4:

$$p_{W_2}(w) = (2w-1)^{-1/2}, \qquad \tfrac{1}{2} \leq w \leq 1, \qquad (31.68a)$$

$$p_{W_3}(w) = \begin{cases} \dfrac{2\pi}{\sqrt{3}}, & \tfrac{1}{3} \leq w \leq \tfrac{1}{2} \\ \dfrac{2\pi}{\sqrt{3}} - 2\sqrt{3}\cos^{-1}(6w-2)^{-1/2}, & \tfrac{1}{2} \leq w \leq 1, \end{cases} \qquad (31.68b)$$

$$p_{W_4}(w) = \begin{cases} 3\pi(4w-1)^{-1/2}, & \tfrac{1}{4} \leq w \leq \tfrac{1}{3} \\ 2\sqrt{3}\pi - 3\pi(4w-1)^{-1/2}, & \tfrac{1}{3} \leq w \leq \tfrac{1}{2} \\ 2\sqrt{3}\pi - 3\pi(4w-1)^{-1/2} - 6\sqrt{3}\tan^{-1}(6w-3)^{1/2} & \\ \quad + 18(4w-1)[(2w-1)/(4w-1)]^{1/2}, & \tfrac{1}{2} \leq w \leq 1. \end{cases} \qquad (31.68c)$$

REFERENCES

Airey, J. R. (1931). Table of Hh functions, In *British Association Mathematical Tables.* vol. 1, British Association. London. Table XV, pp. 60–72. (2nd ed., 1946; 3rd ed., 1951.)

Amos, D. E. (1964). Representations of the central and noncentral *t* distributions, *Biometrika. 51. 451–458.*

Anscombe, F. J. (1950). Table of the hyperbolic transformation $\sinh^{-1}\sqrt{x}$, Journal of the Royal Statistical Society, Series A, 113, 228–229.

Azorín, P. F. (1953). Sobre la distribución t no central. I, II, Trabajos de Estadistica, 4, 173–198 and 307–337.

Bagui, S. C. (1993). CRC Handbook of Percentiles of Noncentral t-Distributions, Boca Raton, FL: CRC Press.

Bargmann, R. E., and Ghosh, S. P. (1964). Noncentral statistical distribution programs for a computer language, IBM Research Report, RC-1231.

Bartlett, M. S. (1935). The effect of non-normality on the t-distribution, Proceedings of the Cambridge Philosophical Society, **31,** 223–231.

Belobragina, L. S., and Eliseyev, V. K. (1967). Statistical estimation of a recognition error based on experimental data, Kibemetika, 4, 81–89. (In Russian).

Bowman, K. O., Lam, H. K., and Shenton, L. R. (1986). Series for Student's non-central t under exponential sampling with comments due to H. P. Mulholland, Communications in Statistics—Simulation and Computation, 15, 697–708.

Bruscantini, S. (1968). Origin, features and use of the pseudonormal distribution, *Statistica* (Bologna), 28, 102–124.

Bulgren, W. G., and Amos, D. E. (1968). A note on representations of the doubly noncentral t distribution, *Journal* of the American Statistical Association, 63, 1013–1019.

Carey, M. B. (1983). Evaluation of the doubly non-central t cumulative distribution function, Computer Science and Statistics, Proceedings of 15th Symposium on the Interface, pp. 339–343.

Chattamvelli, R., and Shanmugam, R. (1994). An enhanced algorithm for noncentral t-distribution, Journal of Statistical *Computation* and Simulation, 49, 77–83.

Chowdhury, J. V., and Stedinger, J. R. (1991). Confidence interval for design floods with estimated skew coefficients, Journal of Hydraulic Engineering, **117,** 811–831.

Cooper, B. E. (1968). Algorithm AS-5. The normal integral; the integral of Student's t-distribution; an auxiliary function for distribution integrals; the integral of the non-central t-distribution, Applied Statistics, 17, 186–194.

Craig, C. C. (1941). Note on the distribution of noncentral t with an application, Annals of Mathematical Statistics, 12, 224–228.

Croarkin, M. C. (1962). Graphs for determining the power of Student's t-test, Journal of Research of the National Bureau of Standards, **66B**, 59–70, [Correction: In Mathematics of Computation, (1963), **17,** 83 (334).]

Dasgupta, S., and Lahiri, K. (1992). A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data, Journal of Business and Economic Statistics, 10, 391–400.

David, H. A., and Paulson, A. S. (1965). The performance of several tests for outliers, Biometrika, 52, 429–436.

Davies, O. L. (1954). The Design and Analysis of Industrial Experiments, New York: Hafner.

Deutler, T. (1984). A series expansion for the cumulants of the χ-distribution and a Cornish-Fisher expansion for the noncentrality parameter of the noncentral t-distribution, Communications in Statistics—Simulation and Computation, **13, 507–513.**

Eeden, C. van **(1958).** Some approximations to the percentage points of the noncentral t-distribution, Report *S242*, Statistics Department, Mathematics Center, Amsterdam.

Eeden, C. van **(1961).** Ibid., *Revue de l'Institut* Internationale de *Statistique*, ***29,*** **4–31.**

Fisher, *R.* A. (1931). Introduction to "Table of Hh Functions," pp. xxvi–xxxv. In Airey **(1931).**

Gayen, A. K. **(1949).** The distribution of "Student's" *t* in random samples of any size drawn from non-normal universes, Biometrika, **36, 353–369.**

Geary, R. C. **(1936).** The distribution of "Student's" ratio for non-normal samples, Journal of the Royal Statistical Society, Series B, **3, 178–184.**

Geary, R. C. **(1947).** Testing for normality, Biometrika, **34, 209–242.**

Ghurye, S. G. **(1949).** On the use of Student's t-test in an asymmetrical population, Biometrika, **36, 426–430.**

Griffiths, P., and Hill, I. D. (eds). (1985). Applied Statistics Algorithms, Chichester: Ellis Horwood.

Guenther, W. C. **(1975).** Evaluation of noncentral distribution integrals with a desk calculator, Research Paper No. 80 *S-1975-538*, College of Commerce and Industry, University of Wyoming, Laramie.

Guirguis, G. H. **(1990).** A note on computing the noncentrality parameter of the noncentral F distribution, Communications in Statistics—Simulation and Computation, ***19,*** **1497–1511.**

Gupta, A. K., and Kabe, D. G. (1992). On the derivation of a certain noncentral t-distribution, Journal of the Korean Statistical Society, 19, 182–185.

Hall, I. J., and Sampson, C. B. (1973). Tolerance limits for the distribution of the product and quotient of normal variates, Biometries, ***29,*** **109–119.**

Halperin, **M. (1963).** Approximations to the noncentral *t*, with applications, *Technometrics*, 5, 295–305.

Harley, B. I. (1957). Relation between the distributions of noncentral t and a transformed correlation coefficient, Biometrika, **44, 219–224.**

Hawkins, D. M. (1975). From the noncentral *t* to the normal integral, The American Statistician, **29(1), 42–43.**

Helms, R. **W. (1992).** Intentionally incomplete longitudinal designs, Statistics in Medicine, **11, 1889–1913.**

Hodges, J. L., and Lehmann, E. L. (1965). Moments of chi and power of t, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and *Probability*, 1, **187–201.**

Hodges, J. L., and Lehmann, E. L. **(1968).** A compact table for power of the t-test, Annals of Mathematical Statistics, 39, **1629–1637.**

Hogben, D., Pinkham, R. S., and Wilk, M. **B. (1961).** The moments of the non-central t distribution, Biometrika, **48, 465–468.**

Hogben, D., Pinkham, R. S., and Wilk, M. B. (1964). An approximation to the distribution of q (a variate related to the noncentral *t*), Annals of Mathematical Statistics. **35. 315–318.**

Ifram, A. F. (1970). On mixture of distributions with applications to estimation, Journal of the American Statistical Association, 65, 749–754.

Iglewicz, B., Myers, R. H., and Howe, R. B. (1968). On the percentage points of the sample coefficient of variation, Biometrika, 55, 580–581.

IMSL Statistics Library (1987), Version 1.0, vol. 3, 942–943, Houston, TX.

Jennett, W. J., and Welch, B. L. (1939). The control of proportion defective as judged by a single quality characteristic varying on a continuous scale, Journal of the Royal Statistical Society, Series B, 6, 80–88.

Jilek, M., and Likar, O. (1959). Coefficients for the determination of one-sided tolerance limits of normal distributions, Annals of the Institute of Statistical Mathematics, 11, 45–48.

Johnson, N. L., and Welch, B. L. (1940). Applications of the noncentral *t* distribution, Biometrika, 31, 362–389.

Kocherlakota, K., and Kocherlakota, S. (1991). On the doubly noncentral *t* distribution, Communications in Statistics—Simulation and Computation, 20, 23–31.

Kraemer, H. C. (1978). A central t-approximation to the noncentral t-distribution, Technical Series No. 107, Laboratory of Stress and Conflict, Dept. of Psychiatry and *Behavioral* Sciences, Stanford University, Stanford, CA.

Krishnan, M. (1967). The moments of a doubly noncentral t-distribution, Journal of the American Statistical Association, 62, 278–287.

Krishnan, M. (1968). Series representations of the doubly noncentral t-distribution, Journal of the American Statistical Association, 63, 1004–1012.

Kruskal, W. H. (1954). The monotonicity of the ratio of two noncentral t density functions, Annals of Mathematical Statistics, 25, 162–164.

Kiihlmeyer, M. (1970). Die *Nichtzentrale t-Verteilung*, Lecture Notes in Operations Research and Mathematical Systems, New York: Springer-Verlag.

Lahiri, K., and Teigland, C. (1987). On the normality of probability distributions of inflation and GNP forecasts, International Journal of Forecasting, 3, 269–279.

Laubscher, N. F. (1960). Normalizing the noncentral t and F distributions, Annals of Mathematical Statistics, 31, 1105–1112.

Lenth, R. V. (1989). Cumulative distribution function of the non-central t distribution, Algorithm AS 243, Applied Statistics, 38, 185–189.

Locks, M. O., Alexander, M. J., and Byars, B. J. (1963). New tables of the noncentral t-distribution, Report *AR63-19*, Wright-Patterson Air Force Base, Dayton, OH.

Lord, E. (1947). The use of range in place of the standard deviation in the t-test, *Bometrika*, 34, 41–67, (Correction: 39, 442.)

Lord, E. (1950). Power of modified t-test (u-test) based on range, Biometrika, 37, 64–77.

Majumder, K. L., and Bhattacharjee, G. P. (1973). Algorithm AS63. The incomplete beta integral, Applied Statistics, 22, 409–411.

Malcolm, S. (1984). A note on the use of the noncentral t-distribution in setting numerical microbiological specifications for foods, Journal of Applied Bacteriology, 57, 175–177.

Masuyama, M. (1951). An approximation to the non-central t-distribution with the stochastic paper, Reports of Statistical Applied Research, JUSE, 1(3), 28–31.

McKay, A. T. (1932). Distribution of the coefficient of variation and the extended t distribution, Journal of the Royal Statistical Society, **95, 695–698.**

Merrington, M., and Pearson, E. S. (1958). An approximation to the distribution of noncentral t, Biometrika, **45, 484–491.**

Miller, R. W. (1989). Parametric empirical Bayes tolerance intervals, Technometrics, **31, 449–459.**

Mulholland, H. P. (1977). Private communication [see Bowman, Lam, and Shenton (1986)].

Narula, S. C., and Weistroffer, H. R. (1986). Computation of probability and noncentrality parameter of noncentral F distribution, Communications in Statistics—Simulation and Computation, **15**, 871–878.

Neyman, J., Iwaszkiewicz, K., and Kolodziejczyk, S. (1935). Statistical problems in agricultural experimentation, Journal of the Royal Statistical Society, Series B, 2, **107–180.**

Neyman, J., and Tokarska, B. (1936). Errors of the second kind in testing "Student's" hypothesis, Journal of the American Statistical Association, 31, **318–326.**

Owen, D. B. (1963). Factors for one-sided tolerance limits and for variables sampling plans, Sandia Corporation Monograph SCR-607, Albuquerque, New Mexico.

Owen, D. B. (1965a). The power of Student's t-test, Journal of the American Statistical Association, 60, **320–333.**

Owen, D. B. (1965b). A special case of bivariate non-central t-distribution, Biometrika, **52, 437–446.**

Owen, D. B. (1968). A survey of properties and applications of the noncentral t-distribution, Technometrics, 10, **445–478.**

Owen, D. B., and Amos, D. E. (1963). Programs for computing percentage points of the noncentral t-distributions, Sandia Corporation Monograph SCR-551, Albuqurque, New Mexico.

Park, J. H. (1964). Variations of the noncentral t- and beta-distributions, Annals of Mathematical Statistics, **35, 1583–1593.**

Patnaik, P. B. (1955). Hypotheses concerning the means of observations in normal samples, *Sankhyā*, **15**, 343–372.

Pearson, E. S. (1958). Note on Mr. Srivastava's paper on the power function of Student's test, Biometrika, **45, 429–430.**

Pearson, E. S. (1963). Some problems arising in approximating to probability distributions, using moments, Biometrika, 50, 95–111. (Appendix: p. 112.)

Pearson, E. S., and Adyanthàya, N. K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations, Biometrika, **21, 259–286.**

Pearson, E. S., and Hartley, H. O. (1954). Biometrika Tables for Statisticians, vol. 1, Cambridge: Cambridge University Press. (2nd ed., 1958; 3rd ed., 1966.)

Phillips, K. F. (1993). A log-normal model for individual bioequivalence, Journal of *Biopharmaceutical* Statistics, **3**, 185–201.

Posten, H. O. (1986). Algorithms for the beta distribution function, In Proceedings in Computational Statistics (eds., F. de Antoni, N. Lauro, and A. Rizzi), Vienna: Physica, pp. 309–319.

Posten, H. O. (1993). A new algorithm for the noncentral t distribution function, University of Connecticut, Department of Statistics, Technical Report No. 93-15.

Resnikoff, S. J., and Lieberman, G. J. (1957). Tables of the Non-Central t-Distribution, Stanford: Stanford University Press.

Rukhin, A. L. (1992). Estimating the noncentrality parameter of a t-distribution, Systems Science and Mathematical Sciences, 5, 1–8.

Scheuer, E. M., and Spurgeon, R. A. (1963). Some percentage points of the non-central t-distribution, Journal *of* the American Statistical Association, 58, 176–182.

Singh, K. P., Relyea, G. E., and Bartolucci, A. A. (1992). On the tail probabilities of the noncentral t-distribution, Computational Statistics, 7, 67–80.

Srivastava, A. B. L. (1958). Effect of non-normality on the power function of t-test, Biometrika, 45, 421–429.

Stedinger, J. R. (1983a). Confidence intervals for design events, Journal of Hydraulic Engineering, ASCE, 109, 13–27.

Stedinger, J. R. (1983b). Design events with specified flood risk, Water Resources Research, 19, 511–522.

Steffens, F. E. (1968). Probability integrals of doubly noncentral F- and t-distributions with regression applications, Research Report No. 267, Council for Scientific and Industrial Research, Pretoria, South Africa.

Voit, E. O., and Rust, P. F. (1990). Evaluation of the noncentral t-distribution in S-systems, Biometrical Journal, 32, 681–695.

Wolfowitz, J. (1946). Confidence limits for the fraction of a normal population which lies between two given limits, Annals *of* Mathematical Statistics, 17, 483–488.

Zaludová, A. H. (1960). The noncentral I-test (q-test) based on range in place of standard deviation, *Acta Technica*, 5, 143–185.

Zhou, Y. (1987). Fiducial and Bayes models for the design of structural reliability, Chinese Journal *of* Mechanical Engineering, 23, 89–96. (In Chinese).

CHAPTER 32

Distributions of Correlation Coefficients

1 INTRODUCTION AND GENESIS

The statistic known as the sample (product-moment) correlation coefficient, based on n pairs of observed values of two characters, represented by random variables (X_t, Y_t) $(t = 1, \ldots, n)$ is

$$R = \frac{\sum_{t=1}^{n} (X_t - \bar{X})(Y_t - \bar{Y})}{\left[\sum_{t=1}^{n} (X_t - \bar{X})^2 \sum_{t=1}^{n} (Y_t - \bar{Y})^2\right]^{1/2}}, \tag{32.1}$$

where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$; $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$. There are many alternative expressions. One of the most useful is

$$R = \frac{\mathbf{X}^* \cdot \mathbf{Y}^*}{|\mathbf{X}^*|\,|\mathbf{Y}^*|} = \cos(\mathbf{X}^*, \mathbf{Y}^*) = \cos\theta, \tag{32.1$'$}$$

where $\mathbf{X}^* = (X_1^*, \cdots, X_n^*)$ and $\mathbf{Y}^* = (Y_1^*, \cdots, Y_n^*)$ with $X_t^* = X_t - \bar{X}$ and $Y_t^* = Y_t - \bar{Y}$, and $|\mathbf{X}^*|$, $|\mathbf{Y}^*|$ are the norms of $\mathbf{X}^*$ and $\mathbf{Y}^*$, respectively. Kass (1989) observes that

$$|R| = \frac{L_1 - L_2}{L_1 + L_2}, \tag{32.1$''$}$$

where L_1, L_2 are the eigenvalues of the Gram matrix

$$\mathbf{G} = \begin{pmatrix} |\mathbf{X}^*|^2 & \mathbf{X}^* \cdot \mathbf{Y}^* \\ \mathbf{X}^* \cdot \mathbf{Y}^* & |\mathbf{Y}^*|^2 \end{pmatrix},$$

and *provided* that $|\mathbf{X}^*| = |\mathbf{Y}^*|$,

$$G = \begin{pmatrix} 1 & \cos\theta \\ \cos\theta & 1 \end{pmatrix} |\mathbf{X}^*|.$$

(This shows that R is less suitable as a measure of association unless $|\mathbf{X}^*| = |\mathbf{Y}^*|$.) See Rodgers and Nicewander (1988), and also Section 3, for several other representations of R. The distribution that is the main topic of this chapter is the distribution of R when

1. (X_i, Y_i) and (X_j, Y_j) are mutually independent if $i \neq j$,
2. the joint distribution of X_t and Y_t has probability density function

$$p_{X_t,Y_t}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\xi}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\xi}{\sigma_X}\right)\left(\frac{y-\eta}{\sigma_Y}\right) + \left(\frac{y-\eta}{\sigma_Y}\right)^2\right\}\right] \qquad (32.2)$$

for each $t = 1, 2, \ldots, n$ $(\sigma_X > 0;\ \sigma_Y > 0;\ -1 < \rho < 1)$.

Formula (32.2) is that of the ***general bivariate normal distribution,*** which will be studied in a planned volume on ***Continuous Multiuariate Distributions.*** We discuss the distribution of R before the parent distribution (32.2) because R is univariate, while (32.2) is a multivariate distribution and so is appropriate to the volume on multivariate distributions. However, we will use a few of the properties of (32.2) in order to pursue our analysis of the distribution of R. The first is that p is the ***population correlation coefficient:***

$$\rho = \frac{E[\{X_t - E[X_t]\}\{Y_t - E[Y_t]\}]}{\sqrt{\mathrm{var}(X_t)\mathrm{var}(Y_t)}}. \qquad (32.3)$$

The second is that X_t, Y_t each have normal distributions, with $E[X_t] = \xi$; $E[Y_t] = \eta$; $\mathrm{var}(X_t) = \sigma_X^2$; $\mathrm{var}(Y_t) = \sigma_Y^2$. The conditional distribution of Y_t, given X_t, is normal, with expected value $\eta + (\rho\sigma_Y/\sigma_X)(X_t - \xi)$ and variance $(1-\rho^2)\sigma_Y^2$.

Rao (1983), while reminiscing on the origin and development of the correlation coefficient, has pointed out that the source of symbol R for the correlation coefficient was really the first letter of ***Reversion*** (according to Karl Peason). Galton himself referred to his measure R as the ***index of co-relation,*** and Weldon called it the ***Galton function.*** While Karl Pearson and Sheppard derived the large-sample standard error of R, Fisher (1915) derived the exact distribution of R, under normality. Fisher also soon found a simple transformation $R = \tanh Z'$, known as Fisher's ***z'-transformation***, to considerably simplify the sampling distribution of R as well as the inference procedures based on the observed value of R.

In his classic treatise on Natural Inheritance, Galton wrote in 1908 of his discovery with that inspiring phraseology so typical of him: "This part of the inquiry may be said to run along a road on a high level, that affords wide views in unexpected directions, and from which easy descents may be made to totally different goals...". It is important to mention here that the correlation coefficient may be computed for any bivariate distribution with measured variables, but it certainly will not be an appropriate measure of intensity of association if the regression is curved; indeed we do not know at all thoroughly how its value is to be interpreted for many purposes unless the distribution itself is normal. Neglect of this latter point is sometimes quite flagrant.

This chapter is primarily concerned with the distribution of R corresponding to (32.2), but the distributions arising under some other conditions will also be discussed in Section 3. Furthermore Sections 8 and 9 are devoted to distributions of serial correlations, and Section 11 to distribution of multiple correlation.

For a brief recent discussion of some aspects of the distribution of the sample correlation coefficient (such as the derivation of the exact distribution, transformations, approximations, moments, robustness, etc.), interested readers may refer to Stuart and Ord (1994, pp. 556–570).

2 DERIVATION OF DISTRIBUTION OF R

Since the correlation between the standardized variables $(X_i - \xi)/\sigma_X$ and $(Y_i - \eta)/\sigma_Y$ is the same as that between X_i and Y_i, no generality is lost by taking $\xi = \eta = 0$; $\sigma_X = \sigma_Y = 1$. We now consider the conditional distribution of R for fixed values of $X_1, X_2, \cdots, X_n$. Since the conditional distribution of Y_i, given X_i, is normal with expected value ρX_i and variance $(1 - \rho^2)$ (remembering that we are taking $\xi = \eta = 0$; $\sigma_X = \sigma_Y = 1$), it follows that $R(1 - R^2)^{-1/2}$ is distributed as $(n - 2)^{-1/2}$ times noncentral t with $(n - 2)$ degrees of freedom and noncentrality parameter

$$\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \cdot \frac{\rho}{\sqrt{(1 - \rho^2)}}.$$

(See Chapter 31, Section 6.)

To obtain the overall (unconditional) distribution of $R(1 - R^2)^{-1/2}$, we must calculate the expected value of the density function so obtained, over the distribution of $X_1, X_2, \cdots, X_n$. Since the density function depends on the X's only through the statistic $\sum_{i=1}^{n}(X_i - \bar{X})^2$, we need only use the fact that this statistic is distributed as χ^2 with $n - 1$ degrees of freedom, that is, as χ^2_{n-1} (Chapter 18, Section 1). Writing $V = R(1 - R^2)^{-1/2}$, the conditional probability density of V is

$$p_V(v|S) = \frac{\exp\left[\dfrac{-\rho^2 S}{2(1-\rho^2)}\right]}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}n - 1\right)}(1 + v^2)^{-(n-1)/2} \times \sum_{j=0}^{\infty} \frac{\Gamma((n-1+j)/2)}{j!}\left\{\frac{2\rho^2 v^2 S}{(1-\rho^2)(1+v^2)}\right\}^{j/2} \qquad (32.4)$$

where $S = \sum_{i=1}^{n}(X_i - \bar{X})^2$.

Since

$$p_S(s) = \left[2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)\right]^{-1} s^{(n-3)/2} e^{-s/2}, \qquad s > 0,$$

and

$$\int_0^{\infty} s^{j/2} \exp\left[-\frac{1}{2}\rho^2 s(1-\rho^2)^{-1}\right] p_S(s)\, ds = \frac{2^{j/2}\Gamma((n-1+j)/2)}{\Gamma((n-1)/2)}(1-\rho^2)^{(n+j-1)/2},$$

it follows that

$$p_V(v) = \frac{(1-\rho^2)^{(n-1)/2}}{\sqrt{\pi}\,\Gamma((n-1)/2)\Gamma\left(\frac{1}{2}n - 1\right)}(1+v^2)^{-(n-1)/2} \times \sum_{j=0}^{\infty} \frac{(2\rho)^j[\Gamma((n-1+j)/2)]^2}{j!}\left(\frac{v^2}{1+v^2}\right)^{j/2}. \qquad (32.5)$$

Finally, making the transformation $V = R(1 - R^2)^{-1/2}$, we obtain

$$p_R(r) = \frac{(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}(n-1)\right)\Gamma\left(\frac{1}{2}n - 1\right)} \times \sum_{j=0}^{\infty} \frac{\left[\Gamma\left(\frac{1}{2}(n-1+j)\right)\right]^2}{j!}(2\rho r)^j, \qquad -1 \le r \le 1. \quad (32.6a)$$

[The constant multiplier may be expressed in an alternative form by using the identity $\sqrt{\pi}\,\Gamma(\tfrac{1}{2}(n-1))\Gamma(\tfrac{1}{2}n-1) = 2^{-(n-3)}\pi(n-3)!$]

There are a number of other forms in which the right-hand side of (32.6a) may be expressed. These include

$$p_R(r) = \frac{(n-2)(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\pi}\int_0^\infty \frac{dw}{(\cosh w - \rho r)^{n-1}}, \tag{32.6b}$$

$$p_R(r) = \frac{(n-2)(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\pi}\int_1^\infty \frac{dw}{(w-\rho r)^{n-1}(w^2-1)^{1/2}}, \tag{32.6c}$$

$$p_R(r) = \frac{(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\pi(n-3)!}\,\frac{d^{n-2}}{d(\rho r)^{n-2}}\left\{\frac{\cos^{-1}(-\rho r)}{(1-\rho^2 r^2)^{1/2}}\right\}, \tag{32.6d}$$

$$p_R(r) = \frac{(n-2)(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\sqrt{2}(n-1)B(\tfrac{1}{2}, n-\tfrac{1}{2})(1-\rho r)^{n-(3/2)}} \times {}_2F_1\left(\frac{1}{2},\frac{1}{2};n-\frac{1}{2};\frac{1}{2}(1+\rho r)\right) \quad \text{[Hotelling (1953)]}, \tag{32.6e}$$

$$p_R(r) = \frac{(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}}{\pi(n-3)!}\left(\frac{\partial}{\partial\theta}\right)^{n-2}\frac{\theta}{\sin\theta} \tag{32.6f}$$

with $0 = \cos^{-1}(\rho r)$ [Fisher (1915)].

In all cases $-1 \le r \le 1$. [${}_2F_1(\ldots)$ is the Gauss hypergeometric function defined in Chapter 1, Eq. (1.104).]

Formulas (32.6b) and (32.6c) are obtainable, each from the other, by simple transformation of the variable in the integral. Equation (32.6e) is a direct consequence of (32.6a) and, even for moderately large n, the hypergeometric series converges rapidly. Formulas (32.6d) and (32.6f) are notable in that they express the probability density by a finite number of terms which involve elementary functions only.

Fisher (1915) obtained the distribution of R in the form (32.6f) motivated by a geometrical argument. [Earlier investigations were made by Student (1908) and Soper (1913).] This distribution also served Fisher as the initial model for introducing the "fiducial method of inference" (see Chapters 1 and 13, and Chapter 28, Section 7) and has been, in this connection, the subject of numerous discussions in the literature [e.g., Fraser (1963); Williams (1993)].

There are several elementary derivations of $p_R(r)$ for the case p = 0. Here we note a purely geometric derivation by Chance (1984), which is valid for any spherically symmetric distribution and was inspired by Fisher's (1915) derivations. [Fisher treated each of X and Y as a point on an n-dimensional sphere and noted that the correlation coefficient corresponds to the cosine of the angle (θ) between the radii to the two points. For the remainder of his paper, however, he relied upon an analytical treatment of the bivariate normal distribution to derive a general expression.] An elementary derivation, after a change of variables, was also given by Srivastava and Khatri (1979).

For p = 0 we have the so-called null pdf of R:

$$p_R(r) = \frac{\Gamma[(n-1)/2]}{\Gamma(\frac{1}{2})\Gamma[(n-2)/2]}(1-r^2)^{(n-4)/2} \quad \text{for } -1 < r < 1. \tag{32.7}$$

(The distribution is symmetric around 0.) The corresponding **moment-generating** function is

$$M_R(t) = \Gamma(\tfrac{1}{2}(n-1))2^{(n-3)/2}t^{-(n-3)/2}I_{(n-3)/2}(t) \qquad \text{for } n > 2, \tag{32.8a}$$

where

$$I_{(n-3)/2}(t) = \frac{(\frac{1}{2}t)^{(n-3)/2}}{\Gamma[\frac{1}{2}(n-1)]}\left[1 + \frac{t^2}{2(n-1)} + \frac{t^4}{2^3(n-1)(n+1)} + \cdots\right]$$

is a modified Bessel function of the second kind of order $(n-3)/2$. The corresponding (real-valued) characteristic function, expressed in terms of the Bessel function $J_{(n-3)/2}(t)$, where $I_\nu(z) = i^{-\nu}J_\nu(iz)$, is

$$\varphi_R(t) = \Gamma(\tfrac{1}{2}(n-1))2^{(n-3)/2}t^{-(n-3)/2}J_{(n-3)/2}(t) \qquad \text{[Bhatti (1990)]}. \tag{32.8b}$$

For small values of n, simple explicit formulas for the cumulative distribution (cdf) of R were obtained by Garwood (1933). A few of these are shown in Table 32.1, with y_n denoting the pdf of R [$p_R(r;\rho,n)$], and

$$Q(r\rho) = (1-r^2\rho^2)^{-1/2}\cos^{-1}(-r\rho). \tag{32.9}$$

The values of y, and y_4 are

$$y_3 = \pi^{-1}(1-r^2)^{-1/2}(1-\rho^2)(1-r^2\rho^2)^{-1}\{1 + r\rho Q(r\rho)\},$$

$$y_4 = \pi^{-1}(1-\rho^2)^{3/2}(1-r^2\rho^2)^{-2}\{3r\rho + (1+2r^2\rho^2)Q(r\rho)\}.$$

Table 32.1 Formulas for the cdf of R

n	$F_R(r;\rho,n)$
3	$\pi^{-1}[\cos^{-1}(-r) - \pi^{-1}\rho(1-r^2)Q(r\rho)]$
4	$\rho^{-1}(1-\rho^2)^{1/2}(1-r^2)^{1/2}y_3 - (\pi\rho)^{-1}(1-\rho^2)^{1/2} + \pi^{-1}\cos^{-1}\rho,$
5	$(2\rho)^{-1}(1-\rho^2)^{1/2}(1-r^2)^{1/2}y_4 - \frac{1}{2}r(1-r^2)y_3$ $-(2\pi\rho)^{-1}(1+\rho^2)(1-r^2)^{1/2}Q(r\rho) + \pi^{-1}\cos^{-1}(-r)$
6	$(3\pi\rho^3)^{-1}(1-\rho^2)^{1/2}(1-4\rho^2) + \pi^{-1}\cos^{-1}\rho$ $-(3\rho^3)^{-1}(1-\rho^2)^{3/2}(1-r^2)^{1/2}y_3$ $+r(3\rho^2)^{-1}(1-\rho^2)y_4 + (3\rho)^{-1}(1-\rho^2)^{1/2}(1-r^2)^{1/2}y_5$
7	$(4\rho)^{-1}(1-r^2)^{1/2}(1-\rho^2)^{1/2}y_6 + r(4\rho^2)^{-1}(1-\rho^2)y_5$ $-r^2(8\rho)^{-1}(1-r^2)^{1/2}(1-\rho^2)^{1/2}(2-\rho^2)y_4$ $-r(8\rho^2)^{-1}(1-r^2)(4-3\rho^2+3\rho^4)y_3$ $-(8\pi\rho)^{-1}(1-r^2)^{1/2}(3+6\rho^2-\rho^4)Q(r\rho) + \pi^{-1}\cos^{-1}(-r)$

Note. The authors are grateful to Drs. O. Öksoy and L. A. Aroian for pointing out a misprint in these formulas in the first edition of this book. See Oksoy and Aroian (1982).

Values of y_n for $n > 4$ can be calculated from the recurrence formula

$$y_n = (1 - r^2\rho^2)^{-1}\{(1+r^2)(1-\rho^2)\}^{1/2}$$
$$\times [(n-3)^{-1}(2n-5)r\rho y_{n-1}$$
$$[illegible] - 2)^{-1}(n-1)\{(1-r^2)(1-\rho^2)\}^{1/2} y_{n-2}]$$

[Soper et al. (1917)]. (3.10)

Garwood (1933) also obtained the general formula, for odd n ($= 2s+3$),

$$F_R(r;\rho,2s+3) = \pi^{-1}\cos^{-1} r - (1-r^2)^{1/2}[(2s!)]^{-1}\pi^{-1}(1-\rho^2)^{s+1}$$
$$\times\left\{\Delta^s\rho^{2s} - \binom{s}{1}\frac{\partial^2}{\partial\rho^2}\Delta^{s-1}\rho^{2s-2}\right.$$
$$+\binom{s}{2}\frac{\partial^4}{\partial\rho^4}\Delta^{s-2}\rho^{2s-4}$$
$$\left. + \cdots + (-1)^s\frac{\partial^{2s}}{\partial\rho^{2s}}\right\}\frac{\rho}{1-\rho^2}Q(r\rho). \qquad (32.11a)$$

Greco (1992) suggest that evaluation can be facilitated by using the formulas

$$F_R(r;\rho,n)$$

$$= \begin{cases} \pi^{-1}\left\{\cos^{-1}(-r) + (1+\rho^2)^{1/2} \sum_{i=1}^{(n-3)/2} L_{2i-1} - \rho(1-r^2)^{1/2} \sum_{i=0}^{(n-3)/2} L_{2i}\right\} & \text{for } n \text{ odd,} \\ \pi^{-1}\left\{\cos^{-1}\rho + (1-\rho^2)^{1/2} r \sum_{i=0}^{(n-4)/2} L_{2i} - \rho(1-r^2)^{1/2} \sum_{i=0}^{(n-2)/2} L_{2i-1}\right\} & \text{for } n \text{ even,} \end{cases} \tag{32.11b}$$

where $L_k = \{(1-\rho^2)(1-r^2)\}^{1/2} d^k Q(r\rho)/d(r\rho)^k$, which may be evaluated from the recurrence relation

$$L_k = (1-r^2\rho^2)^{-1}(1-\rho^2)(1-r^2)$$
$$\times\left[(2-k^{-1})\left\{(1-\rho^2)(1-r^2)^{-1/2} r\rho L_{k-1}\right\} + (1-k^{-1})L_{k-2}\right].$$

As n increases, the expressions rapidly become more complicated. However, despite the complexity of the formulas, the density function is represented by a simple curve over the range $-1 \le r \le 1$ with a single mode (antimode if $n < 4$).

We also note that the value of $F_R(0) = \Pr[R \le 0]$ can be evaluated rather simply. Since $R \le 0$ is equivalent to

$$\sum_{t=1}^{n} (X_t - \bar{X})Y_t \le 0,$$

we need to evaluate the probability of this latter event. For given $X_1, X_2, \cdots, X_n$ the probability is (using results already quoted in this section):

$$\Phi\left(\frac{-\rho}{\sqrt{1-\rho^2}} \sqrt{\sum_{j=1}^{n}\left(X_j-\bar{X}\right)^2}\right) = \Pr\left[\frac{U}{\sqrt{\sum_{j=1}^{n}\left(X_j-\bar{X}\right)^2}} \le -\frac{\rho}{\sqrt{1-\rho^2}}\right]$$

Averaging over the distribution of $\Sigma_{j=1}^{n}(X_j - \bar{X})^2$ (which is chi-square with n – 1 degrees of freedom), we see that

$$\Pr[R \le 0] = \Pr\left[\frac{t_{n-1}}{\sqrt{n-1}} \le -\frac{\rho}{\sqrt{1-\rho^2}}\right]$$

$$= \Pr\left[t_{n-1} \le \frac{-\rho\sqrt{n-1}}{\sqrt{1-\rho^2}}\right]. \tag{32.12}$$

[This result was noted by Armsen (1956) and Ruben (1963).]

The moments of the distribution of R can be expressed in terms of hypergeometric functions [Ghosh (1966)]:

$$\mu'_1 = c_n \rho\, {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}(n+1); \rho^2\right), \tag{32.13a}$$

$$\mu'_2 = 1 - \frac{(n-2)(1-\rho^2)}{n-1}\, {}_2F_1\left(1, 1; \frac{1}{2}(n+1); \rho^2\right), \tag{32.13b}$$

$$\mu'_3 = c_n\left[\rho\, {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}(n+1); \rho^2\right) - \rho^{-1}(n-1)(n-2)\right]$$
$$\times\left\{{}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}(n-1); \rho^2\right) - {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{1}{2}(n+1); \rho^2\right)\right\}, \tag{32.13c}$$

$$\mu'_4 = 1 + \frac{(n-2)(n-4)(1-\rho^2)}{2(n-1)}\, {}_2F_1\left(1, 1; \frac{1}{2}(n+1); \rho^2\right)$$
$$- \frac{n(n-2)(1-\rho^2)}{4\rho^2}\left[{}_2F_1\left(1, 1; \frac{1}{2}(n+1); \rho^2\right) - 1\right], \tag{32.13d}$$

where

$$c_n = \frac{2}{n-1}\left[\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\right]^2.$$

Ghosh (1966) also obtained the following expansions in inverse powers of m ($= n + 6$):

$$\mu'_1 = \rho - \frac{1}{2}\rho(1-\rho^2)m^{-1}$$
$$\times\left[1 + \frac{9}{4}(3+\rho^2)m^{-1} + \frac{3}{8}(121 + 70\rho^2 + 25\rho^4)m^{-2}\right] + O(m^{-4}), \tag{32.14a}$$

$$\mu_2 = \frac{(1-\rho^2)^2}{m}\left[1 + \frac{1}{2}(14 + 11\rho^2)m^{-1} + \frac{1}{2}(98 + 130\rho^2 + 75\rho^4)m^{-2}\right]$$
$$+O(m^{-4}), \tag{32.14b}$$

$$\mu_3 = -\frac{\rho(1-\rho^2)^3}{m^2}$$
$$\times\left[6 + (69 + 88\rho^2)m^{-1} + \frac{3}{4}(797 + 1691\rho^2 + 1560\rho^4)m^{-2}\right]$$
$$+O(m^{-5}), \tag{32.14c}$$

$$\mu_4 = \frac{3(1-\rho^2)^4}{m^2}\left[1 + (12 + 35\rho^2)m^{-1} + \frac{1}{4}(436 + 2028\rho^2 + 3025\rho^4)m^{-2}\right]$$
$$+O(m^{-5}). \tag{32.14d}$$

Also

$$\beta_1 = \frac{\rho^2}{m}[36 + 6(12 + 77\rho^2)m^{-1} - (162 - 1137\rho^2 - 6844\rho^4)m^{-2}]$$
$$+O(m^{-4}) \tag{32.15a}$$

(note that the sign of $\sqrt{\beta_1}$ is opposite to that of ρ), and

$$\beta_2 = 3 - \frac{3}{m}\left[2(1 - 12\rho^2) + (10 + 14\rho^2 - 387\rho^4)m^{-1}\right.$$
$$\left. + \frac{1}{2}(100 + 832\rho^2 + 1503\rho^4 - 14202\rho^6)m^{-2}\right]$$
$$+ O(m^{-4}). \tag{32.15b}$$

Extending Ghosh's (1966) results, Iwase (1985) obtained the following explicit expressions (not involving integrals) for $F_R(r;\rho,n)$, valid for $n \geq 3$:

$$F_R(r;\rho,n) = \frac{1}{2} - \frac{\rho(1-\rho^2)^{(n-1)/2}}{B(\frac{1}{2},\frac{1}{2}(n-1))}\,{}_2F_1(1,\tfrac{1}{2}n;\tfrac{3}{2};\rho^2)$$
$$+\frac{(1-\rho^2)^{(n-1)/2}}{B(\frac{1}{2},\frac{1}{2}(n-2))}r\,F^{1:2:1}_{1:1:0}\left(\begin{matrix}\frac{1}{2};\frac{1}{2}(n-1),\frac{1}{2}(n-1);\dfrac{4-n}{2};\\ \frac{3}{2};\frac{1}{2};-;\end{matrix}\quad \rho^2r^2,r^2\right)$$
$$+\frac{(n-2)\rho(1-\rho^2)^{(n-1)/2}}{2B(\frac{1}{2},\frac{1}{2}(n-1))}r^2\,F^{1:2:1}_{1:1:0}\left(\begin{matrix}1;\frac{1}{2}n,\frac{1}{2}n;\dfrac{4-n}{2};\\ 2;\frac{3}{2};-;\end{matrix}\quad \rho^2r^2,r^2\right), \tag{32.16}$$

where

$$F^{A:B:B'}_{C:D:D'}\left(\begin{matrix}a_1,\dots,a_A;b_1,\dots,b_B;b'_1,\dots,b'_{B'};\\ c_1,\dots,c_C;d_1,\dots,d_D;d'_1,\dots,d'_{D'};\end{matrix}\quad x,y\right)$$
$$=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left\{\frac{\prod_{j=1}^{A}(a_j)_{m+n}\prod_{j=1}^{B}(b_j)_m\prod_{j=1}^{B'}(b'_j)_n}{\prod_{j=1}^{C}(c_j)_{m+n}\prod_{j=1}^{D}(d_j)_m\prod_{j=1}^{D'}(d'_j)_n}\right\}x^my^n,$$

with $(g)_h = g(g+1)\cdots(g+h-1)$, and ${}_2F_1(a,b;c;z)$ is the Gauss hypergeometric function defined in Chapter 1 [see Eq. (1.104)].

In particular,

$$\Pr[R\leq 0] = \frac{1}{2} - \frac{\rho(1-\rho^2)^{(n-1)/2}}{B(\frac{1}{2},\frac{1}{2}(n-1))}\,{}_2F_1\left(1,\frac{n}{2};\frac{3}{2};\rho^2\right), \qquad n\geq 3. \tag{32.17}$$

This is of course equivalent to (32.12).

Also the kth moment of R about zero is

$$\mu'_k = \begin{cases}\dfrac{B((k+1)/2,(n-2)/2)}{B(\frac{1}{2},(n-2)/2)}(1-\rho^2)^{(n-1)/2} \\ \quad\times{}_3F_2\left(\dfrac{k+1}{2},\dfrac{n-1}{2},\dfrac{n-1}{2};\dfrac{n+k-1}{2},\dfrac{1}{2};\rho^2\right), & k=0,2,4,\dots,\\ \dfrac{(n-2)B((k+2)/2,(n-2)/2)}{B(\frac{1}{2},(n-1)/2)}\rho(1-\rho^2)^{(n-1)/2} \\ \quad\times{}_3F_2\left(\dfrac{k+2}{2},\dfrac{n}{2},\dfrac{n}{2};\dfrac{n+k}{2},\dfrac{3}{2};\rho^2\right), & k=1,3,5,\dots,\ |\rho|<1,\ n\geq 3,\end{cases} \tag{32.18}$$

[compare with Eqs. (32.13)], and

$$E\left[\left(\frac{R^2}{1-R^2}\right)^k\right] = \frac{B\left(k+\frac{1}{2}, \frac{n}{2}-k-1\right)}{B\left(\frac{1}{2},(n-2)/2\right)}(1-\rho^2)^{(n-1)/2}$$
$$\times {}_2F_1\left(k+\frac{1}{2}, \frac{n-1}{2}; \frac{1}{2}; \rho^2\right), \qquad (32.19)$$

for any nonnegative real number k, provided that $n \geq 3$ and $|\rho| \neq 1$. [${}_3F_2(\cdot,\cdot,\cdot;\cdot,\cdot;\cdot)$ is a generalized hypergeometric function, as defined in Chapter 1, Eq. (1.140).]

For everyday use, remember that the bias in R as an estimator of ρ is approximately $-\frac{1}{2}\rho(1-\rho^2)n^{-1}$ and that

$$\operatorname{var}(R) \doteq (1-\rho^2)^2 n^{-1} \qquad \text{(see Section 5)}. \qquad (32.20)$$

It is interesting that

$$E[\sin^{-1} R] = \sin^{-1}\rho. \qquad (32.21)$$

[See Harley (1954, 1956); Daniels and Kendall (1958).] Subrahmaniam and Gajjar (1980) claim that this is the only nonconstant function $g(R)$ for which $E[g(R)] = g(\rho)$.

We conclude this section by giving two relationships [(32.22) and (32.23)] satisfied by $p_R(r)$. They are not generally of practical use but may be helpful in specific problems and have some intrinsic interest:

$$r\frac{\partial p_R(r)}{\partial r} + \frac{(n-3)r^2}{1-r^2}p_R(r) = \rho\frac{\partial p_R(r)}{\partial \rho} + \frac{n\rho^2}{1-\rho^2}p_R(r) \qquad (32.22)$$

[Hotelling (1953)]. [The near symmetry of the coefficients in r (left hand) and ρ (right hand) should be noted.] Also

$$(n-1)(n-2)\{1-(\rho R)^2\}p_R(r;\rho,n+1)$$
$$= (2n-1)(n-2)\rho\sqrt{1-\rho^2}\,R\sqrt{1-R^2}\,p_R(r;\rho,n)$$
$$+(n-1)^2(1-\rho^2)(1-r^2)p_R(r;\rho,n-1) \qquad (32.23)$$

[Soper et al. (1917)].

3 HISTORICAL REMARKS

Although our primary interest is in the distributions of R (and other measures of association) rather than in the history of the statistic R in (32.1), we provide for background a condensed account of its history. One of the most popular, and widely used (and abused), descriptive measures of degree of linear relationship between two variables, the product-moment (Pearson) correlation coefficient (32.1) was first described explicitly by Karl Pearson (1896, p. 265). He introduced the formula with the following words: "Thus it appears that the observed result is the most probable when r is given the value $S(xy)/(n\sigma_1\sigma_2)$. This value presents no practical difficulty in calculation, and therefore we shall adopt it. It is the value given by Bravais, but he does not show it is the best."

The formula Pearson described is the one given in modern usage as

$$R = \frac{\sum_{i=1}^{n} X_i^* Y_i^*}{n\sigma_X\sigma_Y} \qquad \text{[cf. (32.1)']}, \tag{32.24}$$

where $X_i^* = X_i - \bar{X}$ and $Y_i^* = Y_i - \bar{Y}$ as in (32.1)'.

There is evidence [see, e.g., Symonds (1926); Tankard (1984)] that K. Pearson's (1896) formula had been in manuscript some time before its publication. K. Pearson (1895) contains a brief reference to the formula, and so does Yule (1895) who was a student of Pearson.

The phrase "coefficient of correlation" in place of Galton's (1886) "index of co-relation" and Weldon's "Galton's function" was introduced by F. Y. Edgeworth in his Newmarch lectures at University College, London, in 1892 [see K. Pearson (1920); Stigler (1978)]. Karl Pearson was greatly influenced by Edgeworth (1892a, b) in his initial groundbreaking work [K. Pearson (1896)].

It was realized that properties of R (or r_{XY}) which are of most practical value can be summed up as follows:

1. $-1 \le R \le 1$.
2. $R = -1$ implies a line sloping downward to the right with all observations lying directly on the line (i.e., a perfect negative linear relation between the observed X's and Y's).
3. $R = 1$ implies a perfect positive ***linear*** relation in ***the*** sample data.
4. R near to zero implies little or no ***linear*** relationship (but there may be a nonlinear relation between the two).
5. If X, Y are independent, $\rho_{XY} = 0$, where ρ_{XY} is the ***population*** correlation coefficient between X and Y

$$\left[= \{E[XY] - E[X]E[Y]\}/\{\mathrm{var}(X)\mathrm{var}(Y)\}^{1/2}\right].$$

6. If X, Y are ***normal*** random variables, then $\rho_{XY} = 0$ implies that X and Y are independent. The same property is also valid if each one of X and Y is a binary random variable.
7. R and ρ_{XY} are invariant to location and scale transformations.

In early days there was considerable emphasis on construction of convenient formulas for computing R. For example, Symonds (1926) collected no fewer than 52(!) such formulas as examples. Among them were three versions involving the "gross scores" X and Y:

$$R = \frac{n^{-1}\Sigma XY - n^{-2}\Sigma X \Sigma Y}{\left\{n^{-1}\Sigma X^2 - (n^{-1}\Sigma X)^2\right\}^{1/2}\left\{n^{-1}\Sigma Y^2 - (n^{-1}\Sigma Y)^2\right\}^{1/2}}, \tag{32.25a}$$

$$R = \frac{\Sigma XY - n^{-1}\Sigma X \Sigma Y}{\left\{\Sigma X^2 - n^{-1}(\Sigma X)^2\right\}^{1/2}\left\{\Sigma Y^2 - n^{-1}(\Sigma Y)^2\right\}^{1/2}}, \tag{32.25b}$$

$$R = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\left\{n\Sigma X^2 - (\Sigma X)^2\right\}^{1/2}\left\{n\Sigma Y^2 - (\Sigma Y)^2\right\}^{1/2}}. \tag{32.25c}$$

These were evidently intended to appeal to calculators of various, slightly different tastes.

These formulas were first published in Harris (1910). They were rediscovered independently by Thurstone (1917) and Ayres (1920). Indeed, these formulas are sometimes referred to as "Ayres's formulas" in educational and psychological literature. As Symonds (1926) pointed out: "This is a good example of scientific men doing research without endeavoring to find out what has been done in the same field in the past." Hull (1925) and Dodd (1926) invented special "automatic correlation machines" to compute ΣX, ΣY, ΣX^2, ΣY^2, and ΣXY.

More general formulas, involving deviations $x' = X - \xi$, $y' = Y - \eta$ from "assumed means" or "arbitrary origin" (ξ, η) simply replace X, Y in (32.25a)–(32.25c) by x', y', respectively. These were essentially given even earlier by Yule (1897), albeit in the slightly disguised form

$$R = \frac{\Sigma x'y' - nc_X c_Y}{\left\{\left(\Sigma x'^2 - nc_X^2\right)\left(\Sigma y'^2 - nc_Y^2\right)\right\}^{1/2}}, \tag{32.26}$$

where c_X is the "distance between the assumed and the true mean." $[c_X = \xi - \bar{X};\ c_Y = \eta - \bar{Y}]$.

The "difference formula"

$$R = \frac{\Sigma x'^2 + \Sigma y'^2 - \Sigma(x' - y')^2}{2\sqrt{(\Sigma x'^2)\Sigma(y'^2)}} \tag{32.27a}$$

was given by K. Pearson (1896). It was independently rediscovered by Boas (1909). (Pearson immediately wrote an indignant rejoinder, criticizing Boas for his negligence in studying the literature.) A variant ("sum formula") of the difference formula,

$$R = \frac{\Sigma(x' + y')^2 - \Sigma x'^2 - \Sigma y'^2}{2\sqrt{(\Sigma x'^2)(\Sigma y'^2)}}, \tag{32.27b}$$

apparently appeared first in Kelley (1923). Yet another variant

$$R = \frac{\Sigma x'^2 + \Sigma y'^2 - \Sigma(X - Y)^2 + n(\overline{X} - \overline{Y})^2}{2\sqrt{(\Sigma x'^2)(\Sigma y'^2)}} \tag{32.28}$$

was devised by Huffaker (1925).

Considerable attention was devoted to the special equivariance case

$$\text{var}(X) = \text{var}(Y) = \sigma^2.$$

Harris (1910) proposed a number of formulas, including

$$R = 1 - \frac{\Sigma(x' - y')^2}{2\Sigma x'^2} \tag{32.29a}$$

for this case. Less familiar formulas, in Symonds (1926), include

$$R = \frac{\Sigma(x' + y')^2 - \Sigma(x' - y')^2}{\Sigma(x' + y')^2 + \Sigma(x' - y')^2}. \tag{32.29b}$$

Printed forms (also called "charts" and "data sheets") facilitating calculation of the correlation coefficient were offered for sale in the United States in the early 1920s.

Early work on the distribution of R was almost exclusively devoted to cases in which X and Y had joint bivariate normal distributions. In later years there has been an increased emphasis on situations where this is not the case, and consideration of robustness arises. This is the topic of the next section.

4 DISTRIBUTION OF *R* IN NONNORMAL POPULATIONS AND ROBUSTNESS

The distribution of R for samples from nonnormal populations has been worked out in detail only for certain special cases. For certain bivariate Edgeworth populations, investigations indicate the kinds of variation one might expect with various departures from normality, as measured by the lower moment ratios. (The assumption that there are ***n independent*** pairs of observations, each with the ***same*** joint distribution, has been retained.)

Quensel (1938) supposed that cumulants (and mixed cumulants) of order higher than four were negligible and that the population value of the correlation coefficient ρ was zero. **Gayen** (1951) extended this work by allowing ρ to be nonzero. He obtained an expansion for the density function in terms of the right-hand side of Eqs. **(32.6a)–(32.6f)** [here denoted by $f(r,\rho)$] and its derivatives with respect to ρ. The formula is

$$p_R(r) = f(r,\rho) + \frac{n-1}{8n(n+1)}\left\{L_{4,1}\frac{\partial f}{\partial \rho} + L_{4,2}\frac{\partial^2 f}{\partial \rho^2}\right\} + \frac{n-2}{12n(n+1)(n+3)}\left\{L_{6,1}\frac{\partial f}{\partial \rho} + L_{6,2}\frac{\partial^2 f}{\partial \rho^2} + L_{6,3}\frac{\partial^3 f}{\partial \rho^3}\right\}, \tag{32.30}$$

where the L's are functions of n, ρ and the cumulant ratios $\gamma_{ij} = \kappa_{ij}\kappa_{20}^{-i/2}\kappa_{02}^{-j/2}$, namely

$$L_{4,1} = 3\rho(\gamma_{40} + \gamma_{04}) - 4(\gamma_{31} + \gamma_{13}) + 2\rho\gamma_{22},$$

$$L_{4,2} = \rho^2(\gamma_{40} + \gamma_{04}) - 4\rho(\gamma_{31} + \gamma_{13}) + 2(2 + \rho^2)\gamma_{22},$$

$$\begin{aligned} L_{6,1} = & -15\rho(\gamma_{30}^2 + \gamma_{03}^2) - 9\rho\left(1 + \frac{2}{n-2}\right)(\gamma_{21}^2 + \gamma_{12}^2) - \frac{6}{n-2}\gamma_{30}\gamma_{03} \\ & + 6\left(2 + \frac{1}{n-2}\right)\gamma_{21}\gamma_{12} + 18(\gamma_{30}\gamma_{21} + \gamma_{03}\gamma_{12}) \\ & + \frac{18\rho}{n-2}(\gamma_{30}\gamma_{12} + \gamma_{03}\gamma_{21}), \end{aligned}$$

$$\begin{aligned} L_{6,2} = & -9\rho^2(\gamma_{30}^2 + \gamma_{03}^2) - 3\left(4 + 5\rho^2 - \frac{2(2-5\rho^2)}{n-2}\right)(\gamma_{21}^2 + \gamma_{12}^2) \\ & - \frac{18\rho}{n-2}\gamma_{30}\gamma_{03} + 18\rho\left(2 + \frac{1}{n-2}\right)\gamma_{21}\gamma_{12} + 30\rho(\gamma_{30}\gamma_{21} + \gamma_{03}\gamma_{12}) \\ & - 6\left(2 + \frac{2-5\rho^2}{n-2}\right)(\gamma_{30}\gamma_{12} + \gamma_{03}\gamma_{21}), \end{aligned}$$

$$\begin{aligned} L_{6,3} = & -\rho^3(\gamma_{30}^2 + \gamma_{03}^2) - 3\rho\left(2 + \rho^2 - \frac{2(1-\rho^2)}{n-2}\right)(\gamma_{21}^2 + \gamma_{12}^2) \\ & + 2\left(1 + \frac{3(1-\rho^2)}{n-2}\right)\gamma_{30}\gamma_{03} + 6\left(1 + 2\rho^2 - \frac{1-\rho^2}{n-2}\right)\gamma_{21}\gamma_{12} \\ & + 6\rho^2(\gamma_{30}\gamma_{21} + \gamma_{03}\gamma_{12}) - 6\rho\left(1 + \frac{1-\rho^2}{n-2}\right)(\gamma_{30}\gamma_{12} + \gamma_{03}\gamma_{21}). \end{aligned}$$

Cook (1951a, b) obtained expressions (up to and including terms in n^{-2}) for the first four moments of R in terms of the cumulants and cross-cumulants of the parent population (without specifying the exact form of this population distribution). The second and third terms, respectively, of (32.30) can be regarded as corrections to the normal density $p_R(r|\rho)$, for kurtosis and skewness. Gayen (1951) gave values of these corrective terms for certain special cases. When ρ is zero, the terms are small, even for n as small as four. However, an example with $\rho = 0.8$ shows that quite substantial corrections can be needed when n is not large.

Gayen further discussed the distribution of $Z' = \tanh^{-1} R$ (see the next section) obtaining expansions for its expected value, variance, β_1, and β_2. He found that β_1 and β_2 of Z' still tend to the normal values of 0 and 3 as n increased, though not so rapidly as when the parent population is normal. Table 32.2 gives the leading terms in expressions for moments and moment ratios of R and Z'. Cheriyan (1945) reported results of sampling experiments on distribution of correlation coefficient in random samples from certain bivariate gamma distributions.

In the last 20 years there has been a proliferation of papers on the distribution of R in samples from nonnormal bivariate distributions. A frequent feature of the methodology has been the use of Cornish-Fisher expansions (Chapter 12, Section 5), following the lines explored by Quensel (1938), Gayen (1951), and Cook (1951a, b). Nakagawa and Niki (1992) have extended these results, obtaining expressions for cumulants of R up to and including terms of order n^{-3}. (The general expression for the approximate fourth cumulant is extremely lengthy, containing some 345 terms!). They present results of simulation, exhibiting the improvement in accuracy achieved by inclusion of the terms of order n^{-3}. Apart from a bivariate normal parent population, they consider two other special cases—of samples from *bivariate* uniform distributions over (1) a parallelogram and (2) a trapezoid—as shown in Figures 32.1*a*, *b*. When $d = 0$, both (1) and (2) of course are the same standard bivariate uniform distribution. The values of the correlation coefficient are, for the parallelogram,

$$d(1 + d^2)^{-1/2}, \tag{32.31a}$$

for the trapezoid,

$$d(12 - d^2)^{1/2}(48 + 24d^2 - d^4)^{-1/2}. \tag{32.31b}$$

Since 1970 there has been a number of papers on the distribution of R in random samples from mixtures of two bivariate normal distributions [Bebbington (1978); Kocherlakota and Kocherlakota (1981); Srivastava (1983); Srivastava and Awan (1982, 1984); and Srivastava and Lee (1984)]. Some of these will now be summarized, using the following common notation, for convenience: The bivariate normal distribution of two random variables X. Y

Table 32.2 Leading terms in expansions

	R	Z'
Expected value	$\rho + \frac{1}{n}\left[-\frac{1}{2}\rho(1-\rho^2) + \frac{1}{8}L_{4,1}\right]$	$\frac{1}{2}\log\frac{1+\rho}{1-\rho} + \frac{1}{n-1} \times \left[\frac{1}{2}\rho + \frac{1}{8(1-\rho^2)^2}\{\rho(3-\rho^2)(\gamma_{40}+\gamma_{04}) - 4(1+\rho^2) \times (\gamma_{31}+\gamma_{13}) + 2\rho(5+\rho^2)\gamma_{22}\}\right]$
Variance	$\frac{1}{n}\left[(1-\rho^2)^2 + \frac{1}{4}L_{4,2}\right]$	$\frac{1}{n-1}\left[1 + \frac{1}{4(1-\rho^2)^2}\{\rho^2(\gamma_{40}+\gamma_{04}) - 4\rho(\gamma_{31}+\gamma_{13}) + 2(2+\rho^2)\gamma_{22}\}\right]$

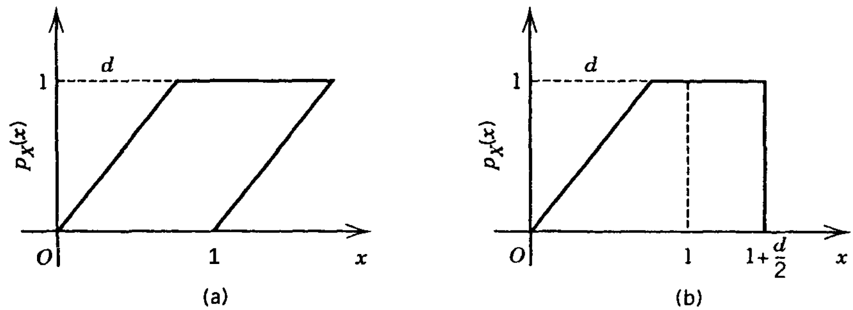


Figure 32.1 Bivariate uniform distributions on parallelogram (a) and on trapezoid (b).

has the pdf

$$\phi(x, y; \xi, \eta; \sigma_X, \sigma_Y; \rho)$$
$$= \left\{2\pi\sqrt{1-\rho^2}\right\}^{-1} \exp\left[-\frac{1}{2}(1-\rho^2)^{-1}\left\{\left(\frac{x-\xi}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\xi}{\sigma_X}\right)\left(\frac{y-\eta}{\sigma_Y}\right) + \left(\frac{y-\eta}{\sigma_Y}\right)^2\right\}\right] \quad \text{[cf. (32.2)]} \quad (32.32a)$$

and the cdf

$$\Phi(x, y; \xi, \eta; \sigma_X, \sigma_Y; \rho) = \int_{-\infty}^{y}\int_{-\infty}^{x} \phi(u, v; \xi, \eta; \sigma_X, \sigma_Y; \rho)\, du\, dv. \quad (32.32b)$$

All the papers just mentioned use two-component mixture models with the cdfs of form

$$\omega\Phi(x, y; \xi_1, \eta_1; \sigma_{X1}, \sigma_{Y1}; \rho_1) + (1-\omega)\Phi(x, y; \xi_2, \eta_2; \sigma_{X2}, \sigma_{Y2}; \rho_2) \quad (32.33)$$

in various specializations.

Bebbington (1978) studied the distribution of R in random samples of size n from the mixture distribution with the cdf (32.33) and

$$\omega = 0.98,$$
$$\xi_1 = \xi_2 = \eta_1 = \eta_2 = 0,$$
$$\sigma_{X1} = \sigma_{Y1} = 1,$$
$$\sigma_{X2} = \sigma_{Y2} = 3,$$
$$\rho_1 = \rho,$$
$$\rho_2 = 0,$$

regarding this as a "contaminated" bivariate normal distribution, the second term representing the "contamination." Bebbington's mixture is the bivariate analog of the univariate mixture used when the effect of outliers is of interest. Simulations indicated that R is (as would be expected) biased toward zero as an estimator of ρ. The effect was greater for large ρ. (For n = 50 and $\rho = 0.8$, the average of the simulated values of R was 0.688.)

Bebbington suggested "hull trimming"—deletion of observed (x, y) values on the bivariate convex hull of the set of n observed values (the vertices of the smallest area convex polyhedron including all observed values)—to improve R as an estimator of ρ. Titterington (1978) suggested a more refined method of trimming, based on minimal covering ellipsoids; see also Tiku and Balakrishnan (1986) for a different robust estimator of ρ based on trimming.

Srivastava and Lee (1984) and Srivastava and Awan (1984) used the mixture distribution (32.33) with

$$\xi_1 = \xi_2 = \eta_1 = \eta_2 = 0,$$

$$\sigma_{X_1} = \sigma_{Y_1} = 3, \quad \sigma_{X_2} = \sigma_{Y_2} = 1, \quad \rho_1 = \rho_2 = 0.$$

Figures 32.2*a*, b, taken from Srivastava and Lee (1984), show the corresponding pdfs of R, for sample sizes n = 6, 10, respectively and $\omega = 0.5, 0.7, 0.9, 0.95$, and 1 (corresponding to a single bivariate normal distribution).

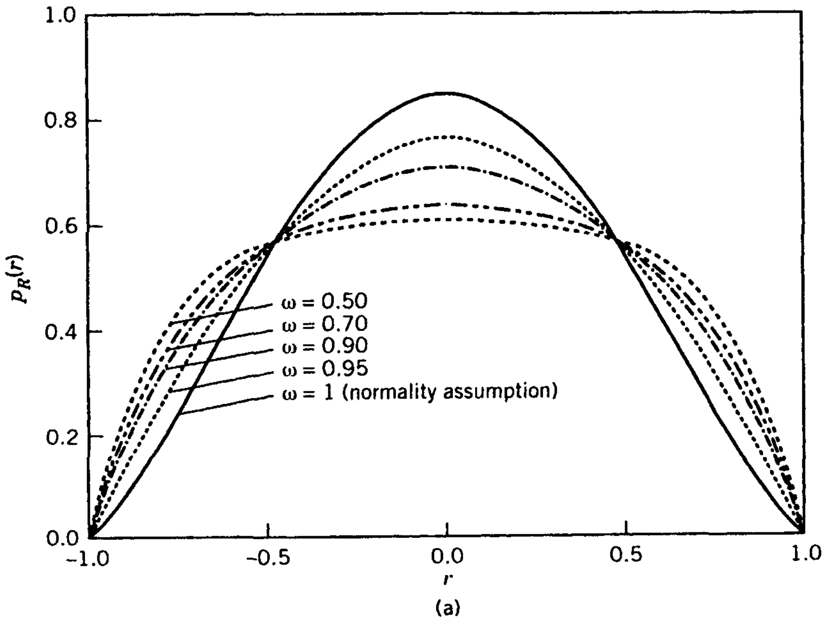


Figure 32.2a The probability density functions of the sample correlation coefficient R ($n = 6$). Note that when $\omega = 1$, the corresponding graph is the probability density function of R under the normality assumption.

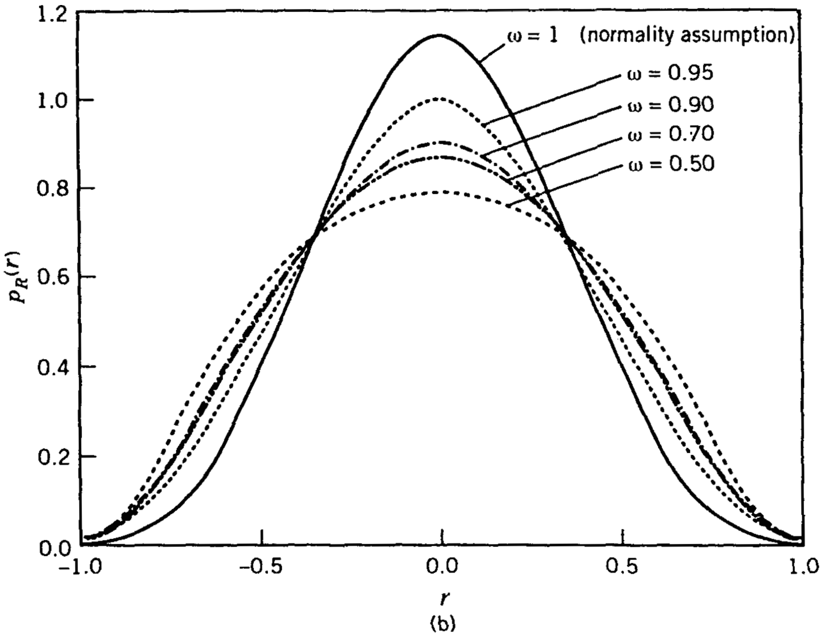


Figure 32.2b The probability density functions of the sample correlation coefficient R ($n = 10$). Note that when $\omega = 1$, the corresponding graph is the probability density function of R under the normality assumption.

These graphs confirm Kowalski's (1972) and Duncan and Layard's (1973) refutation of E. S. Pearson's (1929) blanket claim of robustness of R to the underlying mixture distributions. On the other hand, in the case of common covariance matrix, Srivastava and Awan's (1984) investigations seem to refute Duncan and Layard's (1973) assertion (for the case of different covariance matrices) that when $\rho = 0$ does not imply independence, the R statistic is likely to be sensitive to departures from normality. The results are thus still inconclusive and depend upon the correlation structure of the bivariate normal model assumed.

Srivastava and Awan (1984) also present an explicit formula for the pdf of R when

$$\sigma_{X1} = \sigma_{X2} = \sigma_X,$$

$$\sigma_{Y1} = \sigma_{Y2} = \sigma_Y,$$

$$\rho_1 = \rho_2 = \rho$$

(i.e., there is a common variance-covariance matrix), but it is possible that $\xi_1 \neq \xi_2$ and/or $\eta_1 \neq \eta_2$. The general formula is very complicated. Even in the case $\rho = 0$, Srivastava and Awan obtain

$$p_R(r|\rho = 0) = \sum_{g=0}^{n}\binom{n}{g}\omega^g(1-\omega)^{n-g}\left[\sqrt{\pi}\,\Gamma\left(\frac{1}{2}n-1\right)\right]^{-1}$$

$$\times\exp\left\{\frac{1}{2}\left(\delta_{1g}^2+\delta_{2g}^2\right)\right\}\sum_{h=0}^{\infty}\frac{\left(\delta_{2g}^2/2\right)^h}{h!\Gamma\left(\frac{1}{2}(n-3)+h+1\right)}$$

$$\times\sum_{i=0}^{h}\binom{h}{i}r^i(1-r^2)^{(n-4)/2}\left(\frac{2\delta_{1g}}{\delta_{2g}}\right)^i\sum_{j=0}^{h-i}\binom{h-i}{j}\left(\frac{\delta_{1g}}{\delta_{2g}}\right)^{2j}$$

$$\times\Gamma\left(\frac{n-1+i}{2}+j\right)\Gamma\left(\frac{n-1-i}{2}+h-j\right), \qquad (32.34)$$

where

$$\delta_{1g} = \frac{\xi_2-\xi_1}{\sigma_X}\left\{\frac{g(n-g)}{n}\right\}^{1/2},$$

$$\delta_{2g} = \frac{\eta_2-\eta_1}{\sigma_Y}\left\{\frac{g(n-g)}{n}\right\}^{1/2},$$

which is quite complicated. (Note that without loss of generality one can take $\sigma_X = \sigma_Y = 1$, $\xi_1 = \eta_1 = 0$.)

The authors claim that convergence in (32.34) for "moderate" n (about 20) is achieved if h is truncated around 20, when the accuracy of computation is fixed at 10^{-20}. However, in the general case when $\rho \neq 0$, and especially when $\rho > \frac{1}{2}$, convergence is slower. They also conclude that for two-sided testing purposes, when the contamination is less than 10%, the discrepancy between bivariate normal and mixture of two bivariate normals "is not so serious." Compare with Bebbington's (1978) conclusions, cited above.

A major interest has been the relative stability ("robustness") of certain transforms of R, especially the following transforms:

1. [Fisher (1921)] $Z' = \tanh^{-1} R$.
2. [Nair-Pillai; Pillai (1946)] $(R-\rho)(1-R\rho)^{-1}$.
3. [Samiuddin (1970)] $(n-2)^{1/2}(R-\rho)[(1-R^2)(1-\rho^2)]^{-1/2}$.
4. [Harley (1956)] $\sin^{-1} R$
5. [Ruben (1966)] $(a\tilde{R} + b\tilde{\rho})\{1 + \frac{1}{2}(\tilde{R}^2 + \tilde{\rho}^2)\}^{-1/2}$, where $\tilde{R} = R(1-R)^{-1/2}$, $\tilde{\rho} = \rho(1-\rho^2)^{-1/2}$; $a = (n-\frac{5}{2})^{1/2}$; and $b = (n-\frac{3}{2})^{1/2}$.

Kocherlakota and Kocherlakota (1981) study the distribution of R from samples of size n from the mixture (32.33) with $\xi_1 = \xi_2 = 77, = \eta_2$ $(= 0)$, $\sigma_{X1} = \sigma_{X2} = \sigma_{Y1} = \sigma_{Y2}$ $(= 1)$, but $p, \neq \rho_2$. The skewness of the mixture is zero. The kurtosis is

$$\{\omega(1-\omega)\}^2(3A_1^2 + 3A_2^2 + 2A_1A_2), \tag{32.35}$$

where

$$A_1 = \frac{(1+\rho_2)^2 - (1+\rho_1)^2}{\omega(1+\rho_1)^2 + (1-\omega)(1+\rho_2)^2},$$

$$A_2 = \frac{(1-\rho_2)^2 - (1-\rho_1)^2}{\omega(1-\rho_1)^2 + (1-\omega)(1-\rho_2)^2}.$$

The kurtosis, and also p [= corr(X, Y)] were studied as functions of $p,, \rho_2$, and ω. As ω increases the departure from bivariate normality (zero kurtosis) generally increases at first and then falls off. The maximal departure from normality tends to be high when $p,$ and ρ_2 are apart, and was observed to be the highest for $p, = 0.1$ and $\rho_2 = 0.8$ with $\omega = 0.3$. Formulas are provided for the moments of R, using Cook's (1951a) expansions in the general bivariate case and the robustness of transforms 1 through 4 to this type of nonnormality is studied. Nair and Pillai's and Samiuddin's transforms [2 and 3, respectively] are the most robust, with Fisher's Z', transform 1, nearly as robust. The arcsine transform 4 does not seem to be robust to this type of nonnormality. As in the case of other types of nonnormality, investigated by Subrahmaniam and Gajjar (1979), an increase in the sample size does not greatly reduce the kurtosis. Unlike the situation in the normal case, transforms 1 and 4 remain stubbornly biased with little or no reduction as the sample size increases. Even for the "worst" nonnormal situation ($\rho_1 = 0.1$ and $\rho_2 = 0.8$), Nair-Pillai's transform 2 appears to be quite robust.

Other papers on the effects of nonnormality include Yang (19701, Kowalski (1972), Zeller and Levine (1974), Havlicek and Peterson (1977), Subrahmaniam and Gajjar (1978), Kocherlakota and Singh (1982a, b), Kocherlakota, Kocherlakota, and Balakrishnan (19851, Kocherlakota, Balakrishnan, and Kocherlakota (1986), Fowler (1987), and Shanmugam and Gajjar (1992).

Subrahmaniam and Gajjar (1980) examined all five of transforms, obtaining expansions for the first four moments up to and including terms of order n^{-1}, using the methods of Cook (1951a, b). In particular, they compared the behavior of the five transforms under the assumption of a doubly right-truncated bivariate normal distribution with the pdf

$$\frac{\phi(x, y; 0, 0; 1, 1; \rho)}{\Phi(a, b; 0, 0; 1, 1; \rho)}, \qquad x \le a,\ y \le b. \tag{32.36}$$

As indexes of robustness they used the differences between the expected values and ratios of standard deviations from the values they would have for normal theory (i.e., with a and b each infinite). For expected value Nair-Pillai's transform 2 is the sturdiest, followed closely by Samiuddin's transform 3 and Ruben's transform 5. It is of interest that for Fisher's transform 1 and Harley's transform 4, increase in sample size n does not reduce the effect of nonnormality in this respect.

If comparison is based on the ratios (actual/normal theory) of standard deviations, a notable feature is that higher values of $|\rho|$ lead to higher standard deviations in the nonnormal case if ρ is positive, while the reverse is the case if ρ is negative. Increase in sample size does not affect the comparisons greatly. For large $|\rho|$ the least robust seems to be Harley's transform 4.

These conclusions were based on parent distributions with $\rho = \pm 0.05$, ± 0.25, ± 0.5, ± 0.75, and truncation points $a = b = -2.5, -1.5, 0, 0.5$. [See also Bebbington (1978); Gajjar and Subrahmaniam (1978).]

Shanmugam and Gajjar (1992) compare the transforms 1 through 4 for the Farlie-Gumbel-Morgenstern joint bivariate exponential distributions with pdfs of form

$$p_{X,Y}(x, y; \alpha) = e^{-(x+y)}\{1 + \alpha(2e^{-x} - 1)(2e^{-y} - 1)\},$$
$$x, y > 0,\ |\alpha| < 1. \tag{32.37}$$

[The population correlation coefficient (ρ) has the value $\frac{1}{4}\alpha$ in this case, so it cannot exceed $\frac{1}{4}$ in absolute value.] They find that Fisher's Z' (transform 1) is the most robust for small values of $|\rho| \le 0.05$ as judged in the same way as above. Transforms 2 through 4, on the other hand, are rather severely affected by nonnormality, at least of this kind.

Kocherlakota and Singh (1982b) study the distributions of transforms 1 through 5, using the cumulants/Cornish-Fisher technique, for R based on random samples from two population distributions:

1. ***Bivariate t.*** $X = W_1/\sqrt{V}$, $Y = W_2/\sqrt{V}$ with (W_1, W_2) having a joint bivariate normal distribution with the pdf $\phi(w_1, w_2; 0, 0; 1, 1; \rho)$ and V, independent of (W_1, W_2), distributed as χ^2_ν/ν.
2. ***Bivariate*** χ^2. $X = \sum_{j=1}^{\nu} W_{1j}^2$, $Y = \sum_{j=1}^{\nu} W_{2j}^2$ with (W_{1j}, W_{2j}), $j = 1, 2, \dots, \nu$, mutually independent, each distributed as (W_1, W_2) as given in 1, (namely as a bivariate t).

5 TABLES AND APPROXIMATIONS (ASYMPTOTIC EXPANSIONS)

5.1 Tables

From Section *2* it is clear that the distribution of R is so complicated that it is very difficult to use it without a practicable approximation or an extensive set of tables. We may note that this distribution is important because it is the distribution of the correlation coefficient in random samples from a bivariate normal population, not because it is suitable for fitting purposes.

David *(1938)* prepared a useful set of tables. She gave values of the probability density $p_R(r; p, n)$ and of the cumulative distribution function $F_R(r; p, n)$ to five decimal places for $n = 3(1)25, 50, 100, 200, 400$. For $n \leq 25$, $r = -1.00(0.05)1.00$ for $p = 0.0(0.1)0.4$ and $r = -1.00(0.05)0.600(0.025)1.000$ for $p = 0.5(0.1)0.9$, with additional values for $r = 0.80(0.01)0.900(0.005)1.000$ when $p = 0.9$. For $n > 25$ narrower intervals are used. [Note that the quantity tabulated as "ordinate" is actually $1000\ p_R(r)$.] The Introduction to these tables contain some interesting notes on the distribution of R.

Subrahmaniam and Subrahmaniam (1983) extended David's tables, giving $F_R(r; p, n)$ to five decimal places for $n = 26(1)49$, $p = 0.1(0.1)0.9$, $r = -0.70(0.05)0.65(0.025)0.975$, and also quantiles $r_\alpha(\rho; n)$ satisfying $F_R(r_\alpha(\rho; n)) = a$ for $a = 0.01, 0.02, 0.025, 0.05, 0.1, 0.2, 0.8, 0.9, 0.95, 0.975, 0.98, 0.99$ and $n = 4(1)50, 100, 200, 400$ to four decimal places.

Boomsma *(1975)* noted that in David's (1938) tables the numbers in the "Area" (cdf) column on page *50,* for $n = 100$, $p = 0.4$, have all been displaced one position upward. These errors were also noted by Subrahmaniam and Subrahmaniam (1983) in their extension of David's tables. They gave corrected values of $F_R(r; p, n)$ for $n = 100$, $p = 0.4$, $r = -0.10(0.05)0.70$. [Their computational technique used a 32-point quadrature (DQG-32).]

Öksoy and Aroian (1981) provided values of $F_R(r; p, n)$ and $p_R(r; p, n)$ to six decimal places for $n = 3(1)6, 35, 40(10)60$ and $p = 0, 0.5, 0.98$. Öksoy and Aroian *(1982)* also provided values of $r_\alpha(\rho; n)$ to four decimal places for the same values of n and p, and for $a = 0.0005, 0.001, 0.0025(0.0025)0.0100, 0.0175, 0.0250, 0.0375, 0.05, 0.075, 0.10(0.05)0.90, 0.925, 0.95, 0.9625, 0.975, 0.9825, 0.99(0.0025)0.9975, 0.9990, 0.9995$. They also refer to more detailed, unpublished tables of $F_R(r; p, n)$ for $n = 3(1)10(2)24, 25(5)40(10)100(100)500$, and $p = 0(0.05)0.90, 0.92, 0.94, 0.95, 0.96, 0.98$. They recommend use of Garwood's (1933) formulas [see Table 32.1] for rapid calculation of the density function, the cdf, and quantiles, and they noted that Hotelling's *(1953)* formula (32.6e) for the pdf is also useful in this respect.

Odeh and Owen *(1980,* pp. 228–264) provided tables of confidence limits —values of $\rho_\alpha(r; n)$ satisfying $F_R(r; \rho_\alpha(r; n); n) = a$ for $n = 3(1)30, 40, 60, 100, 120, 150, 200, 400$, $r = -0.95(0.05)0.95$, and $\alpha = 0.005, 0.01,$

0.025, 0.05, 0.1, 0.25, 0.75, 0.90, 0.95, 0.975, 0.99, 0.995. The interval $(\rho_{1-(\varepsilon/2)}(R;n), \rho_{\varepsilon/2}(R;n))$ is a $100(1-\varepsilon)\%$ confidence interval for p, given R. Their methods of calculation (given on pp. 287–294) depend on Garwood's *(1933)* formula, on an adaption of Hotelling's formula for cdf, using a series involving incomplete beta functions, and on a new formulation for the confidence limits of p, using Fisher's transformation as a starting point.

Extensive tables of quantiles $[r_\alpha(\rho;n)]$ were provided by Odeh (1982), giving values to five decimal places for p = 0.0(0.1)0.9, 0.95, n = 4(1)30(2)40(5)50(10)100(20)200(100)1000, and α = 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.75, 0.9, 0.95, 0.975, 0.99, 0.995. Odeh (1983), using the algebraic programming system REDUCE 2 (on an IBM 4341 computer) provided exact expressions for $F_R(r; \mathrm{p}, n)$ for values n = 3(1)10. For n = 5,6 he used essentially the expressions of Garwood (1933) as rewritten by Öksoy and Aroian (1981, 1982, Table 1 et seq.). For n = 7(1)10 Odeh's expressions are obtained using the basic recurrence formulas of Hotelling (1953). For n = 9, 10 the first nine derivatives of

$$Q(r\rho) = (1-r^2\rho^2)^{-1/2}\cos^{-1}(-r\rho) \qquad \text{[see (32.9)]} \qquad (32.38)$$

are presented in Öksoy and Aroian (1982). These derivatives, which occur in the formulas for $F_R(r; \mathrm{p}, n)$, involve polynomials in pr and Q itself in the numerator, together with powers of $1-(\rho r)^2$ in the denominator. For odd n, however, p does not appear in the denominator of these expressions. Guenther (1977), on the other hand, advocated the use of formula (32.6a) for numerical calculation of the cdf of the sample correlation coefficient on "modern desk calculators" [such as Monroe 1930]. He found that the terms in the formula for Pr[0 < R < r], based on (32.6a), converge quite rapidly with good error bounds.

Guenther (1977) derived the formula

$$\begin{aligned}\Pr[0<R<r] = {}& \frac{1}{2}\sum_{j=0}^{\infty} K_1(j)\Pr\left[F_{2j+1,n-2} \le \frac{(n-2)r^2}{1-r^2}\frac{1}{2j+1}\right] \\ & + \frac{1}{2}\sum_{j=0}^{\infty} K_2(j)\Pr\left[F_{2j+1,n-2} \le \frac{(n-2)r^2}{1-r^2}\frac{1}{2j+2}\right], \\ & \qquad 0<r\le 1, \qquad (32.39)\end{aligned}$$

where $F_{v_1,v_2}(\cdot)$ is the cumulative distribution function of a F-variable with (v_1, v_2) degrees of freedom (see Chapter 27),

$$K_1(j) = \frac{\Gamma(\frac{1}{2}(n-1)+j)}{j!\Gamma(\frac{1}{2}(n-1))}(1-\rho^2)^{(n-1)/2}\rho^{2j},$$

$$K_2(j) = \frac{\Gamma(\frac{1}{2}n+j)}{\Gamma(j+\frac{3}{2})\Gamma(\frac{1}{2}(n-1))}(1-\rho^2)^{(n-1)/2}\rho^{2j+1}.$$

[Note that the $K_1(j)$'s are terms in the negative binomial expansion of $(1-\rho^2)^{-(n-1)/2}$.] He combined this with a formula for $\Pr[R>0]$ [(see, e.g., (32.11) or (32.15)] to evaluate

$$\Pr[R>r] = \Pr[R>0] - \Pr[0<R\le r]. \tag{32.40}$$

Guenther (1977) showed that if the first and second series on the right-hand side of (32.39) are terminated at $j=r_1$ and $j=r_2$, respectively, the error in $F_R(r;n,\rho)$ is negative and less than

$$\frac{1}{2}H_{2r_1+1,n-2}\left(\frac{n-1}{2r_1+1}\frac{\rho^2}{1-\rho^2}\right)H_{2r_1,n-1}\left(\frac{n-1}{2r_1}\frac{\rho^2}{1-\rho^2}\right)$$
$$+\frac{1}{2}H_{2r_2+2,n-2}\left(\frac{n-1}{2r_2+2}\frac{\rho^2}{1-\rho^2}\right)H_{2r_2+1,n-1}\left(\frac{n-1}{2r_1+1}\frac{\rho^2}{1-\rho^2}\right), \tag{32.41}$$

where $H_{\nu_1,\nu_2}(y) = \Pr[F_{\nu_1,\nu_2}\le y]$, in absolute value. He presented numerical results demonstrating the accuracy of his formulas.

Stammberger (1968) has constructed a nomogram from which it is possible to determine any one of $\Pr[R\le r]$, r or n, given the other two values.

5.2 Approximations Using Transforms

For most practical purposes, approximations to the distribution of R use Fisher's (1915, 1921) transformation

$$Z' = \tanh^{-1}R = \frac{1}{2}\log\left(\frac{1+R}{1-R}\right). \tag{32.42}$$

This transformation might be suggested as a variance-equalizing transformation, noting that [from (32.14b)]

$$\operatorname{var}(R) \doteqdot (1-\rho^2)^2 n^{-1}, \tag{32.43}$$

$$\int (1-\rho^2)^{-1}\,d\rho = \frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

This approach, however, was not explicitly used by Fisher (1915) in his original suggestion of this transformation.

Approximate values of moments and moment ratios of Z' are

$$\mu'_1(Z') \doteqdot \frac{1}{2}\log\frac{1+\rho}{1-\rho} + \frac{1}{2}\rho(n-1)^{-1}\left[1+\frac{1}{4}(5+\rho^2)(n-1)^{-1}\right], \tag{32.44a}$$

$$\mu_2(Z') \doteqdot (n-1)^{-1}\left[1+\frac{1}{2}(4-\rho^2)(n-1)^{-1} + \frac{1}{6}(22-6\rho^2-3\rho^4)(n-1)^{-2}\right], \tag{32.44b}$$

$$\mu_3(Z') \doteqdot (n-1)^{-3}\rho^3, \tag{32.44c}$$

$$\mu_4(Z') \doteqdot 3(n-1)^{-2}\left[1+\frac{1}{3}(14-3\rho^2)(n-1)^{-1} + \frac{1}{12}(184-48\rho^2-21\rho^4)(n-1)^{-2}\right], \tag{32.44d}$$

$$\beta_1(Z') \doteqdot (n-1)^{-3}\rho^6, \tag{32.45a}$$

$$\beta_2(Z') \doteqdot 3+2(n-1)^{-1}+(4+2\rho^2-3\rho^4)(n-1)^{-2}. \tag{32.45b}$$

These values were given by Fisher (1921) and corrected later by Gayen (1951). [See also Nabeya (1951).] Comparing (32.45a) and (32.45b) with (32.15a) and (32.15b), it can be seen that the (β_1, β_2) values for Z' are much closer to the normal values (0, 3) than are the values for R. Also var(Z') does not depend on ρ up to and including terms of order $(n-1)^{-1}$.

The most commonly used approximation is to regard Z' as normally distributed with expected value $\frac{1}{2}\log((1+\rho)/(1-\rho))$ and variance $(n-3)^{-1}$. The latter value is obtained by noting that

$$\begin{aligned}(n-1)^{-1} + \tfrac{1}{2}(4-\rho^2)(n-1)^{-2} &= n^{-1}+n^{-2}+\cdots+\tfrac{1}{2}(4-\rho^2)n^{-2}\dots\\ &= n^{-1}\left[1+(3-\tfrac{1}{2}\rho^2)n^{-1}\dots\right]\\ &\doteqdot n^{-1}[1-3n^{-1}]^{-1} = (n-3)^{-1}.\end{aligned}$$

To improve the approximation, the expected value may be increased by $(2n-5)^{-1}\rho$. Fowler (1987) took the value of $E[Z']$ from (32.44a), and var(Z') from (32.44b), and treated Z' as a normally distributed variable, resulting in "good approximation in the tails."

Another kind of approximation can be based on consideration of the structure of the random variable R. Recalling the argument used, at the beginning of Section 2, in deriving the distribution of R, we see that for a fixed set of values of $X_1, X_2, \cdots, X_n$, $R(1 - R^2)^{-1/2}$ is distributed as

$$(n-2)^{-1/2} t'_{n-2}\left(\frac{\sqrt{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2}\,\rho}{\sqrt{1-\rho^2}}\right), \tag{32.46}$$

that is, as

$$\frac{U + \sqrt{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2}\,\rho/\sqrt{1-\rho^2}}{\chi_{n-2}},$$

where U is a unit normal variable and U and χ_{n-2} are mutually independent. Averaging over the joint distribution of the X's, we see that $R(1 - R^2)^{-1/2}$ is distributed as

$$\frac{U + \chi_{n-1}\rho\left(1-\rho^2\right)^{-1/2}}{\chi_{n-2}}, \tag{32.47}$$

where u, χ_{n-1}, and χ_{n-2} are mutually independent. This representation was constructed by Ruben (1963, 1966) who used it as the basis for the following approximations. From the representation (32.47), we have

$$\Pr[R \le r] = \Pr\left[U + \rho\left(1-\rho^2\right)^{-1/2}\chi_{n-2} - r\left(1-r^2\right)^{-1/2}\chi_{n-1} \le 0\right]. \tag{32.48}$$

Provided that n is not too small, χ_{n-2} and χ_{n-1} may be approximated by normal variates according to Fisher's approximation that $\{\sqrt{2\chi_n^2} - \sqrt{2n-1}\}$ is approximately distributed as a unit normal variable. Then

$$U + \rho\left(1-\rho^2\right)^{-1/2}\chi_{n-2} - r\left(1-r^2\right)^{-1/2}\chi_{n-1}, \text{ for fixed } r,$$

is approximately distributed normally with expected value

$$\rho\left(1-\rho^2\right)^{-1/2}\left(n-\tfrac{5}{2}\right)^{1/2} - r\left(1-r^2\right)^{-1/2}\left(n-\tfrac{3}{2}\right)^{1/2}$$

and standard deviation

$$\left[1 + \tfrac{1}{2}\rho^2\left(1-\rho^2\right)^{-1} + \tfrac{1}{2}r^2\left(1-r^2\right)^{-1}\right]^{1/2}.$$

Hence from *(32.48)*

$$\Pr[R \le r] \doteq \Phi\left(\frac{r(1-r^2)^{-1/2}(n-\frac{3}{2})^{1/2} - \rho(1-\rho^2)^{-1/2}(n-\frac{5}{2})^{1/2}}{\left[1+\frac{1}{2}r^2(1-r^2)^{-1} + \frac{1}{2}\rho^2(1-\rho^2)^{-1}\right]^{1/2}} \right). \tag{32.49}$$

Muddapur (1988) showed that in the special case when $\sigma_X = \sigma_Y$, the statistic

$$T^* = \frac{R - \rho S}{\{(1-\rho^2)(1-R^2)\}^{1/2}} \sqrt{n-2}, \tag{32.50a}$$

where $S = \frac{1}{2}(S_X S_Y^{-1} + S_X^{-1} S_Y)$ and $S_X = \{(n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2\}^{1/2}$, $S_Y = \{(n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2\}^{1/2}$, has an exact t_{n-2} distribution. If it so happens further that $S_X = S_Y$, then (32.50a) would reduce to

$$T^{**} = \frac{R - \rho}{\{(1-\rho^2)(1-R^2)\}^{1/2}} \sqrt{n-2}. \tag{32.50b}$$

This statistic was proposed by Samiuddin (1970) as having ***approximately*** a t_{n-2} distribution (for ***all*** ρ and ***all*** σ_X/σ_Y) for moderately large n (even if $S_X \neq S_Y$).

Roughly equivalently,

$$V = \frac{R-\rho}{1-R\rho} = T^{**}\left\{\frac{(1-\rho^2)(1-R^2)}{n-2}\right\}^{1/2} \frac{1}{1-R\rho}$$

has an approximate beta $(\frac{1}{2}n - 1, \frac{1}{2}n - 1)$ distribution over $(-1, 1)$, with

$$\Pr[V \le v] = I_{(1+v)/2}(\tfrac{1}{2}n - 1, \tfrac{1}{2}n - 1).$$

This approximation was reported to be good for $n \ge 8$ and even for large ρ.

Muddapur *(1988)* pointed out that

$$G = \frac{(1+R)(1-\rho)}{(1-R)(1+\rho)} \tag{32.51}$$

has an approximate $F_{n-2,n-2}$ distribution. For $\rho = 0$, the distribution is exact. The relations

$$F_{\nu,\nu,\frac{\alpha}{2}} = 1 + 2\nu^{-1}t_{\nu,\alpha}^2 + 2\{\nu^{-1}(1 + t_{\nu,\alpha}^2\nu^{-1})\}^{1/2} t_{\nu,\alpha}, \tag{32.52a}$$

$$F_{\nu,\nu,1-\frac{\alpha}{2}} = 1 + 2\nu^{-1}t_{\nu,\alpha}^2 - 2\{\nu^{-1}(1 + t_{\nu,\alpha}^2\nu^{-1})\}^{1/2} t_{\nu,\alpha}, \tag{32.52b}$$

can be used in conjunction with (32.51) to construct approximate $100(1-\alpha)\%$ confidence intervals for p if values of the appropriate quantiles of $F_{n-2,n-2}$ are not available using the tables of the F-distribution.

Muddapur (1988) tabulated corresponding confidence limits for p for $\alpha = 0.025, 0.005$, and R = 0(0.05)0.95. These intervals seem to be quite wide even for n as large as 25, and especially so for small values of $|R|$.

A better approximation than Fisher's for a one-sided limit is obtained by using

$$T' = \frac{R'\sqrt{n-2}}{\sqrt{(1-R'^2)}} \tag{32.53}$$

as a t_{n-2} variable, where R' (given R and ρ) satisfies

$$\frac{W - E[W|\rho]}{\sqrt{\operatorname{var}(W|\rho)}} - \frac{\log\{(1+R')/(1-R')\}}{\sqrt{\operatorname{var}(W|\rho=0)}} \tag{32.54}$$

with W = $\log\{(1+R)/(1-R)\}$,

$$E[W|\rho] = \log\frac{1+\rho}{1-\rho} + \frac{\rho}{n-1}\left\{1 + \frac{5+\rho^2}{4(n-1)} + \frac{11+2\rho^2+3\rho^4}{8(n-1)^2} + \cdots\right\}, \tag{32.55a}$$

$$\operatorname{var}[W|\rho] = \frac{4}{n-1}\left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{22-6\rho^2-3\rho^4}{4(n-1)^2} + \cdots\right\}. \tag{32.55b}$$

Kraemer (1973) claimed that if the quantity p' $= \rho'(\rho, n)$ satisfies the conditions

1. $|\rho'(\rho,n)| \geq |\rho|$,
2. $\rho'(\rho,n) =$ p when p $= 0, -1$, or 1,
3. $\rho'(\rho,n) = -\rho'(-\rho,n)$,
4. $\lim_{n\to\infty}\rho'(\rho,n) =$ p,

then

$$\frac{\sqrt{(n-2)}\,(R-\rho')}{\sqrt{(1-R^2)(1-\rho'^2)}} = \frac{\sqrt{n-2}\,W}{\sqrt{1-W^2}}, \tag{32.56}$$

with W $= (R-\rho')/(1-R\rho')$, is approximately distributed as t_{n-2}. However, Mi (1990) has pointed out, by means of a counterexample, that these four conditions are not sufficient and that additional conditions of type

5. p' = p $+ o(n^{-1})$,
6. p' of same sign as p, and, possibly,
7. $|\rho'| \leq 1$,

may be needed. The choice of $\rho'(\rho, n) = \rho$ is a possibility [Samiuddin (1970)], but Kraemer (1973) recommended setting $\rho'(\rho, n)$ equal to the median of the distribution of R (given ρ and n) $= \rho^*(\rho, n)$, say. She provided a table of values of ρ^* for $\rho = 0.1(0.1)0.9$, and $n = 11(1)25, 50, 100, 200, 400$, appropriate for $|\rho|$ "not too near" to 1. [Mi (1990) noted that the median satisfies conditions 6 and 7 but was unable to ascertain whether it satisfies condition 5]. Kraemer's approximation is less cumbersome and quite accurate (see also the remark at the end of Section 5.3).

Niki and Konishi (1984) provided an asymptotic expansion for the median ρ^*:

$$\rho^* = \rho + (1 - \rho^2)\rho\left\{\frac{1}{2}n^{-1} + \frac{1}{24}(-7\rho^2 + 15)n^{-2} + \frac{1}{240}(47\rho^4 - 180\rho^2 + 165)n^{-3} + \frac{1}{40320}(-4945\rho^6 + 31227\rho^4 - 55755\rho^2 + 26145)n^{-4}\right\} \quad (32.57)$$

that is accurate to five decimal places for $n > 20$. Kraemer (1973) also used linear interpolation accurate only for three decimal places. She also suggested additional normal approximations, including

$$E[Z'] \doteqdot \frac{1}{2}\log\left(\frac{1+\rho^*}{1-\rho^*}\right),$$

$$\operatorname{var}(Z') \doteqdot \frac{1}{n-1} + \frac{2}{(n-1)(n+1)} + \frac{23}{3(n-1)(n+1)(n+3)} + \cdots, \quad (32.58)$$

and

$$E[Z'] \doteqdot \frac{1}{2}\log\left(\frac{1+\rho^*}{1-\rho^*}\right),$$

$$\operatorname{var}(Z') \doteqdot \frac{1}{n-3}. \quad (32.59)$$

She found the approximation (32.58) to be superior to David's (1938) approximation for small n and $|\rho| \le 0.6$, but for $|\rho| \ge 0.8$ David's approximation is the more accurate one. Thomas (1989) recommended David's approximation and provided a diagram that indicates that even for $n = 15$ and $\rho = 0.8$ the graphs of the approximate and exact cdf's are very close. (The largest discrepancy is in the vicinity of $R = 0.65$.) For $0.4 \le \rho \le 0.6$ the curves are virtually indistinguishable. However, other authors [e.g., Kraemer (1973), as noted above; Konishi (1978)] reach different conclusions.

Konishi (1978) derives the following approximation:

$$\Pr\left[m^{1/2}\left(\frac{1}{2}\log\frac{1+R}{1-R}-\frac{1}{2}\log\frac{1+\rho}{1-\rho}\right)\le x\right]$$

$$\doteqdot \Phi(x)-\frac{1}{2}\left(\rho m^{-1/2}+\frac{x^3}{6m}\right)\phi(x)+O(m^{-3/2}), \quad (32.60)$$

where $m = n - \frac{3}{2} + \frac{1}{4}\rho^2$. The approximation gives high accuracy over the whole range of variation of R for relatively small n. The optimal choice of m (more precisely the correction to n) is still an open problem. The approximation is more accurate than Ruben's (1966) and Kraemer's (1973) approximations [(32.48) and (32.56), respectively] for $\rho \ge 0.3$, and is far better for $\rho \ge 0.7$.

As Chaubey and Mudholkar (1978) have pointed out in this age of advanced computer technology many approximations to the distribution of R are of interest on the grounds of novelty, accuracy, and/or aesthetics rather than practicality. They noted that the reason that Fisher's approximation fails to maintain its accuracy for large values of $|\rho|$, even if n is not too small, is the high kurtosis of the distribution of Z'. From Gayen (1951) we have

$$\beta_1(Z')=\frac{\rho^6}{(n-1)^3}+\cdots, \quad (32.61a)$$

$$\beta_2(Z')=3+\frac{2}{n-1}+\frac{1}{(n-1)^2}(4+2\rho^2-3\rho^4)+\cdots \quad (32.61b)$$

(β_1 decreases more rapidly with n than does $\beta_2 - 3$).

Since $\beta_2 - 3$ of the Student's t-distribution with ν degrees of freedom is $6/(\nu-4)$, the number of degrees of freedom ν of the t distribution with approximately matching kurtosis is

$$\nu=4+\frac{6}{\beta_2-3}\doteqdot\frac{3n^2-2n+7}{n+1}. \quad (32.62)$$

Chaubey and Mudholkar proposed to approximate standardized Z' by a multiple of Student's t variable with $\nu_0 = [\nu]$ degrees of freedom, where ν is given by (32.62). The $100\alpha\%$ percentile of R is approximately given by

$$r_\alpha(\rho;n)\doteqdot\tanh\left[\mu+\sigma t_{\nu_0,\alpha}\left\{\frac{\nu_0-2}{\nu_0}\right\}^{1/2}\right], \quad (32.63)$$

where

$$\mu = \tanh' p + \tfrac{1}{2}(n-1)^{-1}\rho\{1 + f(5+\rho)^2(n-1)^{-1}\},$$

$$\sigma^2 = (n-1)^{-1}\{1 + \tfrac{1}{2}(n-1)^{-1}(4+\rho^2) + \tfrac{1}{6}(n-1)^{-2}(22 - 6\rho^2 - 3\rho^4)\}.$$

Utilizing Wallace's (1959) transformation [Chapter 28, Eq. (28.26)] to normalize the t-distribution, the authors arrived at

$$r_\alpha(\rho;n) \approx \tanh\left[\mu + \sigma\left\{(\nu-2)\left(\left(\exp\frac{1}{\nu}\left(\frac{3+8\nu}{1+8\nu}\xi_\alpha\right)^2\right) - 1\right)\right\}^{1/2}\right], \tag{33.64}$$

where $\Phi(\xi_\alpha) = a$. Numerical computations indicate that Chaubey and Mudholkar's (1978) approximation to cumulative probabilities and quantiles compares favorably with Kraemer's and Ruben's approximations in both simplicity and accuracy. (Ruben's and Kraemer's approximations for quantiles each involve solutions of quadratic equations).

5.3 Asymptotic Expansions of the Distribution of R

Asymptotic expansions of the distribution (under bivariate normality) of

$$Z'(R) = \frac{1}{2}\log\frac{1+R}{1-R} \tag{32.65}$$

were derived by Winterbottom (1980) and Niki and Konishi (1984), among others. Niki and Konishi consider the transformed variable

$$Z^* = n^{1/2}\{Z'(R) - Z'(\rho)\}. \tag{32.66}$$

They list the first ten cumulants of Z*, noting that κ_{2j+1} is of order $n^{-(2j+1)/2}$, instead of $n^{-(2j-1)/2}$. They obtain

$$\Pr[Z' < z] = \Phi(z) + \phi(z)\left[a_1 n^{-1/2} + a_2 n^{-1} + \cdots + a_8 n^{-4}\right] + O(n^{-9/2}), \tag{32.67}$$

where $\Phi(x)$ and $\phi(x)$ are the standard normal cdf and pdf, respectively, and the coefficients a_i depend on z and p. They can be expressed in terms of Hermite polynomials. The expansion (up to a,) involves the first 15 Hermite polynomials, while the corresponding Edgeworth expansion for the distribution function of R is more complicated, involving the first 23 Hermite polynomials.

The expansion including the first eight a_i's guarantees accuracy to five decimal points for a sample of size 11 or larger and to six decimal places for $n \geq 16$. Winterbottom's (1980) Cornish-Fisher form of Edgeworth expansion involves seven Hermite polynomials and is somewhat less accurate.

Mudholkar and Chaubey (1976) represented the distribution of Z' by a mixture of a normal and a logistic distribution, obtaining the approximation

$$\Pr[Z' < z] = \omega\Phi(x) + (1-\omega)L(x), \tag{32.68}$$

where

$$L(x) = \left\{1 + \exp\left(-\frac{\pi x}{\sqrt{3}}\right)\right\}^{-1},$$

with

$$x = \frac{z-\mu}{\sigma},$$

$$\mu = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) + \frac{1}{2}\rho n^{-1} + \frac{1}{8}\rho(5+\rho^2)n^{-2} + \cdots,$$

$$\sigma^2 = n^{-1} + \frac{1}{2}(4-\rho^2)n^{-2} + \cdots,$$

$$\omega = 1 - \frac{5}{6}(\beta_2 - 3),$$

$$\beta_2 = 3 + 2n^{-1} + (4 + 2\rho^2 - 3\rho^4)n^{-2} + \cdots.$$

The idea is to account for the excess kurtosis of the distribution of Fisher's Z' (which is of order n^{-1}) as compared with skewness β_1 (which is of order n^{-3}). The approximation seems to be of comparable accuracy to that of Chaubey and Mudholkar's (1978) approximation (32.64), involving a t-distribution, although the authors do not compare these approximations explicitly. For detailed comparison of Ruben's, Kraemer's, Mudholkar and Chaubey's (1976), and Winterbottom's (1980) approximations, the reader is referred to Winterbottom (1980); for those of Ruben (1966), Kraemer (1973), Winterbottom (1980), and Niki and Konishi, see Niki and Konishi (1984).

Niki and Konishi's (1984) approximation is extremely accurate and very complex. Despite the tremendous advances in computational procedures and analytic methodology in statistics during the last 50 years, the yardstick, so far as the distribution of correlation coefficient is concerned, is still the remarkable work of F. N. David in the 1930s, carried out with the rather primitive computational facilities available at that time but highly ingenious insofar as the approach to the numerical approximations of complicated functions is concerned. For small values of p, Kraemer's (1973) t-distribution approximation seems to be better than the more complicated omnibus approximations.

6 ESTIMATION OF p: ADDITIONAL ROBUSTNESS CONSIDERATIONS

6.1 General Remarks

The literature on estimation of correlation coefficients in the last 20 years is very extensive, and comes from mathematically oriented statistical journals, on the one hand, and sociological, psychological, and educational ones, on the other (the latter including numerous empirical investigations) with little coordination among them, causing substantial duplications and triplications. [Only the works of H. C. Kraemer (1973, 1980) seem to have appeared in both statistical and educational journals, providing a narrow, but a most welcome, bridge.] Contradictory, confusing, and uncoordinated floods of information on the "robustness" properties of the sample correlation coefficient R are scattered in dozens of journals. (Some evidence of this has been apparent in Sections 4 and 5). Various partial simulations and empirical conclusions seem to be directed, on the whole, toward justifying the application of R even in nonnormal situations and when $p \neq 0$. However, results of studies of this issue can be found in the literature, supporting both extreme points of view—of practical robustness, and unsettling volatility.

Sometimes there is failure to distinguish between nonnormality robustness (of the distribution of R and of properties of formal inference concerning ρ) and robustness with respect to outlying observations. Devlin, Gnanadesikan, and Kettenring (1975) warn us that R is very sensitive to outliers. They point out that the influence function* for R is unbounded which may yield "potentially catastrophic effects of a small fraction of deviant observations." Their Monte Carlo results provide empirical verification of the nonrobust properties of R. Bias and mean square error increase markedly as the tails of the parent distribution (e.g., Cauchy) become heavier. In many cases the bias is substantial and accounts for 99% of the mean square error; see also Tiku and Balakrishnan (1986). On the other hand, Zellner and Levine (1974), based on extensive simulations, claim that "R is (an) efficient estimator of the population p when the underlying distribution is not normal, as it is when the underlying distribution is normal," and also, when p is high the standard errors are lower for platykurtic distributions than for normal ones. Kowalski (1972) and Duncan and Layard (1973) arrive at contradictory results, so far as the behavior of R from a bivariate exponential distribution is concerned. Earlier investigations by Rider (1932) for uniform distributions, Hey (1938) using data collected in agricultural tests, Nair (1941) for exponential distribu-

*The "influence function" of a statistic may informally be described as an index of the effect on the distribution of the statistic of a single additional observation with value x. It is of course a mathematical function of x. For details, see Hampel (1974) and Huber (1977, p. 9).

tions, and Norris and Hjelm (1961) (normal, uniform, peaked, slightly skewed, and markedly skewed distributions) all seem to conclude that 0.05 and 0.01 quantiles are very close to those under the bivariate normal assumption. Havlicek and Peterson (1977) went even further, claiming that the Pearson R is robust to rather extreme violations of basic assumptions of bivariate normality and type of scale used. Failure to meet the basic assumptions separately or in various combinations had little effect on the resultant distributions of R. For the 216 distributions of R computed in their study, there were no significant deviations from the theoretical expected proportions of R at the 0.005, 0.01, 0.025, or 0.05 quantiles. Thus it was concluded that the effect of violations of these two assumptions has little effect upon the distribution of R and that probability statements in regard to testing the hypothesis that $\rho = 0$ would be accurate. Both Guilford and Frischter (1973) and McNemar (1962) have indicated that some of the basic assumptions, such as normality, do not have to be met. On the other hand, Nunnaly (1967) has maintained that interpretations of an obtained R may not be correct if there are violations of the basic assumptions. Fowler (1987) concludes that R is "remarkably robust" and retains power (as a test of $\rho = 0$) even under extreme violations of distributional assumptions (however, Spearman's R_s may sometimes provide a more powerful test of the null hypothesis $\rho = 0$). In the applied literature, the transformation $t = R\{(n-2)/(1-R^2)\}^{1/2}$ is quite popular and is claimed to be insensitive to violations of the normality assumption when $\rho = 0$ (see also page 582).

When $\rho \neq 0$, the robustness of R has been examined both analytically [Kraemer (1980); see below] and empirically [Kowalski (1972)] with less agreement as to the effects of nonnormality. It should, however, be noted that although asymptotically

$$\frac{\sqrt{n}\,(R-\rho)}{1-\rho^2} \sim N(0,1),$$

this is a very poor approximation unless n is very large. Indeed David (1938) states: "... up to a sample as large as 400 the distribution curves of R from $\rho = 0.0$ to $\rho = 0.6$ (about) are tending only very slowly to normality, while for $n = 400$ and $\rho > 0.6$ there is a very wide divergence from the normal distribution." Kraemer (1980) points out that if the conditional distribution of Y, given X, is normal, with

$$E[Y|X] = \xi_Y + \rho(X - \xi_X)\sigma_Y\sigma_X^{-1},$$

and $\operatorname{var}(Y|X)$ does not depend on X, while the kurtosis $(\beta_2 - 3)$ of the marginal distribution of X is λ, then the transform

$$\sqrt{n-2}\,(R-\rho)\{(1-\rho^2)(1-R^2)\}^{-1/2} \qquad (32.50b)'$$

has approximately a normal distribution with zero expected value and variance $1 + \frac{1}{4}\rho^2\lambda$.

Thus the asymptotics of R will coincide with those of bivariate normal R if and only if $\lim_{n\to\infty} \operatorname{var}[\sqrt{n}\{(S_X/\sigma_X) - \lambda\}] = (\lambda + 2)/4 = \frac{1}{2}$ (i.e., $\lambda = 0$). On the other hand, when $\rho = 0$, the distribution of

$$\frac{(n-2)^{1/2}R}{(1-R^2)^{1/2}},$$

under the same assumptions, as mentioned above, is unconditionally that of t_ν and thus approaches unit normal as $n \to \infty$. [See also Edgell and Noon (1984); Havlicek and Peterson (1977).]

In summary, the null ($\rho = 0$) distribution of R is approximately that of bivariate normal R (at least for large sample size) provided that the linearity and homoscedasticity conditions hold. The nonnull distribution of R is robust provided that an additional kurtosis condition holds.

The convergence of (32.50b)' to its asymptotic distribution has been shown to be quite rapid. Surprisingly, bootstrap estimation of the correlation coefficient, which was one of the central points of Diaconis and Efron's (1983) investigations, seems to receive rather negative assessment in the applied literature [see, e.g., Rasmussen (1987)]. The bootstrap yields overly liberal Type I error rates, and overly restricted confidence intervals. Furthermore it performs as poorly on bivariate normal as on bivariate nonnormal populations. Finally, Silver and Dunlap (1987) emphasize that there is substantial benefit to be gained by transforming correlation coefficients to Fisher's Z' prior to averaging, then backtransforming the average, especially if sample sizes are small.

As Kowalski (1972) pointed out in an extensive historical survey, "A review of the literature revealed an approximately equal dichotomy of opinion. For every study indicating the robustness of the distribution of R, one could cite another claiming to show just the opposite." A survey of subsequent literature reveals a similar state of affairs. One of the main conclusions of Kowalski's investigations is that E. S. Pearson's (1929) claim that "the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of R" should be modified. "The distribution of R in samples from mixtures of bivariate normal distributions may depart considerably from the corresponding normal density even when $\rho_{XY} = 0$ and even for large sample size."

Duncan and Layard (1973) point out that for a bivariate distribution (not necessarily bivariate normal) having finite fourth moments, the distribution of

$$\sqrt{n}\,(\tanh^{-1} R - \tanh^{-1}\rho) = \sqrt{n}\,\{Z'(R) - Z'(\rho)\}$$

converges in distribution to $N(0, \sigma^2(\rho))$, where

$$\sigma^2(\rho) = 1 + \frac{1}{4(1-\rho^2)^2}\{\rho^2(\gamma_{40}+\gamma_{04}) - 4\rho(\gamma_{31}+\gamma_{13}) + 2(2+\rho^2)\gamma_{22}\}, \tag{32.69}$$

and γ_{ij} is the cumulant ratio of order (i, j) of the bivariate distribution (e.g., $\gamma_{22} = \kappa_{22}/(\sigma_X^2\sigma_Y^2)$, where σ_X^2 and σ_Y^2 are the variances of the marginal distributions). If the distribution is bivariate normal, all of the A's vanish and $\sigma^2(\rho) = 1$. Likewise, if the components are independent, p $= 0$, $\gamma_{22} = 0$, and $\sigma^2(0) = 1$. If independence does not hold, the asymptotic variance of $\tanh^{-1}$ R is not generally n^{-1}, whether or not p $= 0$, and normal theory procedures based on Z' $= \tanh^{-1}R$ may not be valid', even asymptotically, for nonnormal bivariate populations. Duncan and Layard (1973), like Kraemer (1973, 1980), pointed out that the asymptotic variance of T $=$ R$\sqrt{(n-2)}/\sqrt{(1-R^2)}$, when p $= 0$, is $1+\gamma_{22}$, which indicates asymptotic *nonrobustness.*

6.2 Point Estimation

Olkin and Pratt (1958) derived the unique minimum variance unbiased estimator of p in the form

$$R^* = R\,{}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}(n-1); 1-R^2\right), \tag{32.70a}$$

where ${}_2F_1(\alpha, \beta; \gamma; x)$ is the Gaussian hypergeometric function [see Chapter 1, Section A6].

Alternative representations are

$$R^* = R\frac{\Gamma((n-1)/2)}{\Gamma(\frac{1}{2})\Gamma((n-2)/2)}\int_0^1 \frac{t^{-1/2}(1-t)^{((n-2)/2)-1}}{[1-t(1-R^2)]^{1/2}}\,dr; \tag{32.70b}$$

$$R^* = R\frac{\Gamma((n-1)/2)}{\Gamma(\frac{1}{2})\Gamma((n-2)/2)}\int_0^1 \frac{t^{-1/2}(1+t)^{-(n-2)/2}}{(1+tR^2)^{1/2}}\,dr. \tag{32.70c}$$

R* is an odd function of R and is strictly increasing. For p $= \pm 1$, R* $=$ R $= \pm 1$ with probability 1. Hence $-1 \le$ R* ≤ 1, which is also the range of p. Values of R* (and also R^*/R) for R $= 0(0.1)1$ and $n = 2(2)30$ were provided by Olkin and Pratt (1958).

These authors suggested use of the approximation

$$R^+ \doteq R\left\{1 + \frac{1-R^2}{2(n-3)}\right\}, \tag{32.71}$$

which is accurate to within ± 0.01 for $n \ge 8$, and within ± 0.001 for $n \ge 18$. Note that n is equal to N (to the number of observations) if the mean of the variables X and Y are known and to N $-$ 1 when they are unknown and are estimated by $\bar{X}$ and $\bar{Y}$. The variance of R* was derived by Iwase (1981) for the equal variance case for even N and for N = 3. He pointed out that in the equal variance case, this estimator is not good for small n and $|\rho|$ but is asymptotically efficient as n increases.

Pradhan and Sathe (1975) observed that, given independent random variables (X_i, Y_i), i = 1,2 from a bivariate normal distribution with cdf $\Phi(x, y; 0, 0; \sigma_X, \sigma_Y; \rho)$,

$$\Pr[X_1Y_1 + X_2Y_2 > 0] = \tfrac{1}{2}(1 + \rho). \tag{32.72}$$

Defining for $1 \leq i < j \leq n$

$$S_{ij} = \begin{cases} 1 & \text{if } X_iY_i + X_jY_j > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $\binom{n}{2}\bar{S} = \Sigma\Sigma_{i<j}S_{ij}$, the estimator

$$\tilde{\rho} = 2\bar{S} - 1 \tag{32.73}$$

is an unbiased estimator of ρ. The authors did not obtain an exact expression for $\text{var}(\tilde{\rho})$ but point out that from U-statistic properties it follows that

$$\text{var}(\tilde{\rho}) \leq \frac{2}{n}(1 - \rho^2). \tag{32.74}$$

When n = 2, the estimator becomes $\tilde{\rho} = \text{sgn}(X_1Y_1 + X_2Y_2)$.

Sibuya (1964) started by observing that if $\sigma_X = \sigma_Y = 1$,

$$\tilde{\rho}' = \tfrac{1}{4}\sqrt{2\pi}\,\{X_1\,\text{sgn}(Y_1) + Y_1\,\text{sgn}(X_1)\} \tag{32.74$'$}$$

is an unbiased estimator of ρ (based on a sample of size 1!) and

$$\text{var}(\tilde{\rho}') = \tfrac{1}{4}\pi + \tfrac{1}{2}\rho\sin^{-1}\rho + \tfrac{1}{2}\sqrt{1 - \rho^2} - \rho^2. \tag{32.75}$$

Using this in an application of Blackwell-Rao theorem, he derived (for $n \geq 1$) the minimum variance unbiased estimator

$$\tilde{\rho}_n = \begin{cases} \sqrt{S}\,J_n(T/S) & \text{if } S > T, \\ -\sqrt{T}\,J_n(S/T) & \text{if } S < T, \end{cases} \tag{32.76}$$

where

$$S = \sum_{i=1}^{n} \left(\frac{X_i + Y_i}{2} \right)^2,$$

$$T = \sum_{i=1}^{n} \left(\frac{X_i - Y_i}{2} \right)^2,$$

$$J_n(z) = \frac{\Gamma^2(n/2)}{\sqrt{2\pi}\,\Gamma((n-1)/2)\Gamma((n+1)/2)}(1-z)$$

$$\times \int_0^1 (1-zt)^{((n-1)/2)-1} t^{(1/2)-1} (1-t)^{((n-1)/2)-1}\, dt.$$

[Note that W $= 2(1+\rho)^{-1}S$ and $V = 2(1-\rho)^{-1}T$ are independent χ_n^2 random variables.] Use of this estimator depends on knowledge of the standard deviations of X and Y, so it is not often of much practical value.

It turns out that the problem of efficient estimation of the correlation coefficient p in the case when population variances are known is a rather delicate one. When all five parameters are unknown, a complete sufficient statistic is

$$\left(\Sigma X_i, \Sigma Y_i, \Sigma X_i^2, \Sigma Y_i^2, \Sigma X_i Y_i\right).$$

If only the population variances are known,

$$\left(\Sigma X_i, \Sigma Y_i, \Sigma\left(X_i^2 + Y_i^2\right), \Sigma X_i Y_i\right)$$

is a complete sufficient statistic, and if the ***only*** unknown parameter is ρ,

$$\left(\Sigma\left(X_i^2 + Y_i^2\right), \Sigma X_i Y_i\right)$$

or equivalently $(\Sigma(X_i + Y_i)^2, \Sigma(X_i - Y_i)^2)$ is sufficient. However, in this case the statistic is not ***complete,*** yielding infinitely many unbiased estimators of ρ based on the sufficient statistic.

The case where ξ, 77, σ_X, and a, are all known was thoroughly investigated by Iwase (1981) and by Iwase and Setô (1984). Iwase (1981) proposed a family of estimators $\tilde{\rho}_{n,k,l}$ indexed by k and l, for the $\Phi(x, y; 0; 0; 1, 1; \rho)$ model, given by

$$\tilde{\rho}_{n,k,l} = \frac{\sqrt{\pi}\, 2^{k-l-2}\Gamma(n/2)}{\Gamma(l+\frac{3}{2})\Gamma[(n/2)+1-k]} \cdot \operatorname{sgn}(S-T) \frac{|S-T|^{2l+1}}{(S+T)^{k+l}}$$

$$\times {}_2F_1\left(\frac{k+l}{2}, \frac{k+l+1}{2}; \frac{n}{2}; \frac{4ST}{(S+T)^2} \right), \qquad (32.77)$$

[Fréchet (1950); Godwin (1964); Mallows (1956, 1963)] are included to assist readers wanting more details.

Note that for $s = 2r$ the Wald inequality becomes one of the Cantelli inequalities. Also for $\mu'_r \le x^r \le \mu'_{2r}/\mu'_r$ there is no suitable Cantelli inequality. A detailed account of these inequalities is given in Godwin (1964), together with further inequalities applicable to sums of independent variables. This latter topic has been further developed by Hoeffding (1963) and Bennett (1968).

Bennett (1968) showed that if $X_1, \dots, X_n$ are independent with finite expected values $E[X_j]$ and variances $[\sigma(X_j)]^2$, and $\Pr[X_j - E[X_j] > M_j] = 0$ for all j, then (with $t < 1$)

$$\Pr\left[\sum_{j=1}^{n} \{X_j - E[X_j]\} \ge t \sum_{j=1}^{n} M_j\right] \le [f(t,r)]^B, \qquad (33.50)$$

where

$$B = \frac{\sum_{j=1}^{n} M_j}{\max_j M_j},$$

$$r = \min_j \left\{\frac{M_j}{\sigma(X_j)}\right\},$$

and

$$f(t,r) = (1 + tr^2)^{-1}\left(\frac{1 + tr^2}{1 - t}\right)^{r^2(1-t)/(1+r^2)}.$$

In Bennett (1968) there is a table of $f(t,r)$ to six decimal places for $t = 0.00(0.02)1.00$, $r = 2.0(0.5)5.0$. Among further properties of IHR and DHR distributions, we note that $(1 - F(x))^{1/x}$ is a decreasing (increasing) function of x if $F(x)$ is IHR (DHR). The following implications (and no others) hold among the eight classes discussed above:

$$\begin{array}{ccccccccc} \text{IFR} & \Rightarrow & \text{IFRA} & \Rightarrow & \text{NBU} & \Rightarrow & \text{NBUFR} & \Rightarrow & \text{NBUFRA} \\ \Downarrow & & & & \Downarrow & & & & \\ \text{DMRL} & & \Longrightarrow & & \text{NBUE} & \Rightarrow & \text{HNBUE,} & & \end{array}$$

[Kochar and Wiens (1987); Bondesson (1983)], also

$$\text{IFR} \Rightarrow \text{IFR} * \text{IFR}$$

($*$ denotes convolution). The exponential distribution belongs to each of the eight classes, with equality in each defining inequality.

Table 33.3 Bounds on distribution functions

Names	Conditions	Values of x	Limits on $1 - F(x)$
Chebyshev	—	$x^r > \mu_r$	$1 - F(x) \le \mu'_r/x^r$
Cantelli	—	$x^r \ge \mu'_{2r}/\mu'_r$	$1 - F(x) \le (\mu'_{2r} - \mu'^2_r)/[(x^r - \mu'_r)^2 + \mu'_{2r} - \mu'^2_r]$
		$x^r \le \mu'_r$	$1 - F(x) \ge 1 - \{(\mu'_{2r} - \mu'^2_r)/[(\mu'_r - x^r)^2 + \mu'_{2r} - \mu'^2_r]\}$
Wald	$r < s$	$1 \ge \mu'_r/x^r \ge \mu'_s/x^s$	$1 - F(x) \le (\mu'_s - \mu'_r\delta^{s-r})/\{x^r(x^{s-r} - \delta^r)\}$, where $\delta(>0, \ne x)$ satisfies $\mu'_r x^s - \mu'_s x^r + \delta^r(\mu'_s - x^s) + \delta^s(\mu'_r - x^r) = 0$
Gauss-Winkler	$F'(x') \ge F'(x) \ge F'(x'')$	$x^r < r^r(r+1)^{-(r-1)}\mu'_r$	$1 - F(x) \le 1 - x[(r+1)\mu'_r]^{-1/r}$
	if $x' < x < x''$	$x^r > r^r(r+1)^{-(r-1)}\mu'_r$	$1 - F(x) \le [(1 + r^{-1})x]^{-r}\mu'_r$

where ${}_2F_1$ is the hypergeometric function and S and T are given by (32.76). The estimator $\tilde{\rho}_{n,k,l}$ does not satisfy the inequality $|\tilde{\rho}_{n,k,l}| \le 1$ if $l = 0$, n is a positive integer and $2k < n + 1$ or if $l > -1$, n is positive integer and $n - 2k = 1$.

Generally

$$E[\tilde{\rho}_{n,k,l}] = \rho\, {}_2F_1\left(-l, k + \frac{1-n}{2}; \frac{3}{2}; \rho^2\right). \tag{32.78}$$

[If ξ and η are unknown, replace S and T by

$$U = \frac{2}{1+\rho}\sum_{i=1}^{n}\left(\frac{(X_i - \bar{X}) + (Y_i - \bar{Y})}{2}\right)^2$$

and

$$V = \frac{2}{1-\rho}\sum_{i=1}^{n}\left(\frac{(X_i - \bar{X}) - (Y_i - \bar{Y})}{2}\right)^2,$$

which are mutually independent χ^2_{n-1} random variables.]

For $k = \frac{1}{2}$, $l = 0$, we have the Sibuya (1964) estimator. For $n = 2$, $k = 1$, $l = 0$, we arrive at Pradhan and Sathe's (1975) estimator. For $k = 0$, $l = 0$, we have $\tilde{\rho}_{n,0,0} = (\sum_{i=1}^{n} X_iY_i)/n$, the conventional estimator which is locally MVUE at $\rho = 0$. For $k = 1$, $l = 0$, we have

$$\tilde{\rho}_{n,1,0} = R'\, {}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}(n-1); 1 - R'^2\right), \tag{32.79}$$

where $R' = 2\Sigma X_iY_i/\Sigma(X_i^2 + Y_i^2)$ is DeLury's (1938) estimator. This is Olkin and Pratt's estimator, (32.70a), with R replaced by R'. As $n \to \infty$, $\tilde{\rho}_{n,k,0} \to \tilde{\rho}^*_{n,k,0} \equiv (S - T)/(S + T)^k$. Iwase (1985) derived an expression for $\text{var}(\tilde{\rho}_{n,k,0})$ and showed that the asymptotic variance is

$$\text{var}(\tilde{\rho}^*_{n,k,0}) \sim \frac{1}{n}\{1 + (1 - 4k + k^2)\rho^2 + k^2\rho^4\} + O(n^{-2}). \tag{32.80}$$

For $k = (1 + 2\rho^2)^{-1}$ the expression simplifies to

$$\text{var}(\tilde{\rho}^*_{n,k,0}) \sim \frac{n^{-1}(1-\rho^2)^2}{1+\rho^2} + O(n^{-2}).$$

The leading term is the Cramér-Rao lower bound.

Note that $\tilde{\rho}_{n,1,0}$ is the UMVUE for ρ, with variance attaining the Cramér–Rao lower bound $(1 - \rho^2)^2/n$ in the $\Phi(0, 0; \sigma^2, \sigma^2; \rho)$ model, which is larger than the bound $n^{-1}(1 - \rho^2)^2(1 + \rho^2)^{-1}$ for the model $(\xi = \eta = 0,$

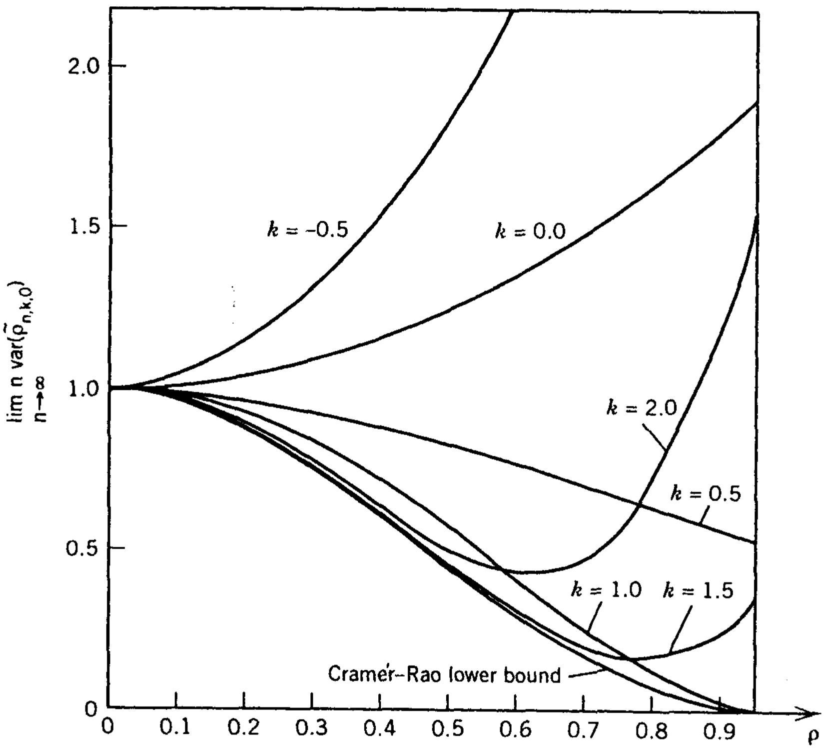


Figure 32.3 $\lim_{n\to\infty} n\,\mathrm{var}(\tilde{\rho}_{n,k,0})$

$\sigma_X = a, = 1$) under consideration. Graphs of $\lim_{n\to\infty}$ n $\mathrm{var}(\tilde{\rho}_{n,k,0})$ against ρ for various values of k are presented in Figure 32.3, indicating that, at least for large ρ, the value k = 1 seems to be optimal.

DeLury (1938) noted that if it is known only that $\xi = \eta$ and $\sigma_X = a,$, then the statistic

$$R' = \frac{2\sum_{i=1}^{n}(X_i - M)(Y_i - M)}{\sum_{i=1}^{n}\{(X_i - M)^2 + (Y_i - M)^2\}}, \tag{32.81a}$$

where $M = \frac{1}{2}\sum_{i=1}^{n}(X_i + Y_i)$, is a slightly better estimator of ρ than is R. If it is only known that $\sigma_X = a,$, DeLury suggested using the estimator

$$R'' = \frac{2\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}\{(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2\}}, \tag{32.81b}$$

which has the pdf

$$p_{R''}(r|\rho) = \{B(\tfrac{1}{2}, \tfrac{1}{2}n)\}^{-1}(1-\rho^2)^{n/2}(1-\rho r)^{-n}(1-r^2)^{(n-2)/2},$$
$$-1 \le r \le 1.$$

Iwase and Setô (1984) generalized Iwase's (1981) results. More precisely, they generalized the unbiased estimator

$$\tilde{\rho}_{n,k,0} = \frac{2^{k-1}\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n+1-k)} \frac{S-T}{(S+T)^k} {}_2F_1\left(\frac{k}{2}, \frac{k+1}{2}; \frac{n}{2}; \frac{4ST}{(S+T)^2}\right), \tag{32.82}$$

for both positive and negative k to

$$\tilde{\rho}_{n,k}(c) = \frac{(1+c)^k}{\Gamma(k)} \int_0^\infty F_n(S,T;u)u^{k-1}e^{-cu}\,du \tag{32.83}$$

(under the condition $n+1 > 2k > 0$ for $c = 0$), where

$$F_n(S,T;u) = 2^{(n-3)/2}\,\Gamma^2\left(\frac{1}{2}n\right)(ST)^{1-(n/4)}u^{-(n-1)/2}$$
$$\times\left\{\frac{1}{\sqrt{T}}J_{n/2}(\sqrt{2uS})J_{(n/2)-1}(\sqrt{2uT})\right.$$
$$\left.-\frac{1}{\sqrt{S}}J_{(n/2)-1}(\sqrt{2uS})J_{n/2}(\sqrt{2uT})\right\}, \tag{32.84}$$

where $J(\cdot)$ is a Bessel function of the first kind. Their aim was to adjust the random variables S and T to $2aS$ and $2bT$, with appropriately chosen a and b [not necessarily $a = (1+\rho)^{-1}$ and $b = (1-\rho)^{-1}$].

It is easy to show that if $c = 0$, the integration can be performed explicitly, yielding $\tilde{\rho}_{n,k,0}$. An alternative form, for $n > 1$ and $c > 0$, is

$$\tilde{\rho}_{n,k}(c) = \frac{1}{n}\left(\frac{1+c}{c}\right)^k (S-T)\int_0^1 (v(1-v))^{(n-3)/2}$$
$$\times {}_2F_1\left(k, n-1; \frac{n}{2}+1; -\frac{1}{2c}\left\{(\sqrt{S}+\sqrt{T})^2 - 4\sqrt{ST}v\right\}\right) dv. \tag{32.85a}$$

For $n = 1$ we have

$$\tilde{\rho}_{1,k}(c) = \frac{1}{2}\left(\frac{1+c}{c}\right)^k XY\left\{\exp\left[-\frac{X^2}{2c}\right]{}_1F_1\left(\frac{3}{2}-k;\frac{3}{2};\frac{X^2}{2c}\right)\right.$$
$$\left.+\exp\left[-\frac{Y^2}{2c}\right]{}_1F_1\left(\frac{3}{2}-k;\frac{3}{2};\frac{Y^2}{2c}\right)\right\}, \qquad (32.85b)$$

and

$$\tilde{\rho}_{1,1/2}(c) = (1+c)^{1/2}\frac{\sqrt{2\pi}}{4}$$
$$\times\left\{X\operatorname{sgn}(Y)\operatorname{erf}\left(\frac{|Y|}{\sqrt{2c}}\right)+Y\operatorname{sgn}(X)\operatorname{erf}\left(\frac{|X|}{\sqrt{2c}}\right)\right\}, \qquad (32.85c)$$

where

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-t^2)\,dt.$$

It is evident that $\tilde{\rho}_{1,1/2}(c)$ approaches the Sibuya estimator $\tilde{\rho}'$ as $c \to 0$. In this sense we can identify $\tilde{\rho}_{1,1/2}(0)$ with $\tilde{\rho}'$.

The authors recommend, for "practical application" (with $n = 1$!) the values $k = \frac{3}{2}$ and $c = 1$, leading to

$$\tilde{\rho}_{1,1.5}(1) = \sqrt{2}\,XY\{e^{-X^2/2}+e^{-Y^2/2}\} \qquad (32.86)$$

with variance

$$\operatorname{var}(\tilde{\rho}_{1,1.5}(1)) = 4(4-2\rho^2+\rho^4)(4-\rho^2)^{-5/2}+4\,3^{-3/2}-\rho^2. \qquad (32.87)$$

The introduction of the parameter c makes it possible to reduce the variance in the case $n = 1$, especially when $\rho_{1,1}(\frac{1}{2})$ is compared with Sibuya's simplified estimator $\tilde{\rho}'$, for all values of ρ. The authors further conjecture that $\tilde{\rho}_{n,k}(c)$ *might* be expected to have small variances when n is greater than 1 and suggest studying the construction of estimators with c a random variable depending on S and T, in the expectation that the variance will be close to the Cramér-Rao lower bound.

6.3 Maximum Likelihood Estimation

The likelihood function for a random sample (X_i, Y_i) $(i = 1, \ldots, n)$ of size n from a bivariate normal distribution (32.2) is

$$l(\mathbf{X}, \mathbf{Y} | \xi, \eta; \sigma_X, \sigma_Y; \rho)$$

$$= (277)^{-n}(1-\rho^2)^{-n/2} \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \sigma_X^{-2} \sum_{i=1}^{n} (X_i - \xi)^2 \right.\right.$$

$$\left.\left. -2\rho\sigma_X^{-1}\sigma_Y^{-1} \sum_{r=1}^{n} (X_i - \xi)(Y_i - \eta) + \sigma_Y^{-2} \sum_{i=1}^{n} (Y_i - \eta)^2 \right\} \right]. \quad (32.88)$$

Equivalently

$$\log l(\mathbf{X}, \mathbf{Y} | \xi, \eta; \sigma_X, \sigma_Y; \rho)$$

$$= -n \log(2\pi) - \tfrac{1}{2} n \log(1-\rho^2) - \tfrac{1}{2}(1-\rho^2)^{-1}$$

$$\times \left\{ \sigma_X^{-2} \sum_{i=1}^{n} (X_i - \xi)^2 - 2\rho\sigma_X^{-1}\sigma_Y^{-1} \sum_{i=1}^{n} (X_i - \xi)(Y_i - \eta) \right.$$

$$\left. + \sigma_Y^{-2} \sum_{i=1}^{n} (Y_i - \eta)^2 \right\}. \quad (32.88)'$$

Solving the maximum likelihood equations

$$\frac{\partial \log \mathrm{I}}{\partial \xi} = \frac{\mathrm{a} \log \mathrm{I}}{\partial \eta} = \frac{\mathrm{a} \log \mathrm{I}}{\partial \sigma_X} = \frac{\mathrm{a} \log \mathrm{I}}{\partial \sigma_Y} = \frac{\mathrm{a} \log \mathrm{I}}{\partial \rho} = 0$$

yields the maximum likelihood estimators

$$\hat{\xi} = \bar{X},$$

$$\hat{\eta} = \bar{Y},$$

$$\hat{\sigma}_X = \left\{ n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right\}^{1/2},$$

$$\hat{\sigma}_Y = \left\{ n^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \right\}^{1/2}, \quad (32.89)$$

and

$$\hat{\rho} = \frac{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\hat{\sigma}_X \hat{\sigma}_Y} = R.$$

But, starting from the pdf of R in the form (32.6b), we obtain the equation

$$\int_0^\infty \frac{R - \hat{\rho}' \cosh x}{(\cosh x - \hat{\rho}' R)^n} \, dx = 0 \tag{32.90}$$

for the corresponding maximum likelihood estimator $\hat{\rho}'$ of R. Of course both $\hat{\rho}$ and $\hat{\rho}'$ have the same asymptotic maximum likelihood properties.

Fisher (1915) showed that from (32.90)

$$\hat{\rho}' \doteqdot R\left(1 + \frac{1 - R^2}{2n}\right)^{-1}. \tag{32.91}$$

Comparison with (32.14a) shows that for p > (<) 0, R has a negative (positive) bias, and consequently $\hat{\rho}'$ will have an even greater negative (positive) bias. If some (or all) of the parameters ξ, η, σ_X, and σ_Y are known, the corresponding maximum likelihood estimators would probably yield better results (at least asymptotically) than $\hat{\rho}$ or $.

The simplest case is when ξ and η are known but not σ_X or a,. Replacing $\bar{X}$ and $\bar{Y}$ in R by ξ and η, respectively, we obtain the MLE

$$\frac{\Sigma_{i=1}^n (X_i - \xi)(Y_i - \eta)}{\left[\left\{\Sigma_{i=1}^n (X_i - \xi)^2\right\}\left\{\Sigma_{i=1}^n (Y_i - \eta)^2\right\}\right]^{1/2}}. \tag{32.92}$$

The associated distribution theory is the same as for R, except that n is replaced by n + 1. If σ_X and a, are known but not ξ or η, the natural estimator

$$R(\sigma_X, \sigma_Y) = \frac{n^{-1}\Sigma_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_X \sigma_Y} \tag{32.93}$$

is unbiased but suffers from the drawback that it can take values outside the range of values of p ($-1 \le$ p ≤ 1). The maximum likelihood equation for a MLE, $\hat{\rho}(\sigma_X, \sigma_Y)$, under these conditions is the real root of the cubic equation

$$\rho^3 - R(\sigma_X, \sigma_Y)\rho^2 + n^{-1}\left\{\sigma_X^{-2} \sum_{i=1}^n (X_i - \bar{X})^2 + \sigma_Y^{-2} \sum_{i=1}^n (Y_i - \bar{Y})^2\right\}\rho$$
$$-R(\sigma_X, \sigma_Y) = 0, \tag{32.94}$$

which maximizes I and lies between -1 and 1 [Kendall (1949); Madansky (1958)]. Madansky (1958) gives a detailed analysis of the conditions for a unique solution of (32.94) satisfying the stated requirements. He shows that asymptotically a unique solution exists with probability 1. He also compares the asymptotic variances of R, $R(\sigma_X, \sigma_Y)$, and $\hat{\rho}(\sigma_X, \sigma_Y)$, which are

$$n \operatorname{var}(R) = (1 - \rho^2)^2,$$

$$n \operatorname{var}(R(\sigma_X, \sigma_Y)) = 1 + \rho^2,$$

$$n \operatorname{var}(\hat{\rho}(\sigma_X, \sigma_Y)) = (1 + \rho^2)^{-1}(1 - \rho^2)^2. \tag{32.95}$$

Therefore the ARE of R is $(1 + \rho^2)^{-1}$ and of $R(\sigma_X, \sigma_Y)$, $(1 - \rho^2)^2(1 + \rho^2)^{-2}$. As in the cases described earlier, if ξ and η are known in addition to σ_X and σ_Y, the appropriate analysis is obtained by replacing $\bar{X}$ and $\bar{Y}$ with ξ and η, respectively. The asymptotic variances and relative efficiencies are unchanged.

Azen and Reed (1973) consider maximum likelihood estimation of p when it is known that $\sigma_X/\xi = \sigma_Y/\eta = c > 0$. (This also implies that $\xi > 0$ and $\eta > 0$.) They present an iterative procedure for calculating the maximum likelihood estimator, $\hat{\rho}(c)$, say. They find that the asymptotic efficiency of R relative to $\hat{\rho}(c)$ is less than 81% for $c < 0.1$ and $|\rho| > 0.5$, although R is almost as efficient as $\hat{\rho}(c)$ for $c > 1$ and $|\rho| < 0.3$, and for large c (≥ 10) the efficiency is practically 100%. The efficiency of R [relative to $\hat{\rho}(c)$] is less for negative p than for the corresponding positive p. Azen and Reed's (1973) general recommendations are: "For samples of size $n = 25$ or larger, $\hat{\rho}(1)$ or $\hat{\rho}$ be used if $c \leq 1$ and it is suspected that $|\rho| \geq 0.5$. This is because, asymptotically, R is at most 80% as efficient as $\hat{\rho}$ for $|\rho| \geq 0.5$. If n is approximately equal to 10, we recommend using $\hat{\rho}(1)$, or $\hat{\rho}$ if $|\rho| \geq 0.7$. On the other hand, for samples as small as $n = 5$, we recommend using Pearson's R. Finally, for $c > 1$ we recommend Pearson's R."

6.4 Estimation of Common ρ Based on Several Samples

Donner and Rosner (1980) compare four different methods of estimating a common correlation coefficient p based on samples for k (≥ 2) bivariate normal populations of sizes $n_1, n_2, \dots, n_k$. In general, since the expected values (ξ_i, η_i) and variances $(\sigma^2_{X_i}, \sigma^2_{Y_i})$ will vary with $i = 1, \dots, k$, it is not valid to use the Pearson product moment sample correlation coefficient based on all n $(= \sum_{i=1}^k n_i)$ pairs (X_{ij}, Y_{ij}) $(i = 1, \dots, k;\ j = 1, \dots, n_i)$ of observed values. The most common procedure is to compute the product-moment coefficients $R_1, \dots, R_k$ for each sample separately, convert to the Z' transform values

$$Z_i' = \frac{1}{2} \log\left(\frac{1 + R_i}{1 - R_i}\right) = \tanh^{-1} R_i,$$

and estimate ρ by

$$\tilde{\rho}_c = \tanh \bar{Z}'_c = \{\exp(2\bar{Z}'_c) - 1\}\{\exp(2\bar{Z}'_c + 1\}^{-1}, \tag{32.96}$$

where

$$\bar{Z}'_c = (n - 3k)^{-1} \sum_{i=1}^{k} (n_i - 3) Z'_i.$$

Explicitly,

$$\tilde{\rho}_c = \frac{\prod_{i=1}^{k}(1 + R_i)^{g_i} - \prod_{i=1}^{k}(1 - R_i)^{g_i}}{\prod_{i=1}^{k}(1 + R_i)^{g_i} + \prod_{i=1}^{k}(1 - R_i)^{g_i}}, \tag{32.96$'$}$$

where g, $= (n - 3k)^{-1}(n_i - 3)$, $i = 1, 2, \ldots, k$.

Donner and Rosner (1980) compare $\tilde{\rho}_c$ with the following estimators:

1. Convert the observed (X_{ij}, Y_{ij}) to standardized values

$$X'_{il} = \frac{X_{I} r - \bar{X}_i}{S_{X_i}},$$

$$Y'_{ij} = \frac{Y_{ij} - \bar{Y}_i}{S_{Y_i}},$$

where $(\bar{X}_i, \bar{Y}_i)$ are the sample means and S_{X_i}, S_{Y_i} the sample standard deviations for the data from the ith population. Then calculate $\tilde{\rho}_S$ as the sample product-moment coefficient for the n pairs (X'_{ij}, Y'_{ij}). Evidently

$$\tilde{\rho}_S = (n - k)^{-1} \sum_{i=1}^{k} (n, - 1) R_i. \tag{32.97}$$

[Donner and Rosner (1980) showed that when $\rho = 0$, $\tilde{\rho}_c$ is a superior estimator (for ρ) than $\tilde{\rho}_S$.]

2. Obtain Pearson's (1933) "maximum likelihood estimator" $\tilde{\rho}_P$ by solving the equation

$$\sum_{i=1}^{k} n_i (R_i - \tilde{\rho}_P)(1 - R_i \tilde{\rho}_P)^{-1} = 0. \tag{32.98}$$

Table 333 Bounds on distribution functions

Names	Conditions	Values of x	Limits on $1 - F(x)$
Chebyshev	—	$x^r > \mu_r$	$1 - F(x) \boldsymbol{/} \mu'_r/x^r$
Cantelli	—	$x^r \geq \mu'_{2r}/\mu'_r$	$1 - F(x)$ $\leq (\mu'_{2r} - \mu'^2_r)/[(x^r - \mu'_r)^2 + \mu'_{2r} - \mu'^2_r]$
		$x^r \leq \mu'_r$	$1 - F(x)$ $\geq 1 - \{(\mu'_{2r} - \mu'^2_r)/[(\mu'_r - x^r)^2 + \mu'_{2r} - \mu'^2_r]\}$
Wald	$r < s$	$1 \geq \mu'_r/x^r \geq \mu'_s/x^s$	$1 - F(x) \boldsymbol{/} (\mu'_s - \mu'_r\delta^{s-r})/\{x^r(x^{s-r} - \delta^r)\}$, where $\delta(> 0, \neq x)$ satisfies $\mu' x^s - \mu' x^r + \delta^r(\mu'_s - x^s) + \delta^s(\mu'_r - x^r) = 0$
Gauss-Winkler	$F'(x') \geq F'(x) \geq F'(x'')$	$x^r < r^r(r + 1)^{-(r-1)}\mu'_r$	$1 - F(x) \boldsymbol{/} 1 - x[(r + 1)\mu'_r]^{-1/r}$
	if $x' < x < x''$	$x^r > r^r(r + 1)^{-(r-1)}\mu'_r$	$1 - F(x) \leq [(1 + r^{-1})x]^{-r}\mu'_r$

3. Replace Z_i' in (32.96) by Z_i'' to obtain Hotelling's (1953) estimator $\tilde{\rho}_{H1}$:

$$Z_i'' = Z_i' - (2n_i - 5)^{-1}R_i \tag{32.99}$$

so that

$$\tilde{\rho}_{H1} = \tanh Z_{H1}', \tag{32.100}$$

with

$$Z_{H1}' = \left(\sum_{i=1}^{k} n_i - \frac{7}{3}k\right)\sum_{i=1}^{k}\left(n_i - \frac{7}{3}\right)Z_i''$$

[since $\operatorname{var}(Z_i'') \simeq (n_i - \frac{7}{3})^{-1}$].

In the special case where $n_1 = n_2 = \cdots = n_k = n_0$, we have

$$\tilde{\rho}_H = \tanh Z_H', \tag{32.101}$$

with

$$Z_H' = \bar{Z}_c' - \left(2n_0 - \tfrac{9}{2}\right)^{-1}\tilde{\rho}_c.$$

Hotelling (1953) [see also Paul (1988)] suggested the further adjustment

$$Z_i''' = Z_i' - \frac{3Z_i' + R_i}{4(n_i - 1)}, \tag{32.102}$$

leading to

$$\tilde{\rho}_{H2} = \tanh Z_{H2}' \tag{32.103}$$

with

$$Z_{H2}' = \left(\sum_{i=1}^{k} n_i - k\right)^{-1}\sum_{i=1}^{k}(n_i - 1)Z_i'''.$$

[The variance of Z_i''' differs from $(n_i - 1)^{-1}$ only by terms of order n_i^{-3} and higher.]

Donner and Rosner (1980) and Paul (1988) studied the performance of these estimators of ρ by means of Monte Carlo simulations, using the "score statistic" $\tilde{\rho}_S$ as the standard of comparison. Paul (1988) carried out a more extensive set of comparisons, using bivariate normal distributions with standardized variables ($\xi_i = \eta_i = 0$; $\sigma_{X_i} = \sigma_{Y_i} = 1$) with $\rho = 0, 0.1(0.2)0.7$, for the following sets of values of n_i:

$k = 2$:

$$n_1 = n_2 = 10, 25, 50,$$
$$n_1 = 10, \quad n_2 = 5, 25.$$

$k = 3$:

$$n_1 = n_2 = n_3 = 10, 25, 50,$$
$$n_1 = n_2 = 10, \quad n_3 = 50,$$
$$n_1 = 10, \quad n_2 = 25, \quad n_3 = 50.$$

$k = 4$:

$$n_1 = n_2 = n_3 = n_4 = 10, 25, 50,$$
$$n_1 = n_2 = 10, \quad n_3 = n_4 = 50$$
$$n_1 = 10, \quad n_2 = 25, \quad n_3 = 50, \quad n_4 = 100.$$

The ratios of mean square errors

$$\frac{\text{MSE}(\tilde{\rho}_S)}{\text{MSE}(\tilde{\rho}_J)}$$

and also relative closeness:

$$\Pr\left[|\tilde{\rho}_S - \rho| - |\tilde{\rho}_J - \rho| > 0\right]$$

for J $= c$, $H1$, $H2$, and P were used as criteria for comparisons. Summarized here are the results:

1. For $\rho < 0.5$, $\tilde{\rho}_{H2}$ performs best; for p > 0.5, $\tilde{\rho}_c$ performs best.
2. Performance of $\tilde{\rho}_{H1}$ is between $\tilde{\rho}_{H2}$ and $\tilde{\rho}_c$.
3. In the neighborhood of $\rho = 0.5$, $\tilde{\rho}_c$, $\tilde{\rho}_{H2}$, and $\tilde{\rho}_S$ perform about equally well.
4. $\tilde{\rho}_P$ is a relatively poor estimator, unless p is small.
5. When sample sizes n_i are equal, the performances of $\tilde{\rho}_c$ and $\tilde{\rho}_P$ are nearly identical.

The ranges of relative efficiencies (inverse MSE) to $\tilde{\rho}_S$, with

$$n_1 = 10, \quad n_2 = 25, \quad n_3 = 50, \quad n_4 = 100,$$

and p increasing from 0 to 0.7 are for $\tilde{\rho}_c$ from 0.97 to 1.09, for $\tilde{\rho}_{H2}$ from 1.01 to 0.97 and for $\tilde{\rho}_{H1}$ from 0.97 to 1.10. The Donner and Rosner (1980) estimator ρ_S is quite good and is simple to calculate.

Mi (1990) uses Kraemer's (1973) suggested approximation,

$$\frac{\sqrt{\nu_i}\,(R_i - \rho)}{\{(1 - R_i^2)(1 - \rho^2)\}^{1/2}} \sim t_{\nu_i},$$

for a suitable value of ν_i (depending on n_i) that implies that R, has an ***approximate*** pdf

$$p_{R_i}(r) = \frac{(1 - \rho^2)^{\nu_i/2}(1 - r^2)^{(\nu_i - 2)/2}}{B(\frac{1}{2}\nu_i, \frac{1}{2})(1 - \rho r)^{\nu_i}}, \qquad -1 \le r \le 1. \quad (32.104)$$

The corresponding likelihood function is proportional to

$$(1 - \rho^2)^{\frac{1}{2}\sum_{i=1}^{R}\nu_i}\left\{\prod_{i=1}^{k}(1 - \rho R_i)\right\}^{-\nu_i}, \qquad (32.105)$$

and the maximum likelihood equation is

$$\sum_{i=1}^{k}\nu_i\left(\frac{\hat{\rho}_A}{1 - \hat{\rho}_A^2} - \frac{R_i}{1 - \hat{\rho}_A R_i}\right) = 0. \qquad (32.106)$$

Mi (1990) recommends using $\nu_i = n_i$ (resulting in an estimator that minimizes $E[(\hat{\rho}_A - \rho)^2]$). The equivalent equation

$$\sum_{i=1}\frac{n_i R_i}{1 - \hat{\rho}_A R_i} = \frac{\hat{\rho}_A}{1 - \hat{\rho}_A^2}\sum_{i=1}^{k} n_i \qquad (32.106)'$$

was given by Hedges and Olkin (1985) and is also equivalent to Pearson's equation (32.98) for $\tilde{\rho}_P$. The mean square error of $\hat{\rho}_A$ is approximately $(1 - \rho^2)^2/\sum_{i=1}^{k} n_i$.

In yet another investigation Viana (1982) studied in detail the three following maximum-likelihood-type estimators of ρ:

1. The only real root of

$$\rho^{*3} + \left(1 + \tfrac{1}{2}\sum_{i=1}^{k} u_i R_i^2\right)\rho^* + \tfrac{1}{2}\sum_{i=1}^{k} v_i R_i = 0, \qquad (32.107)$$

where

$$u_i = \{3 - 2n_i - 9(8n_i - 3)^{-1}(4n_i + 2)^{-1}\}(n - k)^{-1} \quad \left(n = \sum_{i=1}^{k} n_i\right),$$

$$v_i = \{3 - 2n_i - (8n_i + 13)(8n_i - 3)^{-1}(4n_i + 2)^{-1}\}(n - k)^{-1}.$$

This approximation is based on the truncated expansions

$$\rho(1 - \rho^2)^{-1} = \rho + \rho^3 + O(\rho^5), \tag{32.108a}$$

$$R_i(1 - \rho R_i)^{-1} = R_i + \rho R_i^2 + O(\rho^2 R_i^3), \qquad i = 1, \dots, k. \tag{32.108b}$$

2. A linearly combined estimator

$$\tilde{\rho}_L^* = (n - k)^{-1} \sum_{i=1}^{k} (n_i - 1) R_i \left\{1 + \frac{1 - R_i^2}{2(n_i - 4)}\right\}. \tag{32.109}$$

We have $\text{var}(\tilde{\rho}_L^*) \sim (1 - \rho^2)^2 (n - k)^{-1}$ as $n \uparrow \infty$.

3. A doubly transformed linearly combined estimator that is simply $\tilde{\rho}_c$ [Eq. (32.96)].

For $k = 3$, $n \leq 90$, and $n_1 \doteq n_2 = n_3$, $\hat{\rho}^*$ has a smaller mean square error and variance than both $\tilde{\rho}^*$ and $\tilde{\rho}_c^*$ at the expense of a relatively large bias.

We note with regret the marked lack of coordination among the studies by Viana (1982), Paul (1988), and Mi (1990) despite the fact that all three papers appeared in the same journal! Viana does not refer to Donner and Rosner (1980) either, despite similarity in problems and methodology.

6.5 Miscellaneous Estimation Problems

1. Maximum likelihood estimation of ρ from a "broken random sample" was considered by DeGroot and Goel (1980). Specifically each pair of observations (X_i, Y_i) $(i = 1, \dots, n)$ gets "broken" so that all that is available is $(X_1, X_2, \cdots, X_n)$ and $(Y_1', Y_2', \cdots, Y_n')$, where the latter is some (unknown) permutation of $Y_1, Y_2, \cdots, Y_n$. Denoting by $R_{[\theta]}$ the sample correlation coefficient for a specific "pairing" $(X_i, Y_{\theta(i)}')$ $(i = 1, \dots, n)$, where $\theta = (\theta(1), \cdots, \theta(n))$ is a permutation of $(1, \dots, n)$, the maximum likelihood estimator of ρ is

$$\hat{\rho} = \begin{cases} \max_\theta R_{[\theta]} & \text{if } \max_\theta R_{[\theta]} \geq \left|\min_\theta R_{[\theta]}\right|, \\ \min_\theta R_{[\theta]} & \text{if } \max_\theta R_{[\theta]} \leq \left|\min_\theta R_{[\theta]}\right|. \end{cases} \tag{32.110}$$

Unfortunately, this estimator is not reasonable because it is always equal to the maximum or minimum possible sample correlation coefficient that can be calculated from the broken sample. DeGroot and Goel (1980) proposed to use an estimator $\bar{\bar{\rho}}$ which would maximize the "integrated" (average) likelihood defined as

$$(n!)^{-1} \sum_{\theta} \phi(\mathbf{X}, \mathbf{Y} | \xi, \eta; \sigma_X, \sigma_Y; \rho; \theta), \tag{32.111}$$

summation being over all n! permutations θ of $(1, \dots, n)$. The log (likelihood) function is

$$\text{Constant} - \frac{1}{2} n \log(1 - \rho^2) + n(1 - \rho^2)^{-1} + \log\left[\sum_{\theta} \exp\left\{n\rho(1 - \rho^2)^{-1} R_{[\theta]}\right\}\right] \tag{32.112}$$

Equating the derivative with respect to p to zero, we obtain

$$\frac{\Sigma_\theta R_{[\theta]} \exp\left\{n\rho(1 - \rho^2)^{-1} R_{[\theta]}\right\}}{\Sigma_\theta \exp\left\{n\rho(1 - \rho^2)^{-1} R_{[\theta]}\right\}} = \rho. \tag{32.113}$$

This equation has to be solved numerically to obtain $\bar{\bar{\rho}}$.

2. Bayesian estimation of ρ has received relatively little attention in the literature. Gokhale and Press (1982) have mentioned, as possible prior distributions for p, triangular distributions (see Chapter 26, Section 9) with the pdfs

$$p(\rho | \alpha) = \begin{cases} \dfrac{1 - \rho}{1 - \alpha}, & \alpha \le \rho \le 1, \\ \dfrac{1 + \rho}{1 + \alpha}, & -1 \le \rho \le \alpha, \quad \text{with } \alpha^2 < 1, \end{cases} \tag{32.114a}$$

or

$$p(\rho | \alpha = 1) = \tfrac{1}{2}(1 + \rho), \qquad -1 \le \rho \le 1, \tag{32.114b}$$

or

$$p(\rho | \alpha = -1) = \tfrac{1}{2}(1 - \rho), \qquad -1 \le \rho \le 1. \tag{32.114c}$$

They also mention beta priors with the pdfs

$$p(\rho | \alpha, \beta) = 2^{-(\alpha + \beta - 1)} \{B(\alpha, \beta)\}^{-1} (1 + \rho)^{\alpha - 1} (1 - \rho)^{\beta - 1}, \quad -1 \le \rho \le 1; \alpha, \beta > 0. \tag{32.115}$$

However, no justification (or motivation) is presented for choice of these priors.

They do discuss assessing a prior distribution for p, using the ***concordance probability***

$$\tau(\rho) = \Pr[(X_1 < X_2, Y_1 < Y_2) \cup (X_1 > \text{X,, Y,} > Y_2)|\rho] \quad (32.116)$$

or the ***conditional exceedance probability***

$$\psi(\rho) = \Pr[Y_2 > Y_1|x_1, x_2, \rho]. \quad (32.117)$$

7 SAMPLE COVARIANCE

The sample covariance can be defined as

$$n^{-1}\sum_{t=1}^{..}(X_t - \bar{X})(Y_t - \bar{Y}). \quad (32.118)$$

If each of the independent pairs (X_t, Y_t) has a bivariate normal distribution as in (32.2), this is the maximum likelihood estimator of the population covariance $\rho\sigma_X\sigma_Y$. We will find it more convenient to consider the distribution of the sums of products of deviations from sample means:

$$C = \sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \text{Y}). \quad (32.119)$$

The distribution of C has been derived, using various methods, by Mahalanobis, Bose, and Roy (1937), Pearson, Jeffery, and Elderton (1929), Wishart and Bartlett (1932), and Hirschfeld (1937). Remembering that the conditional distribution of Y, given $X_1, X_2, \ldots,$ X,, is normal with expected value $[\eta + (\rho\sigma_Y/\sigma_X)(X_t - \xi)]$ and standard deviation $\sigma_Y\sqrt{1-\rho^2}$, we see that the conditional distribution of C, given $X_1, X_2, \cdots, X_n$ is normal with expected value $(\rho\sigma_Y/\sigma_X)S$ and standard deviation $\sigma_Y\sqrt{(1-\rho^2)S}$, where

$$S = \sum_{t=1}^{n}(\text{X,} - \bar{X})^2.$$

Noting that S is distributed as $\chi^2_{n-1}\sigma^2_X$, it follows that the probability density function of C is

$$p_C(c) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}}\int_0^\infty\left[\frac{S^{(n-4)/2}\exp(-\frac{1}{2}S/\sigma_X^2)}{(2\sigma_X^2)^{(n-1)/2}\Gamma(\frac{1}{2}(n-1))}\right]$$
$$\times\exp\left[-\frac{(c-\rho\sigma_Y S/\sigma_X)^2}{2\sigma_Y^2(1-\rho^2)S}\right]dS. \qquad (32.120)$$

If $\sigma_X = \sigma_Y = 1$, then

$$p_C(c) = \frac{e^{c\rho/(1-\rho^2)}}{2^{n/2}\Gamma(\frac{1}{2}(n-1))\sqrt{\pi(1-\rho^2)}}\int_0^\infty S^{(n-4)/2}\exp\left[-\frac{S+c^2S^{-1}}{2(1-\rho^2)}\right]dS$$
$$= \frac{|c|^{(n/2)-1}e^{c\rho/(1-\rho^2)}}{2^{(n/2)-1}\Gamma(\frac{1}{2}(n-1))\sqrt{\pi(1-\rho^2)}}K_{(n-2)/2}\left(\frac{|c|}{1-\rho^2}\right), \qquad (32.121)$$

where $K_\nu(z)$ denotes the modified Bessel function of the second kind, of order ν (Chapter 12, Section 4). From this formula, distributions of C, with $\sigma_X \neq 1, \sigma_Y \neq 1$, are easily derived. Some other equivalent forms for the density function have been given by Press (1967). He also pointed out that if n is odd, then it is possible to express the density as a finite series of elementary functions, since

$$K_{(n-2)/2}(z) = \sqrt{\frac{\pi}{2z}}e^{-z}\sum_{j=0}^{(n-3)/2}\frac{(\frac{1}{2}(n-3)+j)!}{j!(\frac{1}{2}(n-3-j))!}\cdot\frac{1}{(2z)^j}. \qquad (32.122)$$

The characteristic function of C is

$$\{1 - 2it\rho\sigma_X\sigma_Y + t^2(1-\rho^2)\sigma_X^2\sigma_Y^2\}^{-(n-1)/2}$$
$$= \left[\left\{1 - 2it\frac{\sigma_X\sigma_Y(1+\rho)}{2}\right\}\left\{1 + 2it\frac{\sigma_X\sigma_Y(1-\rho)}{2}\right\}\right]^{-(n-1)/2}. \qquad (32.123)$$

Hence the distribution of C is also that of $\frac{1}{2}\sigma_X\sigma_Y\{(1+\rho)Z_1 - (1-\rho)Z_2\}$, where Z_1 and Z_2 are independent random variables each distributed as χ^2 with n − 1 degrees of freedom. Note that this distribution is the K form of the Bessel function distribution described in Chapter 12, Section 4.

From the representation

$$C = \tfrac{1}{2}\sigma_X\sigma_Y\{(1+\rho)Z_1 - (1-\rho)Z_2\}, \qquad (32.124)$$

it follows that

$$\kappa_r(C) = \left(\tfrac{1}{2}\sigma_X\sigma_Y\right)^r\left\{(1+\rho)^r + (-1)^r(1-\rho)^r\right\}\kappa_r\left(\chi^2_{n-1}\right). \quad (32.125)$$

In particular

$$E(C) = (n-1)\rho\sigma_X\sigma_Y, \quad (32.126a)$$

$$\mathrm{var}(C) = (n-1)(1+\rho^2)\sigma_X^2\sigma_Y^2, \quad (32.126b)$$

$$\alpha_3(C) = \sqrt{\beta_1(C)} = 2\rho(3+\rho^2)(1+\rho^2)^{-3/2}(n-1)^{-1/2}, \quad (32.126c)$$

$$\alpha_4(C) = \beta_2(C) = 3 + 6(1+6\rho^2+\rho^4)(1+\rho^2)^{-2}(n-1)^{-1}. \quad (32.126d)$$

The representation (32.124) also shows clearly how the distribution of C/n tends to normality as n tends to infinity.

8 CIRCULAR SERIAL CORRELATION

In formula (32.1) for the correlation coefficient, X_t and Y_t are usually thought of as representing observations of ***different*** characters on the ***same*** individual. Although this is so in very many applications, it is not formally necessary, and some useful models and techniques can be based on modifications of this idea. In particular, one can take $Y_t = X_{t+k}$, with appropriate allowance for end effects. The correlation coefficients so obtained are called ***serial correlations.*** The absolute value of the difference between t and the subscript of X corresponding to Y_t is called the ***lag*** of the correlation. (It is clear that similar results are obtained by taking $Y_t = X_{t+k}$ or $Y_t = X_{t-k}$.)

Serial correlations are most commonly employed when the subscript t in X_t defines the number of units of time elapsed, from some initial moment, when X_t is observed. However, serial correlations can be, and have been used when the ordering is nontemporal (e.g. spatial).

There is occasionally confusion between serial correlation and ***biserial*** correlation coefficients. The latter are discussed in the (first) edition of ***Multicariate Continuous Distributions.*** Here we just note two additional references [Bedrick (1990, 1992)].

It is of interest to note that the mean square successive difference ratio [von Neumann (1941)],

$$d = \frac{\sum_{t=1}^{n-1}(X_t - X_{t+1})^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}, \quad (32.127)$$

is closely related to a serial correlation R_1 of lag 1. In fact

$$d = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})^2 - 2\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X}) + \sum_{t=1}^{n-1}(X_{t+1} - \bar{X})^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

$$= 2(1 - R_1) - \frac{(X_1 - \bar{X})^2 + (X_n - \bar{X})^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}. \tag{32.128}$$

The simplest way of allowing for end effects is to regard the data as consisting of $(n - k)$ pairs

$$(X_1, X_{1+k}), (X_2, X_{2+k}), \cdots, (X_{n-k}, X_n).$$

Sometimes, however, mathematical analysis is facilitated by using the *circular serial correlation*, based on the n pairs

$$(X_t, X_{t+k}), \qquad t = 1, \dots, n,$$

putting $X_{n+j} = X_j$ for $j = 1, 2, \cdots, k$. Note that in this case the denominator of R_k is simply $\sum_{t=1}^{n}(X_t - \bar{X})^2$. If k is small compared with n, it may be hoped that distributions appropriate to circular correlations may be applied to noncircular correlations with relatively small risk of serious error.

In the case of the mean square successive difference ratio, for example, the circular definition would lead to

$$\tilde{d} = \frac{\sum_{t=1}^{n}(X_t - X_{t+1})^2}{\sum_{t=1}^{n}(X_t - \bar{X})^2}, \tag{32.129}$$

while the circular serial correlation of lag 1 is

$$\tilde{R}_1 = \frac{\sum_{t=1}^{n}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{32.130}$$

(with, of course, $X_{n+1} = X_1$). The relationship between $\tilde{d}$ and $\tilde{R}_1$ is

$$\tilde{d} = 2(1 - \tilde{R}_1) \tag{32.131}$$

which is tidier than that between d and R_1.

We will now consider the distribution of circular serial correlation of lag 1 when the X's are related by equations of form

$$X_t = \rho X_{t-1} + Z_t, \qquad |\rho| < 1, \quad t = 1, 2, \dots, n, \tag{32.132}$$

where the Z_t's are mutually independent normal random variables, each having expected value zero and standard deviation a . Further Z, is independent of all X, for j < t. The series is supposed to be started by a random variable X_0, which is distributed normally with expected value zero, and has standard deviation $\sigma(1-\rho^2)^{-1/2}$. The effect of this rather odd-looking assumption is to ensure that each X, has the same variance (σ^2) as well as the same expected value (zero).

The method used by von Neumann (1941) in deriving the distribution of d (for p = 0) may also be used in deriving the distributions of serial correlations (for $\rho = 0$ or $\rho \neq 0$). The results possess the characteristic property that the density functions take different forms for different intervals in the range of variation of the variable.

The density function of $\tilde{R}_1$ for the case p = 0 was obtained by Anderson (1942) [see also Koopmans (1942)] in the simple form

$$\Pr\left[\tilde{R}_1 > r\right] = \sum_{j=1}^{m} \alpha_j^{-1}(\lambda_j - r)^{(n-3)/2}$$

$$(\text{for } \lambda_{m+1} \leq r \leq \lambda_m,\ m = 1,2,\ldots,n-1) \qquad (32.133)$$

where

$$\lambda_j = \cos\left(\frac{2\pi j}{n}\right),$$

$$\alpha_j = \begin{cases} \prod_{i \neq j,\, i=1}^{(n-1)/2} (\lambda_j - \lambda_i) & (n \text{ odd}), \\ \left[\prod_{i \neq j,\, i=1}^{(n-2)/2} (\lambda_j - \lambda_i)\right]\sqrt{1+\lambda_j} & (n \text{ even}). \end{cases} \qquad (32.133)'$$

Anderson also obtained formulas for the distribution of R, (circular serial correlation with lag I). He also considered the distribution of the statistic

$$\tilde{R}'_1 = \frac{\sum_{j=1}^{n} X_j X_{j+1}}{\sum_{j=1}^{n} X_j^2} \qquad (32.134)$$

(where we use our knowledge that $E[X_j] = 0$). Roughly speaking, the distribution of $\tilde{R}'_1$ is close to that of R,, with n increased by 1. For the case n odd (and $\rho = 0$), Anderson showed that this is exactly so, provided that a, is multiplied by $\sqrt{1-\lambda_j}$. Madow (1945) extended these results to the case $\rho \neq 0$. He used the simple, but effective, device of noting that if

$(T_1, T_2, \cdots, T_S) \equiv \mathbf{T}$ is a sufficient set of statistics for parameter(s) θ, then

$$\frac{p_{\mathbf{T}}(\mathbf{t}|\theta)}{p_{\mathbf{T}}(\mathbf{t}|0)} = \frac{p_{X_1,\cdots,X_n}(x_1,\cdots,x_n|\theta)}{p_{X_1,\cdots,X_n}(x_1,\cdots,x_n|0)}.$$

The joint distribution of T for general ρ can then easily be derived for $\rho = 0$. In this case T is composed of the numerator and denominator of $\tilde{R}'_1$ (or R_1, $\tilde{R}_1$, etc.). The density function of $\tilde{R}_1$ so obtained is

$$p_{\tilde{R}_1}(r) = \tfrac{1}{2}(n-3)\frac{\prod_{j=1}^{n}(1+\rho^2-\lambda_j\rho)^{1/2}}{(1+\rho^2-2\rho r)^{(n/2)-1}}\sum_{j=1}^{m}\alpha_j^{-1}(\lambda_j - r)^{(n-5)/2}$$

$$\text{for } \lambda_{m+1} \le r \le \lambda_m, \quad m = 1,\ldots,(n-1). \qquad (32.135)$$

Daniels (1956) obtained the following expansion for the density of $\tilde{R}'_1$:

$$p_{\tilde{R}_1}(r) = \frac{\Gamma(\frac{1}{2}n+1)(1-\rho^n)}{\sqrt{\pi}(1+\rho^2-2\rho r)^{n/2}}$$

$$\times\left\{\frac{(1-r^2)^{(n-1)/2}}{\Gamma(\frac{1}{2}n+\frac{1}{2})} - \frac{3}{2^n\Gamma(\frac{3}{2}n+\frac{1}{2})}\frac{d^n}{dr^n}(1-r^2)^{(3n-1)/2} + \cdots\right\}. \qquad (32.136)$$

The jth term is

$$(-1)^{j-1}\frac{2j-1}{2^{jn}\Gamma((j-\frac{1}{2})n+\frac{1}{2})}\cdot\frac{d^{jn}}{dr^{jn}}(1-r^2)^{((2j-1)n-1)/2}.$$

Kemp (1970) used the first-order approximation from (32.136),

$$p_{\tilde{R}_1}(r) = (n+1)(1-r^2)^{(n-1)/2}, \qquad -1 \le r \le 1, \qquad (32.137)$$

to obtain approximate expressions for the moments of $\tilde{R}_1$, arriving at

$$E[\tilde{R}_1] \doteq \frac{(n-1)\rho - 1 - n(n-1)(n+3)^{-1}\rho^2}{n(1-\rho)+1+\rho} \qquad (32.138a)$$

and

$$\operatorname{var}(\tilde{R}_1) \doteqdot \left[\frac{\begin{array}{c} n - 2n(n-1)(n+3)^{-1}\rho + 2n(n-1)(n-3)(n+3)^{-1}(n+5)^{-1}\rho^3 \\ (n-1)^2(n-3)(n+3)^{-2}(n+5)^{-1}\rho^4 \end{array}}{\{n(1-\rho)+1+\rho\}^2}\right]$$

(32.138b)

The ***randomization*** $(\rho = 0)$ ***distribution*** of $\tilde{R}_1$ (or $\tilde{R}'_1$) is the distribution obtained by considering all possible orderings of the n observed values $X_1, \dots, X_n$. Since the denominator is unchanged by reordering, this is essentially the distribution of the $n!$ possible values of the numerator. Wald and Wolfowitz *(1943)* studied this distribution.

9 NONCIRCULAR SERIAL CORRELATION

The distributions of noncircular serial correlations are generally even more complicated than those of the corresponding circular correlations. It is possible, however, to obtain formulas for the moments which are reasonably easy to comprehend.

We will first consider the noncircular serial correlation of lag 1, as

$$\hat{R}_1 = \frac{\sum_{j=1}^{n} X_{j-1}X_j}{\sum_{j=1}^{n-1} X_j^2}, \tag{32.139}$$

with $X_0 = 0$, and its distribution under model *(32.132)*. The joint density function of $X_1, \dots, X_n$ is

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = (2\pi\sigma)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \rho x_{j-1})^2\right],$$

$$x_0 = 0. \tag{32.140}$$

The distribution of $\hat{R}_1$ clearly does not depend on σ, and we will henceforth take $\sigma = 1$. The joint moment-generating function of the numerator and denominator of $\hat{R}_1$ is then

$$E\left[\theta_1\sum_{j=1}^{n} X_{j-1}X_j + \theta_2\sum_{j=1}^{n-1} X_j^2\right] = [D_n(\theta_1,\theta_2)]^{-1/2},$$

where for $n \geq 2$,

$$D_n(\theta_1,\theta_2) = (1+\rho^2 - 2\theta_2)D_{n-1}(\theta_1,\theta_2) - (\rho+\theta_1)^2 D_{n-2}(\theta_1,\theta_2), \tag{32.141}$$

with $D_0(\theta_1,\theta_2) = D_1(\theta_1,\theta_2) = 1$.

White *(1957)* evaluated the first and second moments of $\hat{R}_1$, using the formula

$$\mu'_m(\hat{R}_1) = \int_{-\infty}^{0}\int_{-\infty}^{t_m}\cdots\int_{-\infty}^{t_2}\frac{\partial^m\{D_n(\theta_1,t_1)\}^{-1/2}}{\partial\theta_1^m}\bigg|_{\theta_1=0} dt_1\,dt_2\,\cdots\,dt_m. \tag{32.142}$$

Shenton and Johnson *(1965)* found it more convenient to use the formula for moments about ρ ($[\frac{1}{2}m] \equiv$ integer part of $\frac{1}{2}m$):

$$E\left[\left(\hat{R}_1-\rho\right)^m\right] = \sum_{j=0}^{[m/2]} a_j^{(m)}\int_0^\infty\left(\frac{d}{d\rho}\right)^{m-2j} H_{m-j}(t)\,dt, \tag{32.143}$$

where

$$H_r(t) = \frac{t^r}{r!}\{D_n(0,t)\}^{-1/2},$$

where $a_j^{(m)}$ are defined by

$$\sum_{j=0}^{[m/2]}(-1)^j a_j^{(m)} y^{m-2j} = e^{-D^2/2} y^{m*} \quad \text{and} \quad a_0^{(m)} = 1.$$

We now quote some of their results, in the form of asymptotic series

$$\begin{aligned}E\left[\hat{R}_1\right] = \rho &- \frac{2(n-2)\rho}{(n+1)(n-1)} - \frac{12\rho^3}{(n+5)(n+3)(n+1)}\\ &+ \frac{18(n+8)\rho^5}{(n+9)^{\{4\}}} + \frac{24(n+10)(n+12)\rho^7}{(n+13)^{\{5\}}}\\ &+ \frac{30(n+12)(n+14)(n+16)\rho^9}{(n+17)^{\{6\}}} + \cdots,\end{aligned} \tag{32.144a}$$

$$\begin{aligned}E\left[\left(\hat{R}_1-\rho\right)^2\right] = &\frac{n^2-4n+7}{(n+1)^{\{3\}}} - \frac{(n^3-6n^2-25n+42)\rho^2}{(n+5)^{\{4\}}}\\ &+ \frac{3(n+19)(n-3)\rho^4}{(n+9)^{\{5\}}} + \frac{3(n^3+32n^2+111n-928)\rho^6}{(n+13)^{\{6\}}}\\ &+ \frac{3(n+12)(n^3+42n^2+215n-1950)\rho^8}{(n+17)^{\{7\}}} + \cdots,\end{aligned} \tag{32.144b}$$

where $n^{\{s\}} = n(n-2)(n-4)\cdots(n-2\cdot\overline{s-1}) = 2^s(\frac{1}{2}n)^{(s)}$.

*Here D denotes the differential operator described in Chapter 1, Section A4.

For n large the following expansions in powers of n^{-1} can be used:

$$E[\hat{R}_1] = \rho - \frac{2\rho}{n} + \frac{4\rho}{n^2} - \frac{2\rho(1-8\rho^2+4\rho^4)}{n^3(1-\rho^2)^2} \qquad (32.145a)$$

$$+ \frac{4\rho(1-30\rho^2+12\rho^4-4\rho^6)}{n^4(1-\rho^2)^3} - \cdots,$$

$$E[(\hat{R}_1-\rho)^2] = \frac{1-\rho^2}{n} - \frac{1-14\rho^2}{n^2} + \frac{5-78\rho^2+76\rho^4}{n^3(1-\rho^2)}$$

$$+ \frac{11+316\rho^2-692\rho^4+344\rho^6}{n^4(1-\rho^2)^2} + \cdots, \qquad (32.145b)$$

$$\operatorname{var}(\hat{R}_1) = \frac{1-\rho^2}{n} - \frac{1-10\rho^2}{n^2} + \frac{5-62\rho^2+60\rho^4}{n^3(1-\rho^2)}$$

$$+ \frac{11+292\rho^2-596\rho^4+296\rho^6}{n^4(1-\rho^2)^2} + \cdots. \qquad (32.145c)$$

(The original expansions for $E[\hat{R}_1]$ and $E[(\hat{R}_1-\rho)^2]$ were given up to and including terms in n^{-6}.)

Shenton and Johnson (1965) also obtained formulas for β_1 and β_2 and gave results of calculations using these formulas. These are included in Table 32.3. The exact distribution of $\hat{R}_1$ was obtained by Pan (1964) for the case where the correlation between X_i and X_j is p for $|i-j| = 1$ and zero otherwise. In this case $\sum_{j=1}^{n}(X_j-\bar{X})^2$ is distributed as a multiple of χ^2, and it is possible to follow the method of von Neumann (1941). It can be shown that $\hat{R}_1$ is distributed as

$$\frac{\sum_{j=1}^{n-1}\lambda_j U_j^2}{\sum_{j=1}^{n-1}U_j^2},$$

where $U_1, U_2, \cdots, U_{n-1}$ are independent unit normal variables and

$$A, > \lambda_2 > \cdots > \lambda_{n-1},$$

Table 32.3 Moment values for the distribution of $\hat{R}_1$

ρ	n	6	8	10	15	50	100	500
0.0	(b)	0.4254	0.3519	0.3109	0.2530	0.1401	0.0995	0.0447
	(d)	4.7331	3.0022	2.7859	2.7440	2.8919	2.9430	2.9881
0.2	(a)	0.0456	0.0380	0.0323	0.0232	0.0077	0.0039	0.0008
	(b)	0.4244	0.3500	0.3086	0.2502	0.1378	0.0977	0.0438
	(c)	0.0267	0.0418	0.0476	0.0477	0.0241	0.0134	0.0029
	(d)	4.8014	3.0767	2.8625	2.8171	2.9302	2.9648	2.9930
0.4	(a)	0.0902	0.0755	0.0642	0.0463	0.0154	0.0078	0.0016
	(b)	0.4217	0.3445	0.3017	0.2419	0.1304	0.0920	0.0410
	(c)	0.1067	0.1719	0.1999	0.2067	0.1085	0.0610	0.0134
	(d)	5.0031	3.3104	3.1096	3.0613	3.0647	3.0422	3.0103
0.6	(a)	0.1321	0.1114	0.0952	0.0689	0.0230	0.0118	0.0024
	(b)	0.4182	0.3358	0.2901	0.2275	0.1172	0.0815	0.0359
	(c)	0.2392	0.4021	0.4872	0.5372	0.3099	0.1774	0.0394
	(d)	5.3160	3.7301	3.5833	3.5721	3.3839	3.2310	3.0535
08	(a)	0.1674	0.1426	0.1228	0.0900	0.0306	0.0157	0.0032
	(b)	0.4155	0.3256	0.2751	0.2067	0.0958	0.0641	0.0272
	(c)	0.4319	0.7466	0.9563	1.1876	0.9171	0.5318	0.1235
	(d)	5.6832	4.3496	4.3732	4.5969	3.8334	3.8032	3.1926
1.0	(a)	0.1893	0.1599	0.1369	0.0998	0.034[a]	0.019[a]	0.004[a]
	(b)	0.4146	0.3167	0.2603	0.1826	0.06[a]	0.03[a]	0.006[a]
	(c)	0.7456	1.2859	1.7182	2.4705	3.89[a]	5.1[a]	6.5[a]
	(d)	6.0955	5.1633	5.5432	6.5897	9.3[a]	11.3[a]	13.2[a]

Source: Shenton and Johnson (1965).

Key: (a) $E[\hat{R}_1 - \rho]$, (b) standard deviation of $\hat{R}_1$, (c) $\beta_1(\hat{R}_1)$, (d) $\beta_2(\hat{R}_1)$. $E[\hat{R}_1 - \rho]$ and $\beta_1(R_1)$ are zero for p = 0.

[a]The values thus marked are tentative due to uncertain round-off error.

and zero are the characteristic roots of the **matrix MAM** with

$$\mathbf{M} = \begin{pmatrix} 1-n^{-1} & -n^{-1} & \cdots & -n^{-1} \\ n^{-1} & 1-n^{-1} & \cdots & -n^{-1} \\ \vdots & \vdots & & \vdots \\ n^{-1} & n^{-1} & \cdots & 1-n^{-1} \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

The values of the A's are, for odd order A's,

$$\lambda_{2k-1} = \cos\frac{2k\pi}{n+1}, \qquad k = 1, 2, \ldots, \left[\frac{1}{2}n\right],$$

while $\lambda_2, \lambda_4, \cdots, \lambda_{2[(n-1)/2]}$ are roots of the equation

$$(1-\lambda)^{-2}\left\{\left(-\tfrac{1}{2}\right)^n - \tfrac{1}{2}D_{n-1}(\lambda) - (n+1-n\lambda)D_n(\lambda)\right\},$$

with

$$D_n(\lambda) = \left(-\tfrac{1}{2}\lambda\right)^n \sum_{j=1}^{[(n/2)+1]} \binom{n+1}{2j-1}(-1)^{j-1}(\lambda^2-1)^{j-1}.$$

The cumulative distribution function of $\hat{R}_1$ is of different form in each of the intervals $\lambda_{j+1} \le \hat{R}_1 \mathbf{I} \lambda_j$, being

$$F_{\hat{R}_1}(r) = 1 + \pi^{-1}\sum_{i=1}^{[j/2]}(-1)^i\int_{-1}^{1}\frac{(y_{2i}(t) - r)^{(n-3)/2}}{\prod_{\substack{k=1\\k\neq j}}^{n-1}|y_j(t) - \lambda_k|^{1/2}}\frac{dt}{\sqrt{1-t^2}}$$

$$+\frac{1-(-1)^j}{\pi}(-1)^{(j+1)/2}\int_{-1}^{1}\frac{(y_j(t) - r)^{(n/2)-1}}{\prod_{\substack{k=1\\k\neq j}}^{n-1}|y_j(t) - \lambda_k|^{1/2}}\frac{dt}{\sqrt{1-t^2}}$$

for λ_{j+1} $\mathbf{I}$ r ≤ A,. (32.146)

The integrals in this formula can be evaluated approximately by means of the formula (with N sufficiently large)

$$\int_{-1}^{1} f(y)\frac{dy}{\sqrt{1-y^2}} \doteq \pi N^{-1}\sum_{j=1}^{N} f\left(y_j^{(N)}\right), \qquad (32.147)$$

where

$$y_j^{(N)} = \cos\frac{(2j-1)\pi}{2N}.$$

Note that the variables $U_j^2[\Sigma_{j=1}^{n-1}U_j^2]^{-1}$ are correlated beta variables with a joint Dirichlet distribution (see Chapter 40 of the first edition of Continuous Multiuariate Distributions) and so R, is a linear function of such variables. Pan **(1964)** also considered a modified noncircular serial correlation coefficient obtained by dividing a sequence of 2n values $X_1, \cdots, X_{2n}$ into two sets—the first n and the last n values, respectively. The coefficient is defined as

$$R_{1,1} = \frac{\Sigma_{j=1}^{n-1}(X_j - \bar{X}_1)(X_{j+1} - \bar{X}_1) + \Sigma_{j=n+1}^{2n-1}(X_j - \bar{X}_2)(X_{j+1} - \bar{X}_2)}{\Sigma_{j=1}^{n}(X_j - \bar{X}_1)^2 + \Sigma_{j=n+1}^{2n}(X_j - \bar{X}_2)^2}, \tag{32.148}$$

where

$$\bar{X}_1 = n^{-1}\sum_{j=1}^{n} X_j; \quad \bar{X}_2 = n^{-1}\sum_{j=n+1}^{2n} X_j.$$

By essentially the same procedure as before it can be shown that $R_{1,1}$ is distributed as $(\Sigma_{j=1}^{n-1}\lambda_j V_j)/(\Sigma_{j=1}^{n-1}V_j)$, where the V's are mutually independent variables, each distributed as χ^2 with two degrees of freedom. The λ's have the same values as for $\hat{R}_1$. The cumulative distribution function now takes the much simpler form

$$\Pr[R_{1,1} \le r] = 1 - \sum_{i=1}^{j}\left[\prod_{\substack{k=1\\k\neq j}}^{n-1}(\lambda_j - \lambda_k)\right]^{-1}(\lambda_j - r)^{n-2} \qquad \text{for } \lambda_{j+1} \le r \le \lambda_j \tag{32.149}$$

Similar investigations of the distribution of $R_{1,1}$ were performed earlier by Watson and Durbin (1951) (who also computed tables of 5% significant points for their well known and widely used exact noncircular test of the existence of serial correlation in a series of n observations). Watson and Durbin's paper was reproduced in Kotz and Johnson (1991). The asymptotic normality of $\hat{R}_l$ for general l was proved by Pan (1964, 1966). The ***randomization*** $(\rho = 0)$ ***distribution*** of $\hat{R}_1$ was studied by David and Fix (1966).

For independent X_i's with a common normal distribution,

$$\text{var}(R_{1,1}) = n^{-2}(n-1)^{-1}(n-2)^2. \tag{32.150}$$

Despite numerous investigations, by highly skilled researchers, in various fields during the last 20 years, the problem of the behavior of sample serial correlation coefficient under nonnormality has not as yet been fully resolved. This is partly due to the uncoordinated and sporadic nature of this research. More organized comprehensive studies (utilizing modern computational technology) are desirable to put this problem at rest.

We note here a few results relating to the distribution of $R_{1,1}$ in nonnormal populations. In each case the X_i's are assumed to be independent. For

Table 32.4 Distributions of first serial correlation coefficient

	Mean	Standard Deviation	Lower Quantiles				
	$\tilde{\mu}$	$\tilde{\sigma}$	$\tilde{X}_{0.001}$	$\tilde{X}_{0.005}$	$\tilde{X}_{0.010}$	$\tilde{X}_{0.050}$	$\tilde{X}_{0.100}$
Exponential	−0.00207	0.04683	−0.13604	−0.11541	−0.10522	−0.07673	−0.06113
	(0.00002)	(0.00001)	(0.00008)	(0.00005)	(0.00006)	(0.00005)	(0.00004)
			−2.861	−2.420	−2.203	−1.595	−1.261
$\frac{1}{2}$ Weibull	−0.00214	0.04483	−0.10169	−0.08798	−0.08123	−0.06238	−0.05187
$K\chi_2$	(0.00002)	(0.00001)	(0.00016)	(0.00008)	(0.00007)	(0.00003)	(0.00003)
			−2.221	−1.915	−1.764	−1.343	−1.109
	Skewness	**Kurtosis**	**Upper Quantiles**				
	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{X}_{0.900}$	$\tilde{X}_{0.950}$	$\tilde{X}_{0.990}$	$\tilde{X}_{0.995}$	$\tilde{X}_{0.999}$
Exponential	0.17041	0.04871	0.05872	0.07712	0.11281	0.12610	0.15439
	(0.00082)	(0.00219)	(0.00005)	(0.00004)	(0.00007)	(0.00008)	(0.00017)
			1.298	1.691	2.453	2.737	3.341
$\frac{1}{2}$ Weibull	1.02894	2.12262	0.05600	0.08064	0.13588	0.15964	0.21524
$K\chi_2$	(0.00209)	(0.01 146)	(0.00006)	(0.00003)	(0.00014)	(0.00026)	(0.00032)
			1.297	1.847	3.079	3.609	4.849

Note: Estimated quantiles, means, standard deviations and coefficients of skewness and kurtosis of the sample serial correlation coefficient of order 1. Series of $n = 450$ *independent* exponential and $\frac{1}{2}$ Weibull variates. Averaged estimates from six synthetic samples each with 750,000 trials. Quantities in parentheses are estimates of the standard deviation of the estimates. The third line in each group gives the estimated quantiles, standardized by subtracting $\tilde{\mu}$ and dividing by $\tilde{\sigma}$.

Source: Lewis (1972).

exponential distributions, Moran (1967b) obtained

$$\operatorname{var}(R_{1,1}) \doteqdot \frac{1}{n} - \frac{7}{n^2} + \frac{52}{n^3} - \frac{398}{n^4} + \cdots . \tag{32.151}$$

Generally, $n^{1/2}R_{1,1}$ converges in distribution to a unit normal variable, if the first moments of the X, exist. The variance of $R_{1,1}$ tends to be smaller for random variables with long tails (platykurtic) than in the normal case.

Very serious departures from normal theory occur for the quantiles of $R'_{1,1}$ when the X_i's have a common Weibull, distribution (Chapter 21) with parameter c = $\frac{1}{2}$, (i.e., a $K\chi_2$ distribution). In this case Moran's approximation to the variance is poor and the normal approximation is considerably off. Table 32.4, from Lewis (1972), illustrates some results for this case and for exponentially distributed X_i's.

Cox (1966), Moran (1967a, b, 1970), Yang (1970), and Lewis (1972), among others, used simulation methods to investigate the properties of

$$R_1^* = \frac{(n-1)^{-1}\sum_{i=1}^{n-1}(X_i - \bar{X})(X_{i+1} - \bar{X})}{n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \tag{32.139$'$}$$

where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, for the following parent distributions: normal, mixture of two normal, uniform, Laplace, Cauchy, gamma (a, 1) with a = 1.5, 2, 3, 10, 26, Weibull, and exponential for n = 10, 20, as reported in Goodman and Lewis (1972).

10 LEIPNIK DISTRIBUTION

The distributions described in Sections 7 and 8 (dealing with sample covariance and circular serial correlation) are rather complicated in form. An ingenious method of "smoothing" the characteristic function [eliminating the discontinuities in the derivatives of $p_{\tilde{R}_1}(r)$] proposed by Dixon (1944) and Rubin (1945) and extended by Leipnik (1947) leads to the much simpler (approximate) formula

$$p_{\tilde{R}_1}(r) = \left[B\left(\tfrac{1}{2}, \tfrac{1}{2}(n+1)\right)\right]^{-1}(1-r^2)^{(n-1)/2}(1+\rho^2-2\rho r)^{-n/2},$$

$$-1 \le r \le 1. \tag{32.152}$$

Daniels (1956) investigated the error of this approximation and obtained an upper bound for it. Note that if p = 0, we have

$$p_{\tilde{R}_1}(r) = \left[B\left(\tfrac{1}{2}, \tfrac{1}{2}(n+1)\right)\right]^{-1}(1-r^2)^{(n-1)/2},$$

which is the distribution of an ordinary (nonserial) correlation for a bivariate normal population with $\rho = 0$, when the sample size is $n + 3$.

White (1957) pointed out that $(1 + \rho^2 - 2\rho r)^{-n/2}$ can be expressed in terms of Gegenbauer polynomials $c_j^{(n/2)}(x)$ as

$$\sum_{j=0}^{\infty} c_j^{(n/2)}(r)\rho^j, \tag{32.153}$$

where

$$c_j^{(n/2)}(x) = \sum_{m=0}^{[j/2]} \frac{\left(-\frac{1}{2}n\right)^{(j-m)}(-2x)^{j-2m}}{m!(j-2m)!}, \tag{32.154}$$

with $x^{(r)} = x(x-1)\cdots(x-r+1)$, and so

$$\begin{aligned}(-x)^{(r)} &= (-x)(-x-1)\cdots(-x-r+1)\\ &= (-1)^r x(x+1)\cdots(x+r-1) = (-1)^r x^{[r]} \left[\equiv (-1)^r (x)_r\right].\end{aligned}$$

The Gegenbauer polynomials are a special kind of Jacobi polynomials, [see Chapter 1, Eq. (1.175)]. They are orthogonal with respect to the weight function $(1 - x^2)^{(n-1)/2}$ over the interval $-1 \le x \le 1$.

The cumulative distribution function is

$$\begin{aligned}\Pr\left[\tilde{R}_1 \le R_0\right] = I_{(R_0+1)/2}\left(\frac{n+1}{2}, \frac{n+1}{2}\right) - \frac{n}{B\left(\frac{1}{2}, (n+1)/2\right)}\\ \times\left(1 - R_0^2\right)^{(n+1)/2} \sum_{j=1}^{\infty} \frac{\rho^j}{j(n+j)} c_{j-1}^{(\frac{1}{2}n+1)}(R_0).\end{aligned} \tag{32.155}$$

The moment-generating function of $\tilde{R}'_1$ is

$$E[e^{it\tilde{R}_1}] = (2t^{-1})^{n/2}\Gamma\left(\frac{1}{2}n + 1\right) \sum_{j=0}^{\infty} (i\rho)^j \frac{\Gamma(n+j)}{j!\Gamma(n)} J_{(n/2)+j}(t), \tag{32.156}$$

where $J_\nu(t)$ is a Bessel function of order ν [see Eqs. (32.8)].

The rth moment about zero is

$$\mu'_r = \frac{r!}{2^r} \frac{\Gamma((n/2)+1)}{\Gamma(n)} \sum_{j=0}^{[r/2]} \frac{\Gamma(n+r-2j)}{\Gamma\left(\frac{1}{2}n + r - j + 1\right)} \frac{\rho^{r-2j}}{j!(r-2j)!}. \tag{32.157}$$

From (32.157) we find that [Jenkins (1956)]

$$\mu_1' = \frac{n}{n+2}\rho, \tag{32.158a}$$

$$\mu_2 = \frac{1}{n+2} - \frac{n(n-2)\rho^2}{(n+2)^2(n+4)}, \tag{32.158b}$$

$$\mu_3 = \frac{1}{(n+2)^2}\left[-\frac{6n\rho}{n+4} + \frac{2n(n-2)(3n-2)\rho^3}{(n+2)(n+4)(n+6)}\right], \tag{32.158c}$$

$$\mu_4 = \frac{12}{(n+2)(n+4)}\left[1 - \frac{2n(n^2-8n-4)\rho^3}{(n+2)^2(n+6)} + \frac{n(n^4-16n^3+40n^2-32n+16)\rho^4}{(n+2)^3(n+6)(n+8)}\right]. \tag{32.158d}$$

Kemp (1970) derived formulas for the jth moment about 1, and the jth central moment, denoted by ${}_n\mu_j'(1)$ and ${}_n\mu_j$, respectively, of distribution (32.152) in several alternative forms. Among them we note that

$$ {}_n\mu_j = (1-\bar{\rho})(1+\rho)^{-n}F_3\left(\frac{1}{2}(n+1), \frac{1}{2}(n+1), \frac{1}{2}n - j; n+1; \frac{4\rho}{(1+\rho)^2}, \frac{2}{1-\bar{\rho}}\right), \tag{32.159}$$

where $\bar{\rho} = n\rho/(n+1)$ is the expected value,

$$F_3(a,b,c,d;e;x,y) = \sum_{m,n} \frac{(a)_m(b)_n(c)_m(d)_n x^m y^n}{(e)_{m+n} m! n!} \tag{32.160}$$

is Appell's hypergeometric function of the third kind, and

$$ {}_n\mu_j'(1) = \frac{(-2)^j\left(\frac{1}{2}(n+1)\right)_j}{(1+\rho)^n(n+1)_j}\, {}_2F_1\left[\frac{n}{2}, \frac{n+1}{2}; j+n+1; \frac{4\rho}{(1+\rho)^2}\right], \tag{32.161}$$

where ${}_2F_1$ is a Gaussian hypergeometric function (Chapter 1, Section A6).

A recurrence relation for (32.161) is

$$(n+2+2j)\rho\ {}_n\mu_{j+1}'(1) = \left\{(n+j)(1-\rho)^2 - 2(1+2j)\rho\right\}{}_n\mu_j'(1) + (n-1+2j)(1-\rho^2){}_n\mu_{j-1}'(1). \tag{32.162}$$

Some 18 years later McCullagh (1989) rediscovered Leipnik's distribution in a different context, as a noncentral version of the symmetric beta family. He provided plots of the density for various ρ and n. The densities are asymptotically normal as $n \to \infty$ for each fixed ρ.

McCullagh extends the definition of the density to values $|\rho| > 1$ by defining

$$p_X(x;\rho,n) = \frac{(1-x^2)^{(n-1)/2}|\rho|^n}{(1-2\rho x+\rho^2)^{n/2}B\left(\frac{1}{2}(n+1),\frac{1}{2}\right)} \qquad -1 \le x \le 1,\ |\rho| > 1. \tag{32.163}$$

He derived expressions for cumulants of this distribution and also those of a distribution with the pdf

$$p_X(x;\rho,\nu) = \frac{(1-x^2)^{(n-1)/2}(1-\rho)^2}{(1-2\rho x+\rho^2)^{(n/2)+1}B\left(\frac{1}{2}(n+1),\frac{1}{2}\right)}, \qquad -1 \le x \le 1, \tag{32.164}$$

which is related to the Leipnik distribution by

$$X-\rho = \frac{(\tilde{R}_1-\rho)(\rho^2-1)}{1-2\rho\tilde{R}_1+\rho^2}. \tag{32.165}$$

McCullagh (1989) also obtained a pivotal statistic for ρ, namely

$$\tau(\rho) = \frac{1-\tilde{R}_1^2}{1-2\rho\tilde{R}_1+\rho^2}, \tag{32.166}$$

which has a beta $\left(\frac{1}{2}(n+1),\frac{1}{2}\right)$ distribution.

The maximum likelihood estimator, $\hat{\rho}$, of ρ, based on n independent random variables Y_i $(i=1,2,\dots,n)$ with common distribution (32.152), satisfies the equation

$$\sum_{i=1}^{n} \frac{Y_i-\hat{\rho}}{1-2\hat{\rho}Y_i+\hat{\rho}^2} = 0. \tag{32.167}$$

Note that this equation depends on n, only as the number of terms in the summation. Equivalently

$$\sum_{i=1}^{n} \frac{Y_i-\hat{\rho}}{1-Y_i^2+(Y_i-\hat{\rho})^2} = 0. \tag{32.167$'$}$$

Note the similarity to the Cauchy maximum likelihood equation [Chapter 16, Eq. (16.35)]

$$\sum_{i=1}^{n} \frac{Y_i - \hat{\theta}}{\lambda^2 + \left(Y_i - \hat{\theta}\right)^2} = 0, \qquad (32.168)$$

where θ and A are the location and (known) scale parameters, respectively.

Viewing p and n = 2ν as two unknown parameters, **McCullagh** calculates the Fisher information matrix, noting that the Fisher information tends to infinity as $\rho \to \pm 1$. The distributions (32.152) and (32.164) arise as exit distributions of Brownian motion in n + 2 space (for integer n), where a particle starts at the point $\rho = (\rho, 0, \dots, 0)$ on x_1-axis. The probability of hitting the unit sphere is 1 for $-1 < \rho < 1$ and $|\rho|^{-n}$ for $|\rho| > 1$. Saw (1984) encountered Leipnik distributions (also without naming them) in connection with decomposition of densities on unit m spheres. It might be expected that the transformation $\tanh^{-1} \tilde{R}_1$ would produce a more nearly normally distributed variable, as is the case for the ordinary product-moment correlation (32.1). Such a transformation was studied by Quenouille (1948).

However, from (32.158b) it can be seen that

$$\operatorname{var}\left(\tilde{R}_1\right) \doteqdot \frac{1 - \rho^2}{n}, \qquad (32.169)$$

whereas $\operatorname{var}(R) = (1 - \rho^2)^2/n$. Equation (32.169) suggests the use of the transformation

$$\tilde{Z} = \text{sin-'}\ \tilde{R}_1, \qquad (32.170)$$

which would be variance equalizing for $\tilde{R}_1$ in the same sense as is Z' [Eq. (32.42)] for R. Indeed Jenkins (1954a) showed that (32.170) is of comparable effectiveness for R,, as Z' is for R. He obtained the formulas

$$E[\tilde{Z}] = \sin^{-1}\rho - \tfrac{3}{2}\rho(1 - \rho^2)^{-1/2} n^{-1} + \tfrac{1}{8}\rho(17 - 2\rho^2)(1 - \rho^2)^{-3/2} n^{-2} + O(n^{-3}), \qquad (32.171a)$$

$$\operatorname{var}(\tilde{Z}) = n^{-1} - \tfrac{1}{2}(2 - 5\rho^2)(1 - \rho^2)^{-1} n^{-2} + O(n^{-3}), \qquad (32.171b)$$

$$\mu_3(\tilde{Z}) = -3\rho(1 - \rho^2)^{-1/2} n^{-2} + O(n^{-3}), \qquad (32.171c)$$

$$\mu_4(\tilde{Z}) = 3n^{-2} - (8 - 29\rho^2)(1 - \rho^2)^{-1} n^{-3} + O(n^{-4}), \qquad (32.171d)$$

and

$$\sqrt{\beta_1} = -3\rho(1-\rho^2)^{-1/2} n^{-1/2}\{1 + \text{f}(2 - 5\rho^2)(1-\rho^2)^{-} \text{h}^{-1}\} + O(n^{-5/2}), \tag{32.172a}$$

$$\beta_2 = 3 + 2(7\rho^2 - 1)(1-\rho^2)^{-1} n^{-1} + O(n^{-2}). \tag{32.172b}$$

11 MULTIPLE CORRELATION COEFFICIENT

The multiple correlation coefficient between a random variable **X,,** (the dependent variable) and variables $X_1, X_2, \cdots, X_k$ (the independent variables), with $k \geq 2$, is defined to be the maximum correlation between X,, and any linear function of the independent variables:

$$\text{P}_{0.12\ldots k} = \max_{a_1, a_2, \ldots, a_k} \rho\left(X_0, \sum_{j=1}^{k} a_j X_j\right). \tag{32.173}$$

(Where there is no risk of confusion the subscripts **0.12...** k can be omitted.) If the variance-covariance matrix of X_0, **X,,. . .**, X_k is

$$\mathbf{V} = \begin{pmatrix} \text{var}(X_0) & \mathbf{V}_0' \\ \mathbf{V}_0 & \mathbf{V}_{(1)} \end{pmatrix}$$

(where $\mathbf{V}_{(1)}$ is the variance-covariance matrix of $X_1, \cdots, X_k$), then

$$\rho\left(X_0, \sum_{j=1}^{k} a_j X_j\right) = \frac{(\mathbf{V}_0'\mathbf{a})}{\{(\mathbf{a}'\mathbf{V}_{(1)}\mathbf{a})\text{var}(X_0)\}^{1/2}}. \tag{32.174}$$

Since (by appropriate choice of signs of the a_j's) $\rho(\cdot)$ can always arranged not to be negative, we can choose **a** to maximize the square

$$\frac{(\mathbf{a}'\mathbf{V}_0\mathbf{V}_0'\mathbf{a})}{(\mathbf{a}'\mathbf{V}_{(1)}\mathbf{a})}.$$

The maximized value of the square of the correlation coefficient is

$$\frac{\mathbf{V}_0'\mathbf{V}^{-1}\mathbf{V}_0}{\text{var}(X_0)},$$

so the multiple correlation coefficient is

$$\sqrt{\frac{\mathbf{V}_0'\mathbf{V}^{-1}\mathbf{V}_0}{\text{var}(X_0)}}. \tag{32.175}$$

(The more common term for the square is the multiple coefficient of determination, but we will not follow this practice.) Suppose now that $X_0, X_1, \cdots, X_k$ have a joint multinormal distribution and that we have available values of n independent sets of these variables. The quantity $R_{0.12\cdots k}$ obtained by replacing in (32.175) the elements of V by their maximum likelihood estimators (i.e., the mean squares and mean products of deviations from sample means) is called the sample multiple correlation coefficient. This is of course a random variable and has a sampling distribution, though the word "sample" is often omitted from its name when the meaning is otherwise clear. Also, just as $\rho_{0.12\cdots}$, is often replaced by p, so $R_{0.12\cdots}$, is replaced by R when this is conveniently possible.

Following the method of Ruben (1966), already described in Section 3, Hodgson (1967) showed that (with $n > k + 1$), $R^2(1 - R^2)^{-1}$ is distributed as

$$\frac{\chi^2_{k-1} + \left\{U + \rho(1-\rho^2)^{-1/2}\chi_{n-1}\right\}^2}{\chi^2_{n-k-1}}, \tag{32.176}$$

where the χ^2's and the unit normal variable U are mutually independent. From the identity (noting that R cannot be negative)

$$\begin{aligned}\Pr[R \le r] &= \Pr\left[R^2(1-R^2)^{-1} \le r^2(1-r^2)^{-1}\right] \\ &= \Pr\left[\chi^2_{k-1} + \left\{U + \rho(1-\rho^2)^{-1/2}\chi_{n-1}\right\}^2 - r^2(1-r^2)^{-1}\chi^2_{n-k-1} < 0\right]\end{aligned} \tag{32.177}$$

and approximating the distribution of the ratio (32.176) or the left-hand side of the last inequality in (32.177) in the braces, Hodgson suggested the following approximations (h being a suitable positive number):

$$\frac{(n-k+h-2)^h\left[R^2(1-R^2)^{-1}\right]^h - \left\{k+2h-2+(n+h-2)\rho^2(1-\rho^2)^{-1}\right\}^h}{\sqrt{2}h\left[(n-k-1)^{2h-1}\left(\dfrac{R^2}{1-R^2}\right)^{2h} + \left(2+\dfrac{\rho^2}{1+\rho^2}\right)\left\{k+\dfrac{(n-1)\rho^2}{1-\rho^2}\right\}^{2h-1}\right]^{1/2}} \tag{32.177)'$$

has approximately a unit normal distribution, and

$$\left\{R^2(1-R^2)^{-1}\right\}^h \tag{32.177)''$$

is approximately normally distributed with expected value

$$[(n-k+h-2)^{-1}\left\{k+2h-2+(n+h-2)\left[\rho^2(1-\rho^2)^{-1}\right\}\right]^h$$

and variance

$$2h^2(n-k-1)^{-(2h+1)}\left\{k+(n-1)\rho^2(1-\rho^2)^{-1}\right\}^{2h-1}(2n-k)(1-\rho^2)^{-1}.$$

The distribution of R^2 was originally obtained by Fisher (1928), using a geometrical method [see also Soper (1929)]. [Special cases had been discussed earlier by Yule (1921) and Isserlis (1917).] Fisher obtained the formula

$$p_{R^2}(r^2) = \frac{\Gamma(\frac{1}{2}n)(1-\rho^2)^{(n-1)/2}}{\pi\Gamma(\frac{1}{2}(k-1))\Gamma(\frac{1}{2}(n-k))}(r^2)^{(k/2)-1}(1-r^2)^{((n-k)/2)-1}$$
$$\times\int_0^{\pi}\int_{-\infty}^{\infty}\frac{\sin^{k-2}\theta}{(\cosh\phi-\rho r\cos\theta)^{n-1}}\,d\phi\,d\theta, \qquad 0<r^2<1. \tag{32.178}$$

The integral can be evaluated by expanding the integrand in powers of cos 8 and integrating term by term. The result can be conveniently expressed in terms of a hypergeometric function

$$p_{R^2}(r^2) = \frac{(1-\rho^2)^{(n-1)/2}(r^2)^{(k/2)-1}(1-r^2)^{((n-k-1)/2)-1}}{B(\frac{1}{2}k,\frac{1}{2}(n-k-1))}$$
$$\times{}_2F_1\left(\frac{n-1}{2},\frac{n-1}{2};\frac{k}{2};\rho^2r^2\right), \qquad 0<r^2<1. \tag{32.179}$$

An alternative expression is

$$p_{R^2}(r^2) = \frac{(1-r^2)^{(n-k-3)/2}(r^2)^{(k-2)/2}(1-\rho^2)^{(n-1)/2}}{\Gamma\{(n-1)/2\}\Gamma\{(n-k-1)/2\}}$$
$$\times\sum_{i=0}^{\infty}\frac{(\rho^2)^i(r^2)^i[\Gamma\{((n-1)/2)+i\}]^2}{i!\Gamma\{(k/2)+i\}}, \qquad 0<r^2<1. \tag{32.179$'$}$$

[This expression was given by Anderson (1984), and used by Ding and Bargmann (1991a) as a basis for a numerical method, and an efficient computer program (Algorithm AS260) for evaluation of the distribution of R^2.]

The null ($\rho^2 = 0$) pdf of R^2 is

$$p_{R^2}(r^2) = \frac{(1-r^2)^{(n-k-3)/2}(r^2)^{(k-2)/2}}{B[(n-k-1)/2, k/2]}, \qquad 0 < r^2 < 1. \quad (32.180)$$

This is a standard beta distribution with parameters $fk, \frac{1}{2}(n-k-1)$.

From the representation (32.179)', or by direct expansion, it is possible to express $p_{R^2}(r^2)$ as a mixture of standard beta distributions with parameters $(\frac{1}{2}k + j), \frac{1}{2}(n-k-1)$, with weights being the terms in the expansion of the negative binomial

$$\left(\frac{1}{1-\rho^2} - \frac{\rho^2}{1-\rho^2}\right)^{-(n-1)/2}$$

That is,

$$p_{R^2}(r^2) = \sum_{j=0}^{\infty} b_j\left[B\left(\tfrac{1}{2}k + j, \tfrac{1}{2}(n-k-1)\right)\right]^{-1}(r^2)^{(k/2)+j-1}(1-r^2)^{((n-k-1)/2)-1}$$

$$r^2 > 0, \quad (32.181)$$

with weights

$$b_j = \frac{\Gamma\left(\frac{1}{2}(n-1)+j\right)}{j!\Gamma\left(\frac{1}{2}(n-1)\right)} \frac{(\rho^2)^j}{(1-\rho^2)^{((n-1)/2)+j}}.$$

This result was obtained by Gurland (1968), using characteristic functions. He also showed that for $n - k$ odd, $R^2(1-\rho^2)(1-p^2R^2)^{-1}$ is distributed as a mixture of standard beta distributions with parameters $(\frac{1}{2}k + j), \frac{1}{2}(n - k - 1)$ and weights given by the terms in the binomial expansion of $\{\rho^2 + (1-\rho^2)\}^{(n-k-1)/2}$ so that

$$\Pr[R \le r] = \sum_{j=0}^{(n-k-1)/2} \binom{\frac{1}{2}(n-k-1)}{j} (\rho^2)^j$$
$$\times (1-\rho^2)^{((n-k-1)/2)-j} I_{R^2(1-\rho^2)/(1-\rho^2R^2)}\left(\tfrac{1}{2}k + j, \tfrac{1}{2}(n-k-1)\right),$$
$$(32.182)$$

where $I_x(a,b)$ is the incomplete beta ratio. Note that there is only a finite number of terms in this expansion. Various methods of derivation of the distribution will be found in Garding (1941), Moran (1950), Soper (1929), and

Wilks (1932). Williams (1978) obtained (32.182) by another method, essentially showing that conditionally on sample values of $(X_1, \cdots, X_k)$, $(R^2/k)\{(1-R^2)/(n-k-1)\}^{-1}$ (the ratio of mean square due to multiple linear regression to residual mean square in the analysis of variance) has a noncentral F-distribution (see Chapter 27) and then averaging over the $(X_2, \cdots, X_k)$ distribution (cf. beginning of Section 2). Detailed numerical investigations of series (32.181) and (32.182) are reported in Gurland and Milton (1970).

In a more recent paper Gurland and Asiribo (1991) provided other representations of the distribution of R^2, utilizing Gurland's (1968) result that $W = R^2(1-R^2)^{-1}$ is distributed as the ratio Y_1/Y_2 of mutually independent random variables with characteristic functions

$$\phi_{Y_1}(t) = (1-2it)^{g}(1-2iat)^{-h},$$

$$\phi_{Y_2}(t) = (1-2it)^{-g}, \tag{32.183}$$

where $g = \frac{1}{2}(n-k)$, $h = \frac{1}{2}(n-1)$, $a = (1-\rho^2)^{-1}$. They concentrated on the distribution of Y, [termed a "kinked" χ^2 by Gurland and Milton (1970)] and offered several alternative representations for the cdf and pdf of Y, in terms of scaled χ^2-distributions and confluent hypergeometric distributions, implying alternative representations of the distribution of R^2.

The mth moment of R about zero can be expressed in a convenient form, due to Banerjee (1952),

$$\mu'_m(R) = \frac{(1-\rho^2)^{(n-1)/2}\Gamma(\frac{1}{2}(k+m))}{\Gamma(\frac{1}{2}(n-1+m))} D'^m {}_2F_1\left(\frac{1}{2}(n-1+m), \frac{1}{2}n; \frac{1}{2}k; \rho^2\right), \tag{32.184}$$

where D' denotes the operator $\frac{1}{2}\rho^3(\partial/\partial\rho)$. The expected value and variance of R^2 are

$$E[R^2] = 1 - \frac{n-k-1}{n-1}(1-\rho^2)\,{}_2F_1\left(1,1;\frac{1}{2}(n+1);\rho^2\right)$$

$$= \rho^2 + \frac{k}{n-1}(1-\rho^2) - \frac{2(n-k-1)}{n^2-1}\rho^2(1-\rho^2) + O(n^{-2}), \tag{32.185a}$$

(note that R^2, as an estimator of ρ^2, has a substantial positive bias), and

$$\operatorname{var}(R^2) = \frac{(n-k)^2 - 1}{n^2 - 1}(1-\rho^2)^2 {}_2F_1\left(2,2;\frac{1}{2}(n+3);\rho^2\right) - \left[E[R^2] - 1\right]^2$$

$$= \begin{cases} \dfrac{4\rho^2(1-\rho^2)^2(n-k-1)^2}{(n^2-1)(n+3)} + O(n^{-2}) & \text{for } \rho \neq 0, \\ \dfrac{2k(n-k-1)}{(n-1)^2(n+1)} & \text{for } \rho = 0 \end{cases} \tag{32.185b}$$

[Wishart (1931)].

The distribution of R (or R^2) is complicated in form, and considerable attention has been devoted to the construction of useful approximations. It is natural to try Fisher's transformation $Z' = \frac{1}{2}\log\{(1+R)/(1-R)\} = \tanh^{-1} R$. However, as for serial correlation, it is clear that this transformation is not very suitable. Gajjar (1967) has shown that the limiting distribution of $\sqrt{n-1}\tanh^{-1} R$, as n tends to infinity, is not normal but is noncentral χ with k degrees of freedom and noncentrality parameter $(n-1)(\tanh^{-1}\rho)^2$. Numerical calculations indicate that $\tanh^{-1} R$ will not give generally useful results.

Khatri (1966) proposed two approximations. The first is to regard

$$\frac{k^{-1}(n-k-1)(1-\rho^2)\omega(\rho)R^2}{1-R^2} \tag{32.186a}$$

with

$$\omega(\rho) = \left[k + \left\{n-k-1+\sqrt{(n-1)(n-k-1)}\right\}\rho^2\right] \times \left[k + (n-k-1)(2-\rho^2)\rho^2\right]^{-1}$$

as being approximately distributed as noncentral F with k, $(n-k-1)$ degrees of freedom and noncentrality parameter

$$\tfrac{1}{2}\rho^2\omega(\rho)\sqrt{(n-1)(n-k-1)}$$

[cf. Williams (1978)]. The second of Khatri's approximations uses a different multiple of $R^2/(1 - R^2)$. The multiplier is

$$(n - k - 1)(1 - \rho^2)\{(n - k - 1)\rho^2 + k\}^{-1}$$

and the approximate distribution is a ***central*** F-distribution with

$$[(n - k - 1)\rho^2 + k]^2[(n - k - 1)\rho^2(2 - \rho^2) + k]^{-1}, \; n - k - 1 \tag{32.186b}$$

degrees of freedom. Independently Gurland (1968) also obtained this approximation.

Khatri suggested that these approximations be used for $n - k - 1 \geq 100$; he preferred the second when ρ^2 is large. Gurland and Asiribo (1991) provided a further approximation to the distribution by applying the Wilson-Hilferty transformation of a χ^2 random variable (Chapter 18). This yields an approximation to the distribution of $U = R^2/(1 - R)^2$ in terms of a normal distribution. Comparisons with the approximation (32.186b) of Khatri (1966) and Gurland (1968) for $k = 6, 10$, $n = 10, 20, 40$, $p = 0, 0.1(0.2)0.9$, and also for values $\rho^2 = 0.1(0.2)0.9$ are presented. They indicate that for the most part, the approximations are of about the same accuracy.

Moschopoulos and Mudholkar (1983) provide a normal approximation to the distribution of R^2 that seems to be very accurate for $p > \frac{1}{2}$, with errors in the cdf only in the fourth decimal place. For smaller values of p, errors may appear in the third decimal place.

The approximation is of form

$$\left\{\frac{-\log(1 - R^2)}{g_1}\right\}^h,$$

which is approximately normal with expected value and variance (and also h and g_1) given by rather complicated functions of n, k, and p.

Srivastava (1983), Gupta and Kabe (1991), and Amey (1990) have discussed the distribution of R^2 based on random samples from a mixture of m $(k + 1)$-variate multinormal populations, and they have obtained expressions for the exact distribution. Srivastava and Gupta and Kabe deal with the case m = 2, whereas Amey deals with the general case.

Amey (1990) programmed this cdf for the case $p = 0$ using IMSL (1987) subroutines and obtained upper percentage points of R^2 for selected values of the parameters. He tabulated the values of R^2_α, where $\Pr[R^2 \leq R^2_\alpha | \rho = 0] = a$ for $a = 0.90, 0.95, 0.99$, $k = 2, 3, 4, 5$, $n = 10, 15, 20$, and for $k = 2, 3, 4, 5$ and $n = 10, 15, 20$ with selected values of differences among the expected values of component distributions.

It can be shown that $\Pr[R \leq r|\rho]$ is a decreasing function of p. Upper and lower limits for $100(1 - \alpha)\%$ confidence intervals for p can be obtained by solving the equations $\int_{-\infty}^{R} p_R(r|\rho)\,dr = a,, 1 - a,,$ with $(a, + \alpha_2) = a$.

Kramer (1963) gave tables for constructing lower 95% limits (i.e., $\alpha_1 = 0.05$, $\alpha_2 = 0$).

R^2 is a biased estimator of ρ^2, as can be seen from (32.185a). Venables (1985) discussed Fisher's (1924) A-transformation of R^2, which is aimed at reducing this bias. It is defined as

$$A = \frac{R^2 - k(n-1)^{-1}}{1 - k(n-1)^{-1}}, \tag{32.187}$$

or, equivalently,

$$1 - A = (n-1)(n-k-1)^{-1}(1 - R^2). \tag{32.187$'$}$$

Venables showed that A is very close to the maximum likelihood estimator of ρ^2, based on marginal likelihood calculated from (32.179). Specifically he proved that (32.179) has a unique maximum value in the parameter space $0 \le \rho^2 \le 1$ at $\hat{\rho}$, say, and that if $R^2 \le k(n-1)^{-1}$, $\hat{\rho}^2 = 0$; on the other hand, if $1 > R^2 > k(n-1)^{-1}$, Venables showed the maximum likelihood estimator to be at the unique maximum of the marginal likelihood function in the range $0 < \rho^2 < 1$. What this implies is that for $k(n-1)^{-1} < R^2 < 1$, the maximum marginal likelihood estimator $\hat{\rho}^2$ satisfies the inequalities

$$A < \hat{\rho}^2 < A + \frac{2(1-A)^2}{n + 2(1-A)}. \tag{32.188}$$

Note, however, that $\hat{\rho}^2$, being a nonnegative estimator, cannot be exactly unbiased, although it has a substantially smaller bias than R^2. Venables (1985) also provided an expansion for $\hat{\rho}^2$ of the type

$$1 - \hat{\rho}^2 = \frac{n-1}{n-1-k}(1 - R^2) + \sum_{j=2}^{\infty} b_j (1 - R^2)^j, \tag{32.189}$$

thus extending the A statistic. Here

$$b_2 = \frac{2(n-1)}{[2(n-1) - k + 2](n-1-k)},$$

$$b_3 = \frac{2(n-1)\{-4k(k+2) + (n-1)(k^2 + 10k + 8) - 2(n-1)^2(k+2)\}}{[2(n-1) - k - 2]^2[2(n-1) - k - 4](n-1-k)^2}.$$

He justifiably pointed out that for higher j the coefficients are much more complicated and probably of doubtful numerical usefulness.

Olkin and Pratt (1958) showed that

$$R^{*2} = 1 - (n-k)^{-1}(n-2)(1-R^2)\,{}_2F_1\left(1,1;\tfrac{1}{2}(n-k+2);(1-R^2)\right) \tag{32.190}$$

is the unique minimum variance unbiased estimator of ρ^2. It differs from R^2 only by quantities of order n^{-1} and is a strictly increasing function of R^2. Note that if $R^2 = 1$, then $R^{*2} = 1$; however, if $R^2 = 0$, then $R^{*2} = -k(n-k+2)^{-1}$ which is negative, as must be the case since $E[R^{*2}|\rho = 0] = 0$.

Sylvan's (1969) MLE of the multiple correlation coefficient when some observations of one variable, say, the last one, x_k, are missing is constructed as follows: Let

$$\mathbf{x}' = (x_1, \cdots, x_k),$$
$$\mathbf{x}^{*\prime} = (x_1, \cdots, x_{k-1}),$$
$$\mathbf{x}'_+ = (x_1, \cdots, x_k, X_0),$$
$$\mathbf{x}^{*\prime}_+ = (x_1, \cdots, x_{k-1}, X_0).$$

The multiple correlation coefficient is

$$\rho = \sqrt{\frac{\mathbf{V}_0'\mathbf{V}^{-1}\mathbf{V}_0}{\operatorname{var}(X_0)}}. \tag{32.175$'$}$$

Hence

$$1 - \rho^2 = \frac{\operatorname{var}(X_0) - \mathbf{V}_0'\mathbf{V}^{-1}\mathbf{V}_0}{\operatorname{var}(X_0)}. \tag{32.191}$$

We partition the ocerall sample variance-covariance matrix Σ into

$$\Sigma = \begin{pmatrix} \Sigma_{x^*x^*} & \Sigma_{x^*x_k} & \Sigma_{x^*x_0} \\ \Sigma_{x_kx^*} & \Sigma_{x_kx_k} & \Sigma_{x_kx_0} \\ \Sigma_{x_0x^*} & \Sigma_{x_0x_k} & \Sigma_{x_0x_0} \end{pmatrix} = \left(\begin{array}{c|c} \Sigma_{\underline{x}\underline{x}} & \Sigma_{\underline{x}x_0} \\ \hline \Sigma_{x_0\underline{x}} & \Sigma_{x_0x_0} \end{array}\right). \tag{32.192}$$

Assume that x_p has c missing observations. Let the matrix of the corrected sums of squares (ss) and products (sp) of observations including all of x's and x_0 be

$$\mathbf{S}_{\mathbf{x}_+\mathbf{x}_+} = \left(\begin{array}{ccc} S_{\underline{x}^*\underline{x}^*} & S_{\underline{x}^*x_k} & S_{\underline{x}^*x_0} \\ S_{x_k\underline{x}^*} & S_{x_kx_k} & S_{x_kx_0} \\ \hline S_{x_0\underline{x}^*} & S_{x_0x_k} & S_{x_0x_0} \end{array}\right) = \left(\begin{array}{c|c} S_{\underline{x}\underline{x}} & S_{\underline{x}x_0} \\ \hline S_{x_0\underline{x}} & S_{x_0x_0} \end{array}\right) \tag{32.193}$$

based on n − c degrees of freedom. The matrix of corrected sum of squares and sum of products of all observations on $\underline{x}^*$ and x_0 is

$$\mathbf{T}_{kk} = \begin{pmatrix} T_{\underline{x}^*\underline{x}^*} & T_{\underline{x}^*x_0} \\ T_{x_0\underline{x}^*} & T_{x_0x_0} \end{pmatrix} \tag{32.194}$$

based on n degrees of freedom.

To obtain Sylvan's (1969) estimator, the estimators

$$\hat{\Sigma}_{x_pk} = \frac{1}{n}T_{kk},$$

$$\hat{\Sigma}_{x_pk} = \frac{1}{n}S_{x_pk}S_{kk}^{-1}T_{kk},$$

$$\hat{\Sigma}_{x_px_p} = \frac{1}{n-c}S_{x_px_p\cdot k} + \frac{1}{n}S_{x_pk}S_{kk}^{-1}T_{kk}S_{kk}^{-1}S_{kx_p}, \tag{32.195}$$

where

$$S_{x_px_p\cdot k} = S_{x_px_p} - S_{x_pk}S_{kk}^{-1}S_{kx_p},$$

$$\mathbf{S}_{kk}^{-1} = \begin{pmatrix} S_{x^*x^*}^{-1} + \dfrac{S_{x^*x^*}^{-1}S_{x^*x_0}S_{x_0x^*}S_{x^*x^*}^{-1}}{S_{x_0x_0\cdot x^*}} & -\dfrac{S_{x^*x^*}^{-1}S_{x^*x_0}}{S_{x_0x_0\cdot x^*}} \\ -\dfrac{S_{y^*x^*}S_{x^*x^*}^{-1}}{S_{x_0x_0\cdot x^*}} & \dfrac{1}{S_{x_0x_0\cdot x^*}} \end{pmatrix},$$

are substituted into (32.175) or (32.175)'.

Gupta (1987) extended Sylvan's result and obtained somewhat simpler estimator $\tilde{R}^2$ when c observations on a variable are missing. To this end, Gupta (1987) recalled the identity

$$1 - R^2 = \left(1 - R^2_{\underline{x}^*}\right)\cdot\left(1 - R^2_{x_k\cdot x^*}\right), \tag{32.196}$$

where the first factor is the contribution of $\underline{x}^*$ to $1 - R^2$, while the second factor is the contribution of x_k conditional on $\underline{x}^*$. Noting that $1 - R^2$ is a particular case of Wilks's Λ [cf. Eq. (32.174)], he obtained the somewhat simpler estimator:

$$1 - \tilde{R}^2 = \left(\frac{T_{x_0x_0\cdot x^*}}{T_{x_0x_0}}\right)\cdot\left(\frac{S_{x_0x_0\cdot x}}{S_{x_0x_0\cdot x^*}}\right)^{(n-c)/n}. \tag{32.197}$$

It appears that the true value of ρ^2 is overestimated by Sylvan's (1969) method but underestimated by Gupta's (1987) method. It was further noted by Gupta, based on the general theory of Wilks's Λ statistics, that

$$-\left\{n - \frac{(k^2 - 1)}{2k} - \frac{n(2k + 1)}{2k(n - c)}\right\}\log_e(1 - \tilde{R}^2), \qquad (32.198)$$

where c is the number of observations missing, has an asymptotic χ^2 distribution with k degrees of freedom.

REFERENCES

Ali, M. A. (1990). Bounds for the mean of the sample coefficient of determination, *Journal of Statistical Research,* **24,** 27–30.

Allan, F. E. (1930). A percentile table of the relation between the true and the observed correlation coefficient from a sample of size 4, *Proceedings of the Cambridge Philosophical Society, 26,* 536–537.

Amey, A. A. K. (1990). Robustness of the multiple correlation coefficient when sampling from a mixture of two multivariate normal populations, *Communications in Statistics—Simulation and Computation,* **19,** 1443–1457.

Anderson, R. L. (1942). Distribution of the serial correlation coefficient, *Annals of Mathematical Statistics,* **13,** 1-13.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis,* 2nd ed., New York: Wiley.

Armsen, P. (1956). Note on a simple identity concerned with particular values of the distribution function of correlation coefficients in the bivariate normal populations, *Journal of the National Institute of Personnel Research, 6,* 175–176.

Ayres, L. P. (1920). A shorter method for computing the coefficient of correlation, *Journal of Educational Research, 1,* 216–221, 295–298.

Azen, S. P., and Reed, A. H. (1973). Maximum likelihood estimation of correlation between variates having equal coefficients of variation, *Technometrics,* **15,** 457–462.

Baker, G. A. (1930). The significance of the product-moment coefficient of correlation with special reference to the character of the marginal distributions, *Journal of the American Statistical Association,* **25,** 387–396.

Banerjee, D. P. (1952). On the moments of the multiple correlation coefficient in samples from normal populations, *Journal of the Indian Society of Agricultural Statistics, 4,* 88–90.

Bebbington, A. C. (1978). A method of bivariate trimming for robust estimation of the correlation coefficient, *Applied Statistics,* 27, 221–226.

Bedrick, E. J. (1990). On the large sample distributions of modified sample biserial correlation coefficients, *Psychometrika,* **55,** 217–228.

Bedrick, E. J. (1992). A comparison of generalized and modified sample biserial correlation estimators, *Psychometrika,* **57,** 183–201.

Bhatti, M. I. (1990). Null distribution of the small sample mean correlation coefficient and application to medical research, ***Biometrical Journal,*** **32,** 435–440.

Boas, F. (1909). Determination of the coefficient of correlation, ***Science*** (New ***Series),*** **29,** 823–824.

Boomsma, A. (1975). An error in F. N. David's tables of the correlation coefficient, ***Biometrika, 62,*** 711.

Chance, W. A. (1984). A geometric derivation of the distribution of the correlation coefficient (r) when $= 0$, ***American Mathematical Monthly,*** **93,** 94–96.

Chaubey, Y. P., and Mudholkar, G. S. (1978). A new approximation for Fisher's z, ***Australian Journal of Statistics,*** **20,** 250–256.

Cheriyan, K. C. (1945). Distributions of certain frequency constants in samples from nonnormal populations, ***Sankhyā***, 7, 159–166.

Cheshire, I., Oldis, E., and Pearson, E. S. (1932). Further experiments on the sampling distribution of the correlation coefficient, ***Journal of the American Statistical Association, 27,*** 121–128.

Cook, M. B. (1951a). Bi-variate k-statistics and cumulants of their joint sampling distribution, ***Biometrika,*** **38,** 179–195.

Cook, M. B. (1951b). Two applications of bi-variate k-statistics, ***Biometrika,*** **38,** 368–376.

Cox, D. R. (1966). The null distribution of the first serial correlation coefficient, ***Biometrika,*** **53,** 623–626.

Daniels, H. E. (1956). The approximate distribution of the serial correlation coefficients, ***Biometrika,*** **43,** 169–185.

Daniels, H. E., and Kendall, M. G. (1958). Short proof of Miss Harley's theorem on the correlation coefficient, ***Biometrika,*** **45,** 571–572.

David, F. N. (1937). A note on unbiased limits for the correlation coefficient, ***Biometrika,*** **29,** 157–160.

David, F. N. (1938). ***Tables of the Correlation Coefficient***, Cambridge: Cambridge University Press. [2nd ed., (1954).]

David, F. N., and Fix, E. (1966). Randomization and the serial correlation coefficient, In ***Festschrift for J. Neyman*** (ed., F. N. David), New York: Wiley, pp. 461–468.

DeGroot, M. H., and Goel, P. K. (1980). Estimation of the correlation coefficient from a broken random sample, ***Annals of Statistics,*** **8,** 264–278.

DeLury, D. B. (1938). Note on correlations, ***Annals of Mathematical Statistics,*** **9,** 149–151.

Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients, ***Biometrika,*** 62, 531–548.

Diaconis, P., and Efron, B. (1983). Computer intensive methods in statistics, ***Scientific American,*** **248,** 116–130.

Ding, C. G., and Bargmann, R. E. (1991a). AS 260—Evaluation of the distribution of the square of the sample multiple-correlation coefficient, ***Applied Statistics,*** **40,** 195–198.

Ding, C. G., and Bargmann, R. E. (1991b). AS 261—Quantiles of the distribution of the square of the sample multiple-correlation coefficient, ***Applied Statistics,*** **40,** Dixon, W. J. (1944). Further contributions to the problem of serial correlation, ***Annals of Mathematical Statistics,*** **15,** 119–144.

Dodd, S. C. (1926). A correlation machine, ***Industrial Psychology,*** **1,** 46–58.

Donner, A., and Rosner, B. (1980). On inferences concerning a common correlation coefficient, ***Applied Statistics,*** **29,** 69–76.

Duncan, G. T., and Layard, M. W. J. (1973). A Monte Carlo study of asymptotically robust tests for correlation coefficients, ***Biometrika,*** **60,** 551–558.

Durbin, J. (1957). Testing for serial correlation in systems of simultaneous regression equations, ***Biometrika,*** **44,** 370–377.

Durbin, J. and Watson, G. S. (1950, 1951). Testing for serial correlation in least squares regression I, II, ***Biometrika,*** **37,** 409–428; **38,** 159–178.

Edgell, S. E., and Noon, S. M. (1984). Effects of violation of normality on the t test of the correlation coefficient, ***Psychological Bulletin,*** **95,** 576–583.

Edgeworth, F. Y. (1892a). Correlated averages, ***Philosophical Magazine, 5th Series,*** **34,** 190–207.

Edgeworth, F. Y. (1892b). The law of error and correlated averages, ***Philosophical Magazine, 5th Series,*** **34,** 429–438, 518–526.

Eicker, F. (1960). On the distribution of a noncircular serial correlation coefficient with lag 1 when the mean of the observations is unknown, ***University of North Carolina Mimeo Series,*** No. 265 (Abstract in ***Annals of Mathematical Statistics,*** **31,** 234).

Fancher, R. E. (1989). Galton on examinations: An unpublished step in the invention of correlation, ***ISIS***, 80, 446–455.

Ferguson, T. S. (1958). A method for generating best asymptotically normal estimates with application to the estimation of bacterial densities, ***Annals of Mathematical Statistics,*** **29,** 1046–1062.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, ***Biometrika,*** **10,** 507–521.

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample, ***Metron,*** **1,** 3–32.

Fisher, R. A. (1924). The influence of rainfall on the yield of wheat at Rothamsted, ***Philosophical Transactions of the Royal Society of London, Series B,*** **213,** 89–142.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient, ***Proceedings of the Royal Society of London, Series A,*** **121,** 654–673.

Fowler, R. L. (1987). Power and robustness in product-moment correlation, ***Applied Psychological Measurement,*** **11,** 419–428.

Fraser, D. A. S. (1963). On the definition of fiducial probability, ***Bulletin of the International Statistical Institute,*** **40,** 842–856.

Gajjar, A. V. (1967). Limiting distributions of certain transformations of multiple correlation coefficient, ***Metron,*** 26, 189–193.

Gajjar, A. V., and Subrahmaniam, K. (1978). On the sample correlation coefficient in the truncated bivariate normal, ***Communications in Statistics—Simulation and Computation***, **7,** 455–477.

Galton, F. (1888). Co-relations and their measurement, chiefly from anthropometric data, ***Proceedings of the Royal Society of London,*** **45,** 135–145.

Garding, L. (1941). The distributions of the first and second order moments, the partial correlation coefficients and the multiple correlation coefficient in samples from a normal multivariate population, ***Skandinacisk Aktuarietidskrift,*** **24, 185–202.**

Garwood, F. **(1933).** The probability integral of the correlation coefficient in samples from a normal bi-variate population, ***Biometrika,*** 25, **71–78.**

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes, ***Biometrika,*** **38, 219–247.**

Ghosh, B. K. **(1966).** Asymptotic expansions for the moments of the distribution of correlation coefficient, ***Biometrika,*** **53, 258–262.**

Gokhale, D. V., and Press, S. J. **(1982).** Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution, ***Journal of the Royal Statistical Society, Series A*****, 145,** 237–249.

Goodman, A. S., and Lewis, P. A. W. **(1972).** The null distributions of test statistics based on sample serial correlations and the periodogram for renewal processes with exponential and $\frac{1}{2}$ Weibull intervals, ***IBM Research Report.***

Greco, L. **(1992).** The probability integral of the sample correlation coefficient, ***Journal of the Italian Statistical Society,*** **1, 289–294.**

Guenther, W. C. **(1977).** Desk calculation of probabilities for the distribution of the correlation coefficient, ***The American Statistician,*** **31, 45–48.**

Guilford, J. P., and Frischter, B. **(1973).** ***Fundamental Statistics in Psychology and Education,*** New York: McGraw-Hill.

Gupta, A. K., and Kabe, D. G. **(1991).** On some non-central distribution problems for the mixture of two normal populations, ***Metrika,*** **38, 1–10.**

Gupta, R. P. (1987). Estimation of multiple correlation coefficient when observations are missing on one variable, ***Communications in Statistics—Theory and Methods,*** **16, 353–360.**

Gurland, J. **(1968).** A relatively simple form of the distribution of the multiple correlation coefficient, ***Journal of the Royal Statistical Society, Series B,*** **30, 276–283.**

Gurland, J., and Asiribo, **O. (1991).** On the distribution of the multiple correlation coefficient and the kinked chi-square random variable, ***Statistica Sinica***, 1, **493–502.**

Gurland, J., and Milton, R. C. **(1970).** Further consideration of the distribution of the multiple correlation coefficient, ***Journal of the Royal Statistical Society, Series B,*** **32, 381–394.**

Hampel, F. R. **(1974).** The influence curve and its role in robust estimation, ***Journal of the American Statistical Association, 69,*** **383–393.**

Harley, B. I. **(1954).** A note on the probability integral of the correlation coefficient, ***Biometrika,*** **41,** 278–280.

Harley, B. **I. (1956).** Some properties of an angular transformation for the correlation coefficient, ***Biometrika,*** **43, 219–224.**

Harley, B. I. (1957a). Relation between the distribution of noncentral t and of a transformed correlation coefficient. ***Biometrika.*** **44. 219–224.**

Harley, B. I. (1957b). Further properties of an angular transformation of the correlation coefficient, ***Biometrika*, 44**, 273–275.

Harris, J. A. (1910). Arithmetic of the product moment method of calculating the correlation coefficient, ***American Naturalist***, 44, 693–699.

Havlicek, L. L., and Peterson, N. L. (1977). Effects of the violation of assumptions upon significance levels of the Pearson *r*, ***Psychological Bulletin*, 84**, 373–377.

Hedges, L. V., and Olkin, I. (1985). ***Statistical Methods for Meta Analysis***, San Diego, CA: Academic Press.

Hey, G. B. (1938). A new method of experimental sampling illustrated on certain non-normal populations, ***Biometrika***, 30, 68–80.

Hirschfeld, H. O. (Hartley, H. O.). (1937). The distribution of the ratio of covariance estimates in samples drawn from normal bivariate populations, ***Biometrika***, 29, 65–79.

Hodgson, V. (1967). On the sampling distribution of the multiple correlation coefficient, Presented at the ***IMS Meetings in Washington, DC***, December, 1967.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms (with discussion), ***Journal of the Royal Statistical Society, Series B***, 15, 193–232.

Hsu, P. L. (1946). On the asymptotic distributions of certain statistics used in testing the independence between successive observations from a normal population, ***Annals of Mathematical Statistics*, 17**, 350–354.

Huber, P. J. (1977). ***Robust Statistical Procedures***, Philadelphia: Society for Industrial and Applied Mathematics.

Huffaker, C. L. (1925). A note on statistical method, ***Journal of Educational Psychology*, 16**, 265–266.

Hull, C. L. (1925). An automatic correlation calculating machine, ***Journal of the American Statistical Association***, 20, 522–531.

Institute of Mathematical and Statistical Libraries (1987). Packages, Version 1.0, Vol. 3, Houston, Texas.

Isserlis, L. (1915). On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression, ***Biometrika*, 11**, 185–190.

Isserlis, L. (1917). The variation of the multiple correlation coefficient in samples drawn from an infinite population with normal distribution, ***Philosophical Magazine, 6th Series*, 34**, 205–220.

Iwase, K. (1981). A class of unbiased estimators of the normal correlation coefficient in the incomplete sufficient case, ***Journal of the Japanese Statistical Society*, 11**, 127–141.

Iwase, K. (1985). Moments of the correlation coefficient, ***Memoirs of the Faculty of Engineering, Serial No. 27, Hiroshima University*, 9**(1), 1–3.

Iwase, K., and Setô, N. (1984). A construction of incomplete sufficient unbiased estimators of the normal correlation coefficient, ***Journal of the Japanese Statistical Society*, 14**(1), 49–61.

Jenkins, G. M. (1954a). An angular transformation for the serial correlation coefficient, ***Biometrika*, 41**, 261–265.

Jenkins, G. M. (1954b). Tests of hypotheses in the linear autoregressive model, I, ***Biometrika*. 41.** 405–419.

Jenkins, G. M. (1956). Tests of hypotheses in the linear autoregressive model, II, *Biometrika*, **43**, 186–199.

Kamat, A. R., and Sathe, Y. S. (1962). Asymptotic power of certain test criteria (based on first and second differences) for serial correlation between successive observations, *Annals of Mathematical Statistics*, **33**, 186–200.

Kass, S. (1989). An eigenvalue characterization of the correlation coefficient, *American Mathematical Monthly*, **98**, 910–911.

Kelley, T. L. (1923). *Statistical Methods*, New York: Macmillan.

Kemp, A. W. (1970). General formulae for the central moments of certain serial correlation coefficient approximations, *Annals of Mathematical Statistics*, **41**, 1363–1368.

Kendall, M. G. (1949). Rank and product-moment correlation, *Biometrika*, **36**, 177–193.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation, *Biometrika*, **41**, 403–404.

Kendall, M. G. (1957). The moments of the Leipnik distribution, *Biometrika*, **44**, 270–272.

Kendall, M. G. (1960). The evergreen correlation coefficient, In *Essays in Honor of Harold Hotelling* (ed., I. Olkin), Stanford: Stanford University Press, pp. 274–277.

Khatri, C. G. (1966). A note on a large sample distribution of a transformed multiple correlation coefficient, *Annals of the Institute of Statistical Mathematics*, **18**, 375–380.

Kocherlakota, K., Balakrishnan, N., and Kocherlakota, S. (1986). On the performance of the SPRT for correlation coefficient: Normal and mixtures of normal populations, *Biometrical Journal*, **28**, 323–335.

Kocherlakota, K., and Kocherlakota, S. (1981). On the distribution of r in samples from the mixture of bivariate normal populations, *Communications in Statistics—Theory and Methods*, **10**, 1943–1966.

Kocherlakota, S., Kocherlakota, K., and Balakrishnan, N. (1985). Effects of nonnormality on the SPRT for the correlation coefficient: Bivariate Edgeworth series distribution, *Journal of Statistical Computation and Simulation*, **23**, 41–51.

Kocherlakota, S., and Singh, M. (1982a). On the behaviour of the sample correlation coefficient and its transforms in samples from discrete bivariate distributions, *Communications in Statistics—Theory and Methods*, **11**, 2017–2043.

Kocherlakota, S., and Singh, M. (1982b). On the behaviour of the sample correlation coefficient and its transforms in samples from the bivariate t and the bivariate χ^2 distribution, *Communications in Statistics—Theory and Methods*, **11**, 2045–2060.

Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient, *Biometrika*, **65**, 654–656.

Koopmans, T. C. (1942). Serial correlation and quadratic forms in normal variables, *Annals of Mathematical Statistics*, **13**, 14–33.

Kotz, S., and Johnson, N. L. (eds.) (1991). *Breakthroughs in Statistics*, *vol. 1*, New York: Springer-Verlag.

Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient, *Applied Statistics*, **21**, 1–12.

Kraemer, H. C. (1973). Improved approximation to the non-null distribution of the correlation coefficient, ***Journal of the American Statistical Association,*** **68, 1004–1008.**

Kraemer, H. C. (1975). On estimation and hypothesis testing problems for correlation coefficients, ***Psychometrika,*** 40, 473–485.

Kraemer, H. C. **(1980).** Robustness of the distribution theory of product moment correlation, ***Journal of Educational Statistics, 5,*** **115–128.**

Kramer, K. H. **(1963).** Tables for constructing confidence limits on the multiple correlation coefficient, ***Journal of the American Statistical Association,*** **58, 1082–1085.**

Kullback, S. (1936). A note on the multiple correlation coefficient, ***Metron*****, 12, 67–72.**

Kuzmin, R. O. **(1939).** Sur la loi de distribution du coefficient de correlation dans les tirages d'un ensemble normal, ***Comptes Rendus (Doklady) de l'Académie des Sciences, URSS (New Series), 22,*** **298–301.**

Leipnik, **R.** B. **(1947).** Distribution of the serial correlation coefficient in a circularly correlated universe, ***Annals of Mathematical Statistics,*** **18, 80–87.**

Leipnik, **R.** B. **(1958).** Note on the characteristic function of a serial-correlation distribution, ***Biometrika,*** **45, 559–562.**

Lewis, P. A. W. **(1972).** Recent results in the statistical analysis of univariate point processes, In ***Stochastic Point Processes: Statistical Analysis, Theory and Applications*** (ed., P. ***A.*** W. Lewis), New York: Wiley.

Madansky, A. (1958). On the maximum likelihood estimate of the correlation coefficient, ***Report P-1355,*** Rand Corporation.

Madow, W. G. **(1945).** Note on the distribution of the serial correlation coefficient, ***Annals of Mathematical Statistics,*** **16, 308–310.**

Mahalanobis, P. C., Bose, R. C., and Roy, S. N. **(1937).** Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions, ***Sankhyā*****, 3, 1–40.**

Marriott, F. H. C., and Pope, J. **A.** (1954). Bias in the estimation of auto-correlations, ***Biometrika,*** **41, 390–402.**

McCullagh, P. (1989). Some statistical properties of a family of continuous distributions, ***Journal of the American Statistical Association,*** **84, 125–129.**

McNemar, Q. (1962). ***Psychological Statistics,*** New York: Wiley.

Mehta, J. S., and Gurland, J. **(1969).** Some properties and an application of a statistic arising in testing correlation, ***Annals of Mathematical Statistics,*** **40,** 1736–1745.

Mi, J. **(1990).** Notes on the MLE of correlation coefficient in meta analysis, ***Communications in Statistics—Theory and Methods,*** **19, 2035–2052.**

Moran, P. **A.** P. **(1948).** Rank correlation and product-moment correlation, ***Biometrika,*** **35,** 203–206.

Moran, P. **A.** P. **(1950).** The distribution of the multiple correlation coefficient, ***Proceedings of the Cambridge Philosophical Society,*** **46, 521–522.**

Moran, P. **A.** P. (1967a). Testing for correlation between non-negative variates, ***Biometrika,*** **54, 385–394.**

Moran, P. A. P. (1967b). Testing for serial correlation with exponentially distributed variates. ***Biometrika.*** **54. 395–401.**

Moran, P. A. P. **(1970).** A note on serial correlation coefficients, ***Biometrika***, 57, **670–673.**

Moschopoulos, P. G., and Mudholkar, G. S. **(1983).** A likelihood-ratio based normal approximation for the non-null distribution of the correlation coefficient, ***Communications in Statistics—Simulation and computation, 12,*** **355–371.**

Muddapur, M. V. **(1988).** A simple test for correlation coefficient in a bivariate normal distribution, ***Sankhyā, Series*** B, ***50,*** **60–68.**

Mudholkar, G. S., and Chaubey, Y. P. **(1976).** On the distribution of Fisher's transformation of the correlation coefficient, ***Communications in Statistics—Simulation and Computation, 5,*** **163–172.**

Murthy, V. K. **(1960).** On the distribution of the sum of circular serial correlation coefficients and the effect of non-normality on its distribution, ***Annals of Mathematical Statistics,*** **31, 239–240** (Abstract).

Nabeya, S. **(1951).** Note on the moments of the transformed correlation, ***Annals of Mathematical Statistics, 3,*** 1.

Nair, A. N. K. **(1941).** Distributions of Student's ***"t"*** and the correlation coefficient in samples from non-normal populations, ***Sankhyā***, 5, **383–400.**

Nakagawa, S., and Niki, N. **(1992).** Distribution of the sample correlation coefficient for nonnormal populations, ***Journal of the Japanese Society for Computational Statistics, 5,*** ***1*** **–19.**

Neumann, J. von **(1941).** Distribution of the ratio of the mean square successive difference to the variance, ***Annals of Mathematical Statistics,*** **12,** 367–395.

Niki, N., and Konishi, S. **(1984).** Higher order asymptotic expansions for the distribution of the sample correlation coefficient, ***Communications in Statistics—Simulation and Computation,*** **13, 169–182.**

Norris, R. C., and Hjelm, H. F. **(1961).** Nonnormality and product-moment correlation, ***Journal of Experimental Education,*** **29, 261–270.**

Nunnaly, J. C. **(1967).** ***Psychometric Theory,*** New York: McGraw-Hill.

Odeh, R. E. **(1982).** Critical values of the sample product-moment correlation coefficient in the bivariate normal distribution, ***Communications in Statistics—Simulation and Computation,*** 11, **1–26.**

Odeh, R. E. **(1983).** On formulas for the exact distribution of the correlation coefficient using REDUCE, ***ASA Proceedings of Statistical Computing Section***, **306–308.**

Odeh, R. E., and Owen, D. B. **(1980).** ***Tables for Normal Tolerance Limits, Sampling Plans, and Screening,*** New York: Dekker.

Öksoy, O., and Aroian, L.A. **(1981).** Formulas for the rapid calculation of the ordinates, distribution and percentage points of the coefficient of correlation, ***Technical Report,*** Union College, Schenectady, NY.

Öksoy, O., and Aroian, L. A. **(1982).** ***Percentage Points of the Distribution of the Correlation Coefficient,*** Institute of Administration and Management, Union College, Schenectady, NY.

Olkin, I., and Pratt, J. W. **(1958).** Unbiased estimation of certain correlation coefficients, ***Annals of Mathematical Statistics,*** **32, 448–460.**

Olkin, I., and Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables, ***Annals of Mathematical Statistics,*** **32**, 448–465.

Pan, J.-J. (1964). Distribution of the serial correlation coefficients with non-circular statistics, ***Shuxue Jinzhan***, **7**, 328–337. (In Chinese) [English translation: ***Selected Translations in Mathematical Statistics and Probability,*** **7**, 281–291 (1968), American Mathematical Society, Providence, RI]; ***Ibid;*** 9, 291–295 (1966).

Paul, S. R. (1988). Estimation of and testing significance for a common correlation coefficient, ***Communications in Statistics—Theory and Methods,*** **17**, 39–53.

Pearson, E. S. (1929). Some notes on sampling tests with two variables, ***Biometrika,*** **21,** 337–360.

Pearson, E. S. (1931). The test of significance for the correlation coefficient, ***Journal of the American Statistical Association,*** **26**, 128–134.

Pearson, K. (1895). Mathematical contributions to the theory of evolution (Abstract), ***Proceedings of the Royal Society, London,*** **58**, 240–242.

Pearson, K. (1896). Mathematical contributions to the theory of evolution—111. Regression, heredity and panmixia, ***Philosophical Transactions of the Royal Society, London, Series A***, **187**, 253–318.

Pearson, K. (1909). Determination of the coefficient of correlation, ***Science (New Series),*** **30**, 23–25.

Pearson, K. (1920). Notes on the history of correlation, In ***Studies in the History of Statistics and Probability,*** *1* (eds., E. S. Pearson and M. G. Kendall), London: Griffin. (***Biometrika,*** **13**, 25–45.)

Pearson, K. (1931). Editorial note, ***Biometrika,*** **22,** 362–367.

Pearson, K. (1933). On a method of determining whether a sample of size ***n*** supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random, ***Biometrika,*** **25**, 379–410.

Pearson, K., Jeffery, G. B., and Elderton, E. M. (1929). On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population, ***Biometrika,*** **21**, 164–193. (***Tables*** computed by E. M. Elderton, pp. 194–201.)

Pepper, J. (1929). Studies in the theory of sampling, ***Biometrika,*** **21**, 231–258.

Pillai, K. C. S. (1946). Confidence interval for the correlation coefficient, ***Sankhyā***, **7**, 415–422.

Pradhan, M., and Sathe, Y. S. (1975). An unbiased estimator and a sequential test for the correlation coefficient, ***Journal of the American Statistical Association,*** **70,** 160–161.

Press, S. J. (1967). On the sample covariance from a bivariate normal distribution, ***Annals of the Institute of Statistical Mathematics,*** **19**, 355–361.

Press, S. J. (1969). On serial correlation, ***Annals of Mathematical Statistics,*** **40,** 188–196.

Quenouille, M. H. (1948). Some results in the testing of serial correlation coefficients, ***Biometrika,*** **35,** 261–267.

Quensel, C. E. (1938). The distributions of the second moment and of the correlation coefficient in samples from populations of Type A, ***Lunds Universitet***, ***Årsskrift N. F. Afdelning,*** (2), **34**, (4), 1–111.

Ranney, G. B., and Thigpen, C. C. **(1981).** The sample coefficient of determination in simple linear regression, ***The American Statistician,*** **35, 152–153.**

Rao, C. R. **(1983).** Multivariate analysis. Some reminiscences on its origin and development, ***Sankhyā, Series B,*** **45, 284–299.**

Rasmussen, J. L. **(1987).** Estimating correlation coefficients: Bootstrap and parametric approaches, ***Psychological Bulletin,*** **101, 136–139.**

Rider, P. R. **(1932).** On the distribution of the correlation coefficient in small samples, ***Biometrika,*** **24, 382–403.**

Rodgers, J. L., and Nicewander, A. L. **(1988).** Thirteen ways to look at the correlation coefficient, ***The American Statistician,*** **42, 59–66.**

Romanovsky, V. **(1925).** On the moments of standard deviations and of correlation coefficient in samples from normal population, ***Metron,*** **5 (4), 3–46.**

Ruben, H. **(1963).** Probability of a positive sample correlation (Abstract), ***Annals of Mathematical Statistics,*** **34, 694.**

Ruben, H. **(1966).** Some new results on the distribution of the sample correlation coefficient, ***Journal of the Royal Statistical Society, Series B,*** **28, 513–525.**

Rubin, H. **(1945).** On the distribution of the serial correlation coefficient, ***Annals of Mathematical Statistics,*** **16, 211–215.**

Samiuddin, M. **(1970).** On a test for an assigned value of correlation in a bivariate normal distribution, ***Biometrika,*** **57, 461–464.**

Saw, J. G. **(1984).** Ultraspherical polynomials and statistics on the m-space, ***Journal of Multivariate Analysis,*** **14, 105–113.**

Shanmugam, M., and Gajjar, A. V. **(1992).** On the distribution of sample correlation coefficient and its transformations in samples from Morgenstern bivariate exponential distribution, ***Communications in Statistics—Simulation and Computation,*** **21, 1021–1041.**

Shenton, L. R., and Johnson, W. L. **(1965).** Moments of a serial correlation coefficient, ***Journal of the Royal Statistical Society, Series B,*** **27, 308–320.**

Sibuya, M. **(1964).** An incomplete sufficient unbiased estimate, ***Mimeographed Note,*** University of Tokyo, Tokyo, Japan.

Silver, N. C., and Dunlap, W. P. **(1987).** Averaging correlation coefficients: Should Fisher's z transformation be used? ***Journal of Applied Psychology,*** **72, 146–148.**

Soper, H. E. **(1913).** On the probable error of the correlation coefficient to a second approximation, ***Biometrika,*** **9, 91–115.**

Soper, H. E. **(1929).** The general sampling distribution of the multiple correlation coefficient, ***Journal of the Royal Statistical Society, Series A,*** **92, 445–447.**

Soper, H. E., Young, A. W.,. Cave, B. M., Lee, A., and Pearson, K. **(1917).** On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher, ***Biometrika,*** **11, 328–413.**

Srivastava, M. S. **(1983).** On the distribution of Hotelling's T^2 and multiple correlation R^2 when sampling from a mixture of two normals, ***Communications in Statistics—Theory and Methods,*** **12, 1481–1487.**

Srivastava, M. S., and Awan, H. M. **(1982).** On the robustness of Hotelling's T^2-test and distributions of linear and quadratic forms in sampling from a mixture of two multivariate normal populations, ***Communications in Statistics—Theory and Methods.*** **11. 81–107.**

Srivastava, M. S., and Awan, H. M. (1984). On the robustness of the correlation coefficient in sampling from a mixture of two bivariate normals, *Communications in Statistics—Theory and Methods,* **13,** 374–382.

Srivastava, M. S., and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics,* Amsterdam: North-Holland.

Srivastava, M. S., and Lee, G. S. (1984). On the distribution of the correlation coefficient when sampling from a mixture of two bivariate densities; robustness and the effect of outliers, *Canadian Journal of Statistics,* **12,** 119–133.

Stammberger, A. (1968). Ein Nomogramm zur Verteilung von Korrelationskoeffizienten, *Biometrische Zeitschrift,* **10,** 80–83.

Stigler, S. M. (1978). Francis Ysidro Edgeworth, Statistician, *Journal of the Royal Statistical Society, Series A,* **141,** 287–322.

Stuart, A., and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics—Vol. I,* 6th edition, London: Edward Arnold.

'Student' (1908). Probable error of a correlation coefficient, *Biometrika,* 6, 302–310.

Subrahmaniam, K., and Gajjar, A. V. (1978). Robustness to nonnormality of the sample correlation coefficient, *Technical Report No. 92,* Department of Statistics, University of Manitoba, Winnipeg, Canada.

Subrahmaniam, K., and Gajjar, A. V. (1980). Robustness to non-normality of some transformations of the sample correlation coefficient, *Journal of Multivariate Analysis,* **10,** 69–78.

Subrahmaniam, K., Gajjar, A. V., and Subrahmaniam, K. (1981). Polynomial representations for the distribution of the sample correlation and its transformations, *Sankhyā, Series B,* **43,** 319–351.

Subrahmaniam, K., and Subrahmaniam, K. (1983). Some extensions to Miss F. N. David's *Tables of Sample Correlation Coefficient: Distribution Function and Percentiles, Sankhyii, Series B, 45,* 75–147.

Sylvan, M. (1969). Estimation and hypothesis testing for the Wishart matrices when part of the covariance matrix is known, *Technical Report No. 38,* Department of Statistics, Stanford University, Stanford, CA.

Symonds, P. H. (1926). Variations of the product-moment (Pearson) coefficient of correlation, *Journal of Educational Psychology,* **17,** 458–469.

Tankard, J. W., Jr. (1984). *The Statistical Pioneers,* Cambridge, MA: Schenkman.

Thomas, H. (1989). *Distributions of Correlation Coefficients,* New York: Springer-Verlag.

Thurstone, L. L. (1917). A method of calculating the Pearson correlation coefficient without the use of deviations, *Psychological Bulletin,* **14,** 28–32.

Tiku, M. L., and Balakrishnan, N. (1986). A robust test for testing the correlation coefficient, *Communications in Statistics—Simulation and Computation,* **15,** 945–971.

Titterington, D. M. (1978). Estimation of correlation coefficients by ellipsoidal trimming, *Applied Statistics,* **27,** 227–234.

Venables, W. N. (1985). The multiple correlation coefficient and Fisher's *A* statistic, *Australian Journal of Statistics,* **27,** 172–182.

Viana, M. A. G. (1982). Combined estimators for the correlation coefficient, *Communications in Statistics—Theory and Methods,* **11,** 1483–1504.

Wald, A., and Wolfowitz, J. (1943). An exact test for randomness in the non-parametric case based on serial correlation, *Annals of Mathematical Statistics*, **14**, 378–388.

Wallace, D. L. (1959). Bounds on normal approximations to Student's and the chi-square distribution, *Annals of Mathematical Statistics*, **30**, 1121–1130.

Watson, G. S. (1956). On the joint distribution of the circular serial correlation coefficients, *Biometrika*, **43**, 161–168.

Watson, G. S., and Durbin, J. (1951). Exact tests of serial correlation using non-circular statistics, *Annals of Mathematical Statistics*, **22**, 446–451.

White, J. S. (1957). Approximate moments for the serial correlation coefficient, *Annals of Mathematical Statistics*, **28**, 798–803.

White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient, *Biometrika*, **48**, 85–94.

Wilks, S. S. (1932). On the sampling distribution of the multiple correlation coefficient, *Annals of Mathematical Statistics*, **3**, 196–203.

Williams, E. J. (1978). A simple derivation of the distribution of the multiple correlation coefficient, *Communications in Statistics—Theory and Methods*, **7**, 1413–1420.

Williams, E. J. (1993). Minimum-information probability models with an application to correlation, *Mathematical Scientist*, **18**, 51–59.

Winterbottom, A. (1980). Estimation for the bivariate normal correlation coefficient using asymptotic expansion, *Communications in Statistics—Simulation and Computation*, **9**, 599–609.

Wishart, J. (1931). The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population, *Biometrika*, **22**, 353–361.

Wishart, J., and Bartlett, M. S. (1932). The distribution of second order moment statistics in a normal system, *Proceedings of the Cambridge Philosophical Society*, **28**, 455–459.

Yang, G. L. (1970). Asymptotic tests for serial correlation with exponentially distributed random variables, *Annals of Mathematical Statistics*, **41**, 1390.

Yule, G. U. (1895). On the correlation of total pauperism with proportion of out-relief—I. All ages, *Economic Journal*, **5**, 603–611.

Yule, G. U. (1897). On the theory of correlation, *Journal of the Royal Statistical Society*, **60**, 812–854.

Yule, G. U. (1907). On the theory of correlation for any number of variables, treated by a new system of notation, *Proceedings of the Royal Society of London, Series A*, **79**, 182–193.

Yule, G. U. (1921). On the time-correlation problem with special reference to the variate-difference correlation method, *Journal of the Royal Statistical Society, Series A*, **84**, 497–526.

Zeller, R. A., and Levine, Z. H. (1974). The effects of violating the normality assumption underlying *r*, *Sociological Methods and Research*, **2**, 511–514.

CHAPTER 33

Lifetime Distributions and Miscellaneous Orderings

1 INTRODUCTION

The number of possible different continuous distributions is limitless. In this chapter we restrict our attention to distributions of some importance in statistical practice or theory that were not covered in Chapters 12 through 32.

We will mostly concentrate on lifetime distributions, a field that has been intensively studied in the years since the publication of the first edition of this volume. The basic text by Barlow and Proschan (1981) has substantially influenced the development of the subject. A more recent paper by Barlow and Mendel (1992) is a significant contribution to Bayesian analysis of lifetime distributions. The literature on lifetime (or life) distributions is enormous, so we have again had to restrict ourselves to a small representative sample of the results. Some of the literature is cited in the references without mention in the text.

We will look in some detail at a representative class of life distributions known as ***Birnbaum-Saunders distributions.*** These distributions are generated by a simple transformation of normal variables and have been rather extensively studied. Birnbaum-Saunders distributions can be derived from several models of mechanism of lifetime distribution.

We will also consider results on ***ordering*** of distributions, since this topic is especially relevant to life distributions. Ordering also confers considerable generality and enables some practical applications of distributions. Books by Pečarić, Proschan, and Tong (1992) and Shaked and Shanthikumar (1994) (and their co-authors who contributed to the chapters on applications) both provide a detailed and systematic treatment of stochastic orders, highlight their growing importance, and illustrate their usefulness in numerous applications. A classified bibliography of stochastic orders and applications has also been prepared by Mosler and Scarsini (1993) with the collaboration of R. Dyckerhoff and H. Holz.

We have omitted discussion of Kolmogorov-Smirnov and related distributions involving sample cumulative distribution functions. Although we presented this topic in the first edition, interest in it has grown over the last 25 years and now there is available orderly classification in the literature. The reader is referred to the compendium by D'Agostino and Stephens (1986) for extensive information on this subject. The discussion of circular normal (von Mises) distributions and their extensions (spherical normal distributions, etc.), the subject of a comprehensive monograph by Fisher, Lewis, and Embleton (1989) has been moved to the planned new edition of Continuous Multicariate Distributions.

2 LIFE DISTRIBUTIONS

Every cumulative distribution function F over the nonnegative real line can be a life distribution. If F is absolutely continuous then, provided that $1 - F_T(t) > 0$, the hazard (or failure) rate function

$$h_T(t) = -\left(\frac{d}{dt}\right)\log\{1 - F_T(t)\} \tag{33.1}$$

represents the conditional density of failure at time t, given survival until time t. When $p_T(t)$ is a density function, we also have the hazard rate function to be $h_T(t) = p_T(t)/\{1 - F_T(t)\}$; see Chapter 1. Given $h_T(t)$, $F_T(t)$ can be derived from

$$F_T(t) = 1 - \exp\left[-\int_0^t h_T(x)\,dx\right], \qquad 0 \le t < \infty. \tag{33.2}$$

Note that the hazard rate function is sometimes confused with the conditional failure density function

$$\frac{d\Pr[T \le t|T > x]}{dt} = \frac{p_T(t)}{1 - F_T(x)}, \qquad t > x.$$

We have already encountered a number of more or less commonly used life distributions such as exponential, mixture of exponentials, Weibull, and especially extreme value distributions, whose truncated form is the well-known Gompertz distribution (see Chapter 22). Extensive classes of such distributions, covering developments up to the early 1960s, were discussed by Buckland (1964).

The exponential distribution is used in many situations to represent distribution of lifetimes. If the departure from this distribution is too pronounced to be ignored, a Weibull distribution may be used. Among others used [Barlow (1968); Bain (1974)] there is the linear failure rate distribution (see also Section 4) with density function

$$p_T(t) = (1 + \theta t)\exp\{-(t + \tfrac{1}{2}\theta t^2)\}, \qquad t > 0, \qquad (33.3)$$

and a distribution with density function

$$p_T(t) = [1 + \theta(1 - e^{-t})]\exp\{-[t + \theta(t + e^{-t} - 1)]\}, \qquad t > 0. \quad (33.4)$$

(These distributions are in standard form. A further parameter can be introduced by considering the distribution of αT, with $\alpha > 0$.)

Flehinger and Lewis (1959) discussed two lifetime distributions constructed from specific hazard rate functions (see Section 4). These are given by

$$a + 2b^2t, \quad a + 3c^3t^2 \qquad \text{(Hazard rate)},$$

leading to

$$1 - \exp[-at - (bt)^2], \quad 1 - \exp[-at - (ct)^3]$$

$$\text{(Cumulative distribution function)}. \qquad (33.5)$$

[cf. (33.3)]. The parameters a, b, and the argument t are all positive. It is, of course, also possible to consider a distribution with hazard rate $a + 2b^2t + 3c^3t^2$ and cumulative distribution function

$$1 - \exp[-at - (bt)^2 - (ct)^3],$$

but the extra complexity makes this unattractive. Flehinger and Lewis also discussed the use of a truncated normal distribution (Chapter 13, Section 10.1) as a lifetime distribution.

Greenwich (1992) has suggested use of a distribution with unimodal hazard rate

$$at(b^2 + t^2)^{-1}, \qquad a, b > 0, \qquad (33.6a)$$

which corresponds to

$$F_T(t) = 1 - (1 + b^{-2}t^2)^{-a/2}, \qquad t \geq 0, \qquad (33.6b)$$

and is in fact the distribution of $b(2F_{2,a}/a)^{1/2}$, where $F_{2,a}$ denotes a central F random variable with (2, a) degrees of freedom (see Chapter 27). Shaked (1977) has investigated life distributions with hazard rate functions of the general form

$$h_T(t) = A_1g_1(t) + A_2g_2(t), \qquad 0 \leq t, \qquad (33.7)$$

where the functions $g_1(t)$ and $g_2(t)$ are assumed to be known and not to depend on A, or A,. The form (33.7) includes many types of hazard rate functions that are encountered in practice, including functions that are not everywhere monotonic. As noted above, Bain (1974) dealt with the case $g_j(t) = t^{\theta_j}$ $(j = 1, 2)$ [and the linear hazard rate obtained by taking $\theta_1 = 0$, $\theta_2 = 1$; see (33.3).] Gaver and Acar (1979) gave a thorough discussion of models of this kind. Gaver and Acar discussed bathtub hazard rate models

(i) $XL(X)R(X)$ with X standard exponential, and

$$L(X) \text{ concave in } X,\ L(0) < 1,\ L(\infty) = 1,$$
$$R(X) \text{ convex in } X,\ R(0) = 1,\ R(0) > R(\infty);$$

(ii) Hazard rate of form $h(t) = g(t) + \lambda + k(t)$, where

$$g(t) > 0 \text{ decreasing function of } t,\ \lim_{t \to \infty} g(t) = 0,$$
$$k(t) \text{ increasing (preferably with } k(0) = 0,\ k(\infty) = \infty),$$
$$\text{e.g. } h(t) = \frac{A}{t + \alpha} + Bt + A.$$

Crowder et al. (1991) give, inter *alia,* graphs of hazard functions for

Weibull ($c = 0.5, 1.5, 2.5, 5$); Unit normal; Gumbel;
Log normal [$\log X \to N(\mu, \sigma^2)$ with $\sigma = \frac{1}{4}, \frac{1}{2}, 1, 1\frac{1}{2}$, and μ chosen to make $E[X] = 1$];
Gamma [$a = 0.5, 1.5, 2.5, 5$ with β chosen to make $E[X] = 1$].

Hazard rates of the form (33.7) arise naturally, and commonly for systems with two independent components. Bain and Engelhardt (1991) provide a detailed discussion of estimation procedures for parameters of life-time distributions with polynomial hazard rates. Their method involves the determination of a least-squares polynomial fit to a set of points, and is applicable to both complete and censored samples.

Foster and Craddock (1974) considered (33.7) for the case $g_1(t) = 1$, $g_2(t) = \exp(-at)$. Substituting

$$g_1(t) = I_{[0,K]}(t),$$
$$g_2(t) = I_{[K,\infty)}(t)$$

(for some $K > 0$), in (33.7), we obtain piecewise constant hazard rate functions discussed by Prairie and Ostle (1961) and Colvert and Boardman (1976). Dimitrov, Chukova, and Green (1993) have discussed continuous probability distributions with periodic hazard rates. Shaked (1977) arrived at functions of

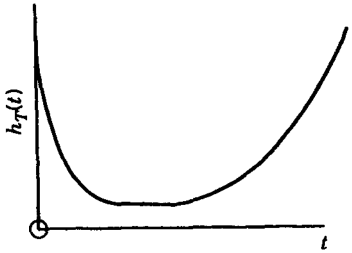


Figure 33.1 A bathtub-shaped hazard function.

form (33.7) with $g_1(t) = 1$, $g_2(t) = \sin t$ when modeling the hazard rate function of an item influenced by periodic fluctuations of temperature. He investigated maximum likelihood estimation of A, and A, when the distribution corresponding to (33.7) is of form IHR (increasing hazard rate), IHRA (increasing hazard rate on average), or NBU (new better than used) (see Section 4). Note that the distributions (33.5) are all of IHR type.

For many living (i.e., human) populations the hazard rate is of the so-called bathtub type (see Figure 33.1). In several practical situations, the effect of age on a component is initially beneficial, but after a certain period the effect of age is adverse. Infant mortality or work hardening of certain tools are typical instances of such aging. This type of monotonic aging is usually modelled using life distributions displaying bathtub failure rates (BFR) and may be defined as follows.

A life distribution F is said to be a BFR (upturned bathtub failure rate, UBFR) distribution if there exists a $t_0 \geq 0$ such that the cumulative hazard rate function $-\log \bar{F}(t)$ is concave (convex) on $[0, t_0]$ and convex (concave) on $[t_0, \infty)$.

This definition of a BFR distribution is quite general and extends the idea of distributions possessing a bathtub-shaped failure rate to situations where the failure rate itself does not exist [see Mitra and Basu (1994)].

MacGillivray (1981) investigated the relation between the number of changes from increasing to decreasing hazard rate and that of

$$\frac{p'_T(t)}{p_T(t)} = \frac{d \log F_T(t)}{dt},$$

the motivation being that the number of changes in concavity of $\log\{1 - F_T(t)\}$ is limited by that of $\log p_T(t)$.

Dhillon (1981) [see also Leemis (1986)] introduced a two-parameter system of life distributions—the exponential power distributions. The two-parameter system can have increasing, decreasing, or bathtub-shaped hazard rates. The survival functions are

$$\bar{F}_T(t) = 1 - F_T(t) = \exp\left[1 - \exp\{(\alpha t)^\beta\}\right], \qquad \alpha, \beta > 0, \quad (33.8a)$$

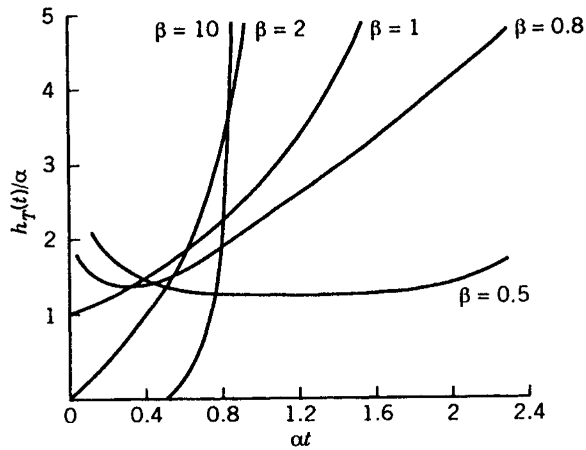


Figure 33.2 Hazard rate (33.8b) for various values of β.

with hazard rates

$$h_T(t) = \alpha\beta(\alpha t)^{\beta-1}\exp\{(\alpha t)^{\beta}\}. \tag{33.8b}$$

This is bathtub shaped if $\beta < 1$, achieving a minimum value for $h_T(t)$ at $t = \{(1-\beta)/(\alpha\beta)\}^{1/\beta}$ (see Figure 33.2). For $\beta = 1$ we have extreme value distributions (see Chapter 22).

Dhillon (1981) also constructed another two-parameter system, with survival functions

$$\bar{F}_T(t) = \exp\left[-\{\log(\lambda t + 1)\}^{\beta+1}\right], \quad \beta \geq 0, \quad \lambda > 0, \quad t \geq 0 \tag{33.9}$$

and hazard rate functions

$$h_T(t) = (\beta + 1)(\lambda t + 1)^{-1}\{\log(\lambda t + 1)\}^{\beta}. \tag{33.10}$$

Typical graphs of $h_T(t)$ are shown in Figure 33.3. Dhillon provided maximum likelihood estimation procedures for both systems of distributions, (33.8a) and (33.9).

Hjorth (1980) proposed a three-parameter system of life distributions with survival functions

$$\bar{F}_T(t) = (1 + \beta t)^{-\gamma/\beta}\exp\left(-\tfrac{1}{2}\alpha t^2\right), \qquad \alpha, \beta > 0, \tag{33.11}$$

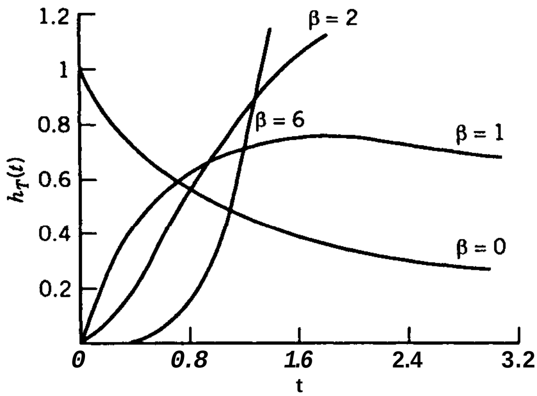


Figure 33.3 Hazard rate (33.10) for various values of β.

and hazard rates

$$h_T(t) = \alpha t + \gamma(1 + \beta t)^{-1}. \tag{33.12}$$

These distributions have bathtub shaped hazard rate functions, if $0 < a < \beta\gamma$.

A mixture of two Weibull distributions with shape parameters β_k satisfying $\beta_1 < 1 < \beta_2$ (see Chapter 21) also produces a distribution with a bathtub hazard rate. Unlike Dhillon's two-parameter bathtub distributions, this distribution involves five parameters; see Nelson (1982) and Lawless (1982).

By noting that the shape of a probability density function restricts that of its hazard function and vice versa, Sibuya (1994) was primarily concerned with the six specific shapes, viz., increasing, decreasing, unimodal, anti-unimodal, increasing-decreasing-increasing, and decreasing-increasing-decreasing. Then, of the six by six total possible shape combinations, Sibuya (1994) has shown that sixteen are impossible and by means of examples has shown the remaining twenty are possible.

Kunitz (1989) suggests using a mixture of gamma distributions (see Chapter 17) as an adequate model for distributions with bathtub-shaped hazard functions. These are defined in terms of the ***total-time-on-test (TTT) transform***

$$\overline{H}_{F^{-1}}(t) = \int_0^{F_T^{-1}(t)} \overline{F}_T(u)\, du. \tag{33.13}$$

These change from convex to concave in the interval (0, 1) and are "extremal" in the power of tests of exponentiality for DFR or IFR alternatives.

The type of monotonic ageing described earlier for bathtub failure rate models can also be studied through the class of IDMRL distributions,

introduced by Guess, Hollander, and Proschan (1986) through the mean residual life function, defined as follows:
A life distribution F is said to be an increasing initially, then decreasing mean residual life (IDMRL) (decreasing initially, then increasing mean residual life, DIMRL) distribution if there exists a $t_0 \geq 0$ such that $\mu_{(F)}(t) = \int_t^\infty \bar{F}(x)dx/\bar{F}(t)$ is nondecreasing (nonincreasing) on [0, t,,] and nonincreasing (nondecreasing) on $[t_0, \infty)$. See Section 4.2 (page 670) for related families introduced by Mitra and Basu (1994).

Zelterman (1992) studied hazard rate functions which rise rapidly and become unbounded at a finite point ψ corresponding to a cap on the maximum lifetime attainable. (The motivation is to estimate an absolute limit on attainable human lifespan conditional on the existence of such a limit.) Specifically:

$$h(t) = \theta\beta(\psi - t)^{\beta-1}, \qquad 0 \leq t < \psi, \qquad \theta > 0$$

with parameter $0 < \beta < 1$ controlling the shape. (Models with $\beta < 0$ correspond to the Type 2 extreme value distribution; see Chapter 22. For $\beta \to 0$, we obtain a generalized Pareto distribution (see Chapter 19) with survival function

$$\bar{F}(t) = \{1 - (\beta t/\sigma)\}^{1/\beta}, \qquad 0 \leq t \leq \sigma/\beta, \sigma > 0, \beta > 0,$$

with the hazard rate $h(t) = (a - \beta t)^{-1}$.) The corresponding survival function is

$$\bar{F}(t|\theta,\beta,\psi) = \begin{cases} 1, & t < 0 \\ \exp\left[-\theta\{\psi^\beta - (\psi - t)^\beta\}\right], & 0 \leq t < \psi \\ 0, & t \geq \psi. \end{cases}$$

The main feature of these survival functions is the discrete jump of $\bar{F}(t)$ to zero at $t = \psi$, representing an abrupt mass extinction of all survivors at the cap $t = \psi$, and the density function $p(t)$ associated with this $h(t)$ and $\bar{F}(t)$ is mixed in the sense that it contains both a continuous and a discrete component:

$$p(t) = p(t|\theta,\beta,\psi) = \begin{cases} \theta\beta(\psi - t)^{\beta-1}\bar{F}(t), & 0 \leq t < \psi \\ \exp(-\theta\psi^\beta), & \text{point mass at } t = \psi \\ 0, & \text{other values of t.} \end{cases}$$

Hazard rates with a change point or phases have been studied by Blackstone, Naftel, and Turner (1980) and Hazelrig, Turner, and Blackstone (1982) among others.

Hazelrig, Turner, and Blackstone (1982) provide some hazard functions for analyzing survival data with a family of continuous-differential survival distributions. (The models are related by a common differential equation; similarity to the earlier ideas of Voit, mentioned in Chapter 12, should be noted.) The generic differential equation, in terms of the cumulative distribution function, $F(t)$, is given by

$$\dot{F}(t) = \mu\{F(t)\}^{1-(1/m)}\left\{\frac{m}{|m|}(1 - F(t))^{1/m}\right\}^{1/(1+\nu)},$$

with the shape parameters $m, \nu > 0$, yielding under the initial condition $F(0) = 0$

$$F(t) = 1 - \frac{m}{|m|}\left(\frac{m + |m|}{2|m|} + \frac{\mu t}{|m|}\right)^{-\nu}.$$

The parameter μ in the time-scale factor is often called the ***modulus of mortality.*** Values of m are naturally classified into three categories:

(i) $m > 1$ or $m < -1/\nu$
(ii) $m = 1$ or $m = -1/\nu$
(iii) $0 < m < 1$ or $-1/\nu < m < 0$.

Hazelrig, Turner, and Blackstone (1982) provide a schematic diagram of special cases of this generic family which includes hyperbolic, hyper logistic, single-hit $[F(t) = (1 - e^{-\mu t/m})^m]$, hyper-Gompertz, and exponential distributions. Figure 33.4 presents graphs of four examples of this generic hazard function $h(\mu t)$. Note that the case $\nu = 3$ and $m = -0.2$ yields a hazard rate with a local maximum and a local minimum.

Hazard rates of the form

$$h(t) = \exp(a_0 + a_1 t + a_2 t^2)$$

were studied by Lewis and Shedler (1976, 1979) using simulation methodology.

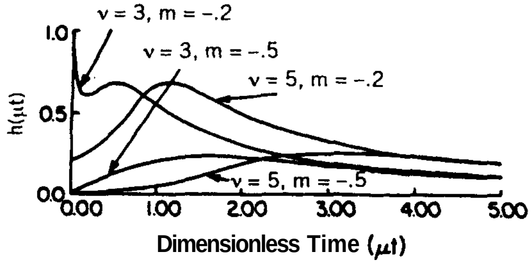


Figure 33.4 Four examples of the generic hazard function $h(\mu t)$.

Makino (1984) introduced the concept of mean hazard rate (MHR):

$$E\left[r_{(F)}(t)\right] = \int_0^{\infty} r_{(F)}(t) p_T(t)\, dt,$$

where $r_{(F)}(t) \equiv h_T(t)$. He showed that

$$E\left[r_{(F)}(t)\right] \geq \left[\int_0^{\infty} \{1 - F_T(t)\}\, dt\right]^{-1}.$$

(The latter quantity is sometimes called mean failure rate.)

For a unit normal distribution, MHR = 0.9048557; a Weibull distribution (Chapter 21) with shape parameter c = 3.4392 has approximately the same MHR. (It is noteworthy that for c = 3.43938 the median and expected value of the Weibull distribution coincide.)

Armero and Bayarri (1993) have discussed Kummer distributions. A random variable X has a Kummer distribution with parameters a, β, y, δ (a > 0, $\beta > 0$, 6 > 0) if it has a continuous distribution with probability density function

$$p_X(x|\alpha, \beta, \gamma, \delta) = Ce^{-\beta x} x^{\alpha - 1}/(1 + \delta x)^{\gamma}, \quad 0 < x < \infty,$$

where the constant C is such that

$$C^{-1} = \frac{\Gamma(\alpha)}{\delta^{\alpha}} U(\alpha, \alpha + 1 - \gamma, \beta/\delta).$$

Here, $U(a$, b, $z)$ is one of Kummer's functions (a confluent hypergeometric function; see Chapter 1) with an integral representation

$$\Gamma(a)U(a, \mathrm{b}, z) = \int_0^{\mathrm{m}} e^{-zt} t^{a-1} (1 + t)^{b-a-1}\, \mathrm{dt}, \quad a > 0, \quad z > 0.$$

The moments of X can be easily expressed in terms of the U function. For example, we have

$$E[X^k] = \frac{\Gamma(\alpha + k)}{\delta^k \Gamma(\alpha)} \frac{U(\alpha + k, \alpha + k + 1 - \gamma, \beta/\delta)}{U(\alpha, \alpha + 1 - \gamma, \beta/\delta)}.$$

As noted by Armero and Bayarri (1993), the Kummer distribution generalizes both gamma and F distributions. For example, the above Kummer distribution, when y = 0, becomes the Gamma (a, β) distribution. Similarly, it may be easily seen that the Kummer distribution with a = $\nu_1/2$, $\beta = 0$, y = $(\nu_1 + \nu_2)/2$ and 6 = 1 becomes the F-distribution with (ν_1, ν_2) degrees of freedom.

Evans and Swartz (1994) have discussed a class of densities useful in life-test analyses formed by taking the product of nonnegative polynomials and normal densities, and termed these distributions as polynomial-normal distributions. A polynomial-normal density function is of the form

$$\frac{C}{\theta} Q\left(\frac{x-\xi}{\theta}\right) Z\left(\frac{x-\xi}{\theta}\right), \quad \theta > 0,$$

where $Z(\cdot)$ is the standard normal density function, $Q(\cdot)$ is a nonnegative polynomial, and the normalizing constant C is given by

$$C = \left\{\int_{-\infty}^{\infty} Q(x) Z(x)\, dx\right\}^{-1}.$$

If $Q(\cdot)$ is a polynomial of degree 2m, the integration needed to determine C can be exactly evaluated by a Gauss-Hermite rule of order $m + 1$; in this case,

$$C^{-1} = \sum_{i=1}^{m+1} w_i Q(x_i)$$

where the w_i and x_i are the Gauss-Hermite weights and points, respectively. Similarly, the rth moment about zero can be evaluated exactly by a Gauss-Hermite rule of order $[(r + 2m + 1)/2]$. Observing that this family provides a rich class of distributions that can be used in modeling when faced with non-normal characteristics such as skewness and multimodality, Evans and Swartz (1994) have discussed some inference procedures for these distributions.

Creedy and Martin (1994) have used a generalized gamma distribution as a model of the distribution of prices. The probability density function of this distribution is

$$p_X(x) = \exp\{\theta_1 \log x + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 - \eta\}, \quad x > 0.$$

The normalizing constant η needs to be determined numerically. This generalized gamma distribution includes many well-known distributions as special cases. For example, the case $\theta_1 = \theta_3 = \theta_4 = 0$ gives an exponential distribution, the case $\theta_2 = \theta_4 = 0$ gives a Rayleigh distribution, and the case $\theta_3 = \theta_4 = 0$ yields a gamma distribution. In addition, this distribution is related to the generalized gamma distribution of McDonald (1984) and also to the generalized lognormal distribution discussed by Lye and Martin (1993).

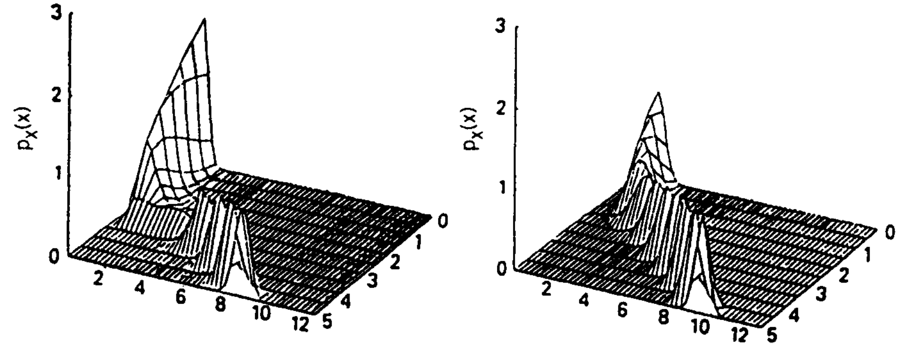


Figure 33.5 Two generalized gamma densities of Creedy and Martin.

Two plots of Creedy and Martin's generalized gamma densities are presented in Fig. 33.5.

(a) $$p_X(x) = \exp\{6 \log x - 11x + \theta_3 x^2 - (x^3/3) - \eta\}$$

(b) $$p_X(x) = \exp\{6 \log x - 5x + \theta_3 x^2 - (x^3/3) - \eta\}$$

Sobolev (1994) introduced the following family of life distributions, termed $q - r$ distributions, and investigated its properties as well as estimation procedures:

$$p(x; p, q, r) = |q| \frac{p^q x^{q-1}}{\Gamma(1 + \frac{1}{2}qr)} e^{-(px)^{2/r}}, \qquad x \geq 0,$$

where p is a scale parameter. He considered the following cases: 1. $q, r > 0$; 2. $q, r > 0$ and $r \to 0$; 3. $q, r < 0$; and 4. $q, r < 0$ and $|r| \to 0$. Case 2 yields

$$p(x) = qp^q x^{q-1}, \qquad 0 \leq x \leq \frac{1}{p} \quad \text{(power function)}$$

while Case 4 yields

$$p(x) = qp^q x^{-q-1}, \qquad p \leq x \leq \infty \quad \text{(Pareto)}$$

Values of (q, r) equal to $(-\frac{1}{2}, -2)$, $(1, 0)$, $(1, 1)$, $(2, 1)$, $(3, 1)$, $(2, 2)$ and $\left(\frac{2}{r}, r\right)$ correspond respectively to stable Levy, rectangular, folded normal, Rayleigh, Maxwell, exponential, and Weibull distributions. Note the resemblance with the generalized distributions of McDonald (see Chapter 27) and with the generalizations of the gamma distribution (see Chapter 17). The modified Laplace transform of the Sobolev family

$$\int_0^\infty p(x) \exp\{-(sx)^{2/r}\} dx = p\left[(p^{2/r} + s^{2/r})^{r/2}\right]^{-q}$$

for Case 1, and similarly for the other cases it is of the form ψ^q, ψ being a function of the remaining parameters.

3 BIRNBAUM-SAUNDERS DISTRIBUTIONS AND TRANSFORMATIONS

In this section we will consider some other distributions that have been suggested to represent lifetime based on more or less realistic conjectures about the mechanism leading to conclusion of a "life." This treatment is far from exhaustive.

We have already discussed a number of simple transformations of normal variables in Chapter 12, from the viewpoint of construction of general systems of distributions. Birnbaum and Saunders (1968a) were led to a discussion of the distribution of

$$T = \beta\left[\tfrac{1}{2}U\alpha + \sqrt{\left(\tfrac{1}{2}U\alpha\right)^2 + 1}\right]^2 \tag{33.14}$$

(where β and α are positive parameters and U is a unit normal variable) from a model representing time to failure of material subjected to a cyclically repeated stress pattern. They supposed that the jth cycle leads to an increase X_j in length of a "crack" and that the sum $\Sigma_{j=1}^{n} X_j$ is approximately normally distributed with expected value $n\mu_0$ and standard deviation $\sigma_0\sqrt{n}$. Then the probability that the crack does not exceed a critical length ω, say, is

$$\Phi\left(\frac{\omega - n\mu_0}{\sigma_0\sqrt{n}}\right) = \Phi\left(\frac{\omega}{\sigma_0\sqrt{n}} - \frac{\mu_0\sqrt{n}}{\sigma_0}\right).$$

It was supposed that "failure" occurs when the crack length exceeds ω. If T denotes the lifetime (in number of cycles) until failure, then the cdf of T is approximately

$$F_T(t) \doteqdot 1 - \Phi\left(\frac{\omega}{\sigma_0\sqrt{t}} - \frac{\mu_0\sqrt{t}}{\sigma_0}\right) = \Phi\left(\frac{\mu_0\sqrt{t}}{\sigma_0} - \frac{\omega}{\sigma_0\sqrt{t}}\right) \tag{33.15}$$

(it being assumed that probability of negative values of X_j's can be neglected). If (33.15) is regarded as an exact equation, then it follows that

$$U = \frac{\mu_0\sqrt{T}}{\sigma_0} - \frac{\omega}{\sigma_0\sqrt{T}} \tag{33.16}$$

has a unit normal distribution. Equation (33.16) can be rewritten

$$T = \frac{\omega}{\mu_0}\left[\frac{U\sigma_0}{2\sqrt{\omega\mu_0}} + \sqrt{\left(\frac{U\sigma_0}{2\sqrt{\omega\mu_0}}\right)^2 + 1}\right]^2, \tag{33.16$'$}$$

which is of the same form as (33.14), with

$$\beta = \frac{\omega}{\mu_0},$$

$$\alpha = \frac{\sigma_0}{\sqrt{\omega\mu_0}}. \qquad (33.17)$$

Thus T has a Birnbaum-Saunders distribution if

$$\frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right) \qquad (33.18)$$

has a unit normal distribution. Clearly cT $(c > 0)$ also has a Birnbaum-Saunders distribution, with parameters βc and a [Birnbaum and Saunders (1969)]. Also the distribution of T^{-1} is the same as that of T, with β replaced by β^{-1} and the parameter a unchanged in value.

From Eq. (33.18), Chang and Tang (1994a, b) recently proposed a simple random variate generation algorithm for the Birnbaum-Saunders distribution as follows: "Suppose Z is a standard normal variable and U is a uniform (0, 1) random variable. Let t_1 and t_2 be the roots of the quadratic equation

$$t^2 - \beta(2 + \alpha^2 z^2)t + \beta^2 = 0,$$

where z is a value of Z. Then, a Birnbaum-Saunders random variable with α and β as its shape and scale parameters, respectively, is given by

$$t_1 I_{(U > 0.5)} + t_2 I_{(U < 0.5)},$$

where $I_{(A)}$ is the indicator function of event A." Observe that the last step is along the lines of Michael, Schucany, and Haas (1976) ensuring that the conditional probability of selecting either of the two roots t_1 and t_2, given that $Z = z$, is 1/2.

The parameter β is simply a multiplier and does not affect the shape of the distribution of T. The moment ratios of T depend only on a, and the (β_1, β_2) points lie on a line. The situation is analogous to that for lognormal distributions [see Chapter 14, Eqs. (14.9)].

The expected value of $(T/\beta)^r$ is

$$E\left[\left(\frac{T}{\beta}\right)^r\right] = \sum_{j=0}^{r}\binom{2r}{2j}E\left[\left\{\left(\frac{1}{2}U\alpha\right)^2+1\right\}^j\left(\frac{1}{2}U\alpha\right)^{2(r-j)}\right]$$

$$= \sum_{j=0}^{r}\binom{2r}{2j}\sum_{i=0}^{j}\binom{j}{i}E\left[U^{2(r-j+i)}\right]\left(\frac{1}{2}\alpha\right)^{2(r-j+i)}$$

$$= \sum_{j=0}^{r}\binom{2r}{2j}\sum_{i=0}^{j}\binom{j}{i}\frac{[2(r-j+i)]!}{2^{r-j+i}(r-j+i)!}\left(\frac{1}{2}\alpha\right)^{2(r-j+i)}. \qquad (33.19)$$

Note that $E[(\frac{1}{2}U\alpha)^s\{(\frac{1}{2}U\alpha)^2+1\}^r] = 0$ if s is odd. From (33.19) we find that

$$E[T] = \beta\left(\frac{1}{2}\alpha^2+1\right), \qquad (33.20a)$$

$$\operatorname{var}(T) = \beta^2\alpha^2\left(\frac{5}{4}\alpha^2+1\right), \qquad (33.20b)$$

$$\beta_1(T) = \frac{16\alpha^2(11\alpha^2+6)}{(5\alpha^2+4)^3}, \qquad (33.20c)$$

$$\beta_2(T) = 3 + \frac{6\alpha^2(93\alpha^2+41)}{(5\alpha^2+4)^2}. \qquad (33.20d)$$

As α tends to zero, the distribution tends to normality. The ratio

$$\frac{\beta_2(T)-3}{\beta_1(T)}$$

is remarkably stable, varying between *2.88* (as $\alpha \to \infty$) and *3.42* (as $\alpha \to 0$).

T is a monotonically increasing function of U. The median of T corresponds to $U = 0$ and is equal to β. Since

$$\left(U+\sqrt{U^2+1}\right)^{-1} = (-U)+\sqrt{(-U)^2+1},$$

it follows that the relation

$$T_\alpha T_{1-\alpha} = \beta^2 \qquad (33.21)$$

holds between the lower and upper $100\alpha\%$ points of the distribution of T.

The average hazard rate $\nu(t) = (1/t)\int_0^t h(s)\,ds$ of the Birnbaum-Saunders distribution approaches a constant as $t \to \infty$. This rate is nearly non-decreasing. Mann, Schafer, and Singpurwalla (1974) provide graphs of the density and the corresponding $\nu(t)$. A comparison between the hazard rates of the Birnbaum-Saunders and the lognormal distributions is given in Nelson *(1990).* While the hazard rate of Birnbaum-Saunders is zero at $t = 0$, then increases to a maximum for some t_0 and finally decreases to a finite value, the hazard rate of the lognormal distribution decreases to zero (see Chapter *14,* Section 1).

An alternative expression for the Birnbaum-Saunders distribution prevalent in the literature is

$$F_T(t;\alpha,\beta) = \Phi\left[\frac{1}{\alpha}\xi\left(\frac{t}{\beta}\right)\right], \quad t > 0, \tag{33.21$'$}$$

where $\xi(t) = t^{1/2} - t^{-1/2}$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. [See also the remarks made after Eq. (33.22).]

The pdf of T is

$$p_T(t;\alpha,\beta) = \frac{\exp(\alpha^{-2})}{2\alpha\sqrt{(2\pi\beta)}}t^{-3/2}(t+\beta)\exp\left\{-\frac{1}{2\alpha^2}\left(\frac{t}{\beta}+\frac{\beta}{t}\right)\right\},$$
$$t > 0;\ \alpha,\beta > 0. \tag{33.22}$$

The alternative expression for the probability density function corresponding to *(33.21)'* is

$$p_T(t;\alpha,\beta) = (\alpha\beta)^{-1}(2\pi)^{-1/2}\xi'\left(\frac{t}{\beta}\right)\exp\left\{-\frac{1}{2\alpha^2}\xi^2\left(\frac{t}{\beta}\right)\right\},$$
$$t > 0, \quad \alpha > 0, \quad \beta > 0. \tag{33.22$'$}$$

Desmond *(1985, 1986)* noted that this distributional form, derived by Birnbaum and Saunders (1969), had been previously obtained by Freudenthal and Shinozuka *(1961)* with a somewhat different parametrization.

He observed that this model is based on the following assumptions about the "fatigue" process:

1. Fatigue failure is due to repeated application of a common cyclic stress pattern.
2. Under the influence of this cyclic stress a dominant crack in the material grows until a critical size w is reached at which point fatigue failure occurs.
3. The crack extensions in each cycle are random variables with the same mean and variance.
4. The crack extensions in each cycle are statistically independent.
5. The total extension of the crack, after a large number of cycles, is approximately normally distributed by the central limit theorem.

It was noted above that the crack length is assumed to be normally distributed in Birnbaum and Saunders' original derivation, so it can, in principle, take on negative values with nonzero probability. This possibility is ignored in the derivation. Desmond (1985) provided a more general derivation of the Birnbaum-Saunders distribution without assuming normality for the distribution of crack length:

1. A variety of distributions for crack size (many of which are on the positive real line) are possible which still result in a Birnbaum-Saunders distribution for the fatigue failure time.
2. It is possible to allow the crack increment in a given cycle to depend on the total crack size at the beginning of the cycle and still obtain a fatigue life distribution of the Birnbaum-Saunders type.

Desmond (1985) also derived the Birnbaum-Saunders distribution, using a biological model due to Cramér (1946, p. 219) of which the "law of proportional effects" [see, e.g., Mann, Schafer, and Singpurwalla (1974)] is a special case. He points out that the application of Cramér's argument in the fatigue context leads to a Birnbaum-Saunders type distribution, rather than a lognormal distribution, as suggested by Birnbaum and Saunders (1969) and Mann, Schafer, and Singpurwalla (1974).

He further provided failure models in random environments described by stationary continuous-time Gaussian processes, for which the Birnbaum-Saunders distribution is appropriate. These include failure due to the response process being above a fixed level for too long a period of time. The damage in an excursion above the fixed level is related to the so-called Z_n-exceedance measures of Cramér and Leadbetter (1967). (The case $n = 1$ corresponds to using areas cut off by the process above the fixed level of measures of damage.) Fatigue failure may also be due to the stress history, that is proportional to the response of a lightly damped single degree of freedom oscillator excited by stationary Gaussian white noise. All of these models lead to distributions of the Birnbaum-Saunders type. Setô, Iwase, and Oohara (1993), analyzing the rainfall characteristics of the city of Hiroshima, reached the conclusion that for these data the distribution of periods of continuous rainfall is best fitted by a Birnbaum-Saunders distribution with a cdf of form (33.15). Their analysis employs a so-called Tsukatani-Shigemitsu test which is based on the (δ_2, δ_3) chart where

$$\delta_2 = \frac{\mu_2}{(\mu_1')^2} \tag{33.23a}$$

and

$$\delta_3 = \frac{\mu_3}{\mu_1'\mu_2}, \tag{33.23b}$$

μ_i being the ith moment about the mean μ'_1 ($i = 2, 3$). For the distribution (33.22), these parameters are functions of α^2 only:

$$\delta_2 = \frac{5\alpha^4 + 4\alpha^2}{(\alpha^2 + 2)^2}, \tag{33.24a}$$

$$\delta_3 = \frac{44\alpha^4 + 24\alpha^2}{(5\alpha^2 + 4)(\alpha^2 + 2)}. \tag{33.24b}$$

Saunders and Birnbaum (1969) have proposed the following methods for calculating maximum likelihood estimators of the parameters. The maximum likelihood estimator of β is the unique positive root of

$$\hat{\beta}^2 - \hat{\beta}\{2H + K(\hat{\beta})\} + H\{\bar{T} + K(\hat{\beta})\} = 0, \tag{33.25}$$

where

$$\bar{T} = n^{-1}\sum_{j=1}^{n} T_j,$$

$$H = \left[n^{-1}\sum_{j=1}^{n} T_j^{-1}\right]^{-1},$$

$$K(\hat{\beta}) = \left[n^{-1}\sum_{j=1}^{n}\left(\hat{\beta} + T_j\right)^{-1}\right]^{-1}.$$

[See also Achcar (1993).] Having calculated $\hat{\beta}$, the maximum likelihood estimator $\hat{\alpha}$ of a can be obtained directly from the formula

$$\hat{\alpha} = 2\left[\tfrac{1}{2}\left(\bar{T}\hat{\beta}^{-1} + H^{-1}\hat{\beta}\right) - 1\right]^{1/2}. \tag{33.26}$$

As an initial value to use in the iterative solution of (33.25), the"mean mean"

$$(\bar{T}H)^{1/2} = \left\{\left(\sum_{i=1}^{n} T_i\right)\left(\sum_{i=1}^{n} T_i^{-1}\right)^{-1}\right\}^{1/2} \tag{33.27}$$

is suggested. It is further shown that if

$$2\bar{T} < 3H + \min(T_1, T_2, \cdots, T_n),$$

the Newton-Raphson iteration method will converge to $\hat{\beta}$ (for *any* initial value between $\bar{T}$ and H). If 2H > $\bar{T}$, then iterative calculation of

$$H + \tfrac{1}{2}K(\hat{\beta}) - \sqrt{\tfrac{1}{4}\left[K(\hat{\beta})\right]^2 - H(\bar{T} - H)}$$

will converge to $\hat{\beta}$.

The "mean mean," $(\bar{T}H)^{1/2}$, referred to above, has approximate variance (for n large)

$$n^{-1}\left[\tfrac{1}{8}\left(1 + \tfrac{3}{8}\alpha^2\right)\left(1 + \tfrac{1}{4}\alpha^2\right)^{-1}\beta^2\alpha^2\right]. \tag{33.28}$$

The expected value is approximately

$$\beta\left[1 + \tfrac{1}{4}\left(1 + \tfrac{3}{8}\alpha^2\right)\left(1 + \tfrac{1}{4}\alpha^2\right)^{-1}\alpha^2 n^{-1}\right].$$

For a < $\sqrt{2}$ it is suggested that $(\bar{T}H)^{1/2}$ can be used in place of the maximum likelihood estimator of β.

Ahmad (1988) noted that while the "mean-mean" estimator is asymptotically unbiased, it does have a positive bias in finite size samples. He suggested the following jackknife estimator, based on a random sample $T_1, \cdots, T_n$ of size n = mg (m and g being integers). Divide the sample into m groups B,,. . ., B„ of size g each and calculate

$$\beta_i^* = \frac{\Sigma_{(i)} T_j}{\Sigma_{(i)} T_j^{-1}}, \tag{33.29}$$

where $\Sigma_{(i)}$ denotes summation over all $j = 1, \dots, n$ except j = i (i = 1,. . . , m). The jackknife estimator is

$$\beta^* = m\beta_0^* - (1 - m^{-1}) \sum_{i=1}^{m} \beta_i^*, \tag{33.30}$$

where

$$\beta_0^* = \left(\frac{\sum_{j=1}^{n} T_j}{\sum_{j=1}^{n} T_j^{-1}} \right)^{1/2} \qquad \text{[the "mean mean;" see (33.27)]}.$$

Ahmad (1988) also showed that β^*, as defined in **(33.30)**, is a consistent estimator of β (as is also the mean-mean estimator). He evaluated the mean square errors of both estimators, concluding that β^* has the smaller error, although both have the same limiting distribution.

An alternative approach, based on the idea of equating first moments, leads to a simple estimator. Since $\alpha^{-1}\{(T/\beta)^{1/2} - (\beta/T)^{1/2}\}$ has a unit normal distribution [see (33.18)],

$$E\left[\left(\frac{T}{\beta}\right)^{1/2} - \left(\frac{\beta}{T}\right)^{1/2}\right] = 0.$$

Equating the observed average value

$$n^{-1}\sum_{i=1}^{n}\left[\left(\frac{T_i}{\beta}\right)^{1/2} - \left(\frac{\beta}{T_i}\right)^{1/2}\right]$$

to zero, we obtain the estimator

$$\beta' = \frac{\sum_{i=1}^{n} T_i^{1/2}}{\sum_{i=1}^{n} T_i^{-1/2}} \qquad \text{[cf. (33.27)].} \tag{33.31}$$

Corresponding to any estimator β_e, say, of β it is natural to use the "moment estimator" of a:

$$\begin{aligned}\alpha_e &= \left[n^{-1}\sum_{i=1}^{n}\left\{\left(\frac{T_i}{\beta_e}\right)^{1/2} - \left(\frac{\beta_e}{T_i}\right)^{1/2}\right\}^2\right]^{1/2} \\ &= \left\{n^{-1}\sum_{i=1}^{n}\left(\beta_e^{-1}\sum_{i=1}^{n}T_i + \beta_e\sum_{i=1}^{n}T_i^{-1}\right) - 2\right\}^{1/2}.\end{aligned} \tag{33.32}$$

Note that, if the value of β is known, the statistic obtained by replacing β_e by β in (33.32) is a maximum likelihood estimator of a, distributed as $\alpha\chi_n/\sqrt{n}$.

Comparing (33.22) with Eq. (15.4a) of Chapter 15 (and reparametrizing the latter with $\mu = \beta$, A $= \beta\alpha^{-2}$), we see that there is a clear relationship between the pdfs of Birnbaum-Saunders and inverse Gaussian distributions. In fact the Birnbaum-Saunders pdf (33.22) is a mixture (with equal weights) of the inverse Gaussian $IG(\beta, \beta\alpha^{-2})$ pdf and the pdf of the reciprocal $IG(\beta, \beta^{-1}\alpha^2)$ of this variable. This result was noted by Desmond (1985, 1986), and by Jørgensen, Seshadri, and Whitmore (1991), who have noted that the Birnbaum-Saunders distribution belongs to a two-parameter exponential family [contrary to the opinion of Bhattacharyya and Fries (1982)], but it is not an exponential dispersion model.

Engelhardt, Bain, and Wright (1981) point out that the pdf (33.22) satisfies regularity conditions that ensure that the maximum likelihood estimators

$(\hat{\alpha}, \hat{\beta})$ of (α, β) have a joint asymptotically bivariate normal distribution with mean (α, β) and asymptotic variance-covariance matrix

$$\|nB_{ij}(\alpha, \beta)\|^{-1}, \qquad i = 1, 2, j = 1, 2, \tag{33.33}$$

where

$$B_{11}(\alpha, \beta) = 2\alpha^{-2},$$
$$B_{12}(\alpha, \beta) = B_{21}(\alpha, \beta) = 0,$$
$$B_{22}(\alpha, \beta) = \beta^{-2}\{\tfrac{1}{4} + \alpha^{-2} + I(\alpha)\}$$

with

$$I(\alpha) = 2\int_0^\infty \left[\{1 + \xi^{-1}(\alpha z)\}^{-1} - \tfrac{1}{2}\right]^2 \phi(z)\, dz, \qquad \phi(z) = \Phi'(z),$$

where, as before, $\xi(t) = t^{1/2} - t^{-1/2}$. Engelhardt, Bain, and Wright (1981) utilized this result to construct large-sample inference procedures on α and β. The variable $\sqrt{n}\,(\hat{\alpha}\alpha^{-1} - 1)$ is approximately a pivotal statistic (has a parameter-free distribution) as also is $n^{1/2}(\hat{\beta}\beta^{-1} - 1)\{\frac{1}{4} + \hat{\alpha}^{-2} - I(\hat{\alpha})\}^{-1/2}$. Hence approximate independent $100(1 - \varepsilon)\%$ confidence intervals

$$\left(\hat{\alpha}\left\{1 + (2n)^{-1/2}u_{1-\frac{\varepsilon}{2}}\right\}^{-1}, \hat{\alpha}\left\{1 + (2n)^{-1/2}u_{\frac{\varepsilon}{2}}\right\}^{-1}\right) \qquad \text{for } \alpha \tag{33.34a}$$

and

$$\left(\hat{\beta}\left[1 + n^{-1/2}\left\{(1 + (\tfrac{1}{4} + \hat{\alpha}^{-2} - I(\hat{\alpha}))^{-1/2}u_{1-\frac{\varepsilon}{2}}\right\}\right]^{-1},\right.$$
$$\left.\hat{\beta}\left[1 + n^{-1/2}\left\{1 + (\tfrac{1}{4} + \hat{\alpha}^{-2} - I(\hat{\alpha}))^{-1/2}u_{\frac{\varepsilon}{2}}\right\}\right]^{-1}\right) \qquad \text{for } \beta, \tag{33.34b}$$

can be constructed. Engelhardt, Bain, and Wright (1981) also discussed some asymptotic results.

Achcar (1993) has developed Bayesian estimation procedures based on two "noninformative" joint prior distributions for α and β, with the pdfs

$$p_{\alpha,\beta}(a, b) \propto (ab)^{-1}, \qquad a, b > 0, \tag{33.35}$$

and

$$p_{\alpha,\beta}(a, b) \propto (ab)^{-1}(\tfrac{1}{4} + a^{-2})^{1/2}, \qquad a, b > 0. \tag{33.36}$$

(Both distributions are improper.) For (33.36) the joint posterior distribution of α and β, given $\mathbf{T} = (T_1, \cdots, T_n)$, has the pdf

$$p_{\alpha,\beta}(a, b|\mathbf{T}) \propto a^{-(n+1)} b^{-(n/2)-1} \left(\tfrac{1}{4} + a^{-2}\right)^{1/2} \left\{ \prod_{i=1}^{n} (b + T_i) \right\}$$
$$\times \exp\left\{ -\tfrac{1}{2} a^{-2} \left(b^{-1} \sum_{i=1}^{n} T_i + b \sum_{i=1}^{n} T_i^{-1} - 2n \right) \right\}, \qquad a, b > 0. \tag{33.37}$$

The corresponding pdf in case (33.35) is obtained by omitting the factor $(\frac{1}{4} + a^{-2})^{1/2}$.

The posterior density function of β, for case (33.35), is

$$p_\beta(b|\mathbf{T}) \propto b^{-(n/2)-1} \left\{ \prod_{j=1}^{n} (b + T_i) \right\} \left(b^{-1} \sum_{i=1}^{n} T_i + b \sum_{i=1}^{n} T_i^{-1} - 2n \right)^{-n/2}. \tag{33.38}$$

For the case (33.36) Achcar (1993) obtained an ***approximate*** formula for $p_\beta(b|\mathbf{T})$ by multiplying the right-hand side of *(33.38)* by

$$\left\{ 4 + (n+2)^{-1} \left(b^{-1} \sum_{i=1}^{n} T_i + b \sum_{i=1}^{n} T_i^{-1} - 2n \right) \right\}^{1/2}. \tag{33.39}$$

The posterior density of α is of simpler form, being proportional to

$$a^{-n} \exp\left[-na^{-2} \left\{ \left(\sum_{i}^{n} T_i \right)^{1/2} \left(\sum_{i}^{n} T_i^{-1} \right)^{-1/2} - 1 \right\} \right], \qquad a > 0, \tag{33.40}$$

for case (33.35), and ***approximately*** proportional to

$$a^{-(n+1)} \left(\tfrac{1}{4} + a^{-2}\right)^{1/2} \exp\left[-na^{-2} \left\{ \left(\sum_{i=1}^{n} T_i \right)^{1/2} \left(\sum_{i=1}^{n} T_i^{-1} \right)^{-1/2} - 1 \right\} \right],$$
$$a > 0, \tag{33.41}$$

for case *(33.36)*. If β is known, Achcar (1993) suggested using a prior gamma distribution (Chapter 17) for α^{-2}. This had been discussed by Padgett *(1982)* in connection with Bayesian estimation of the reliability function corresponding to *(33.22)*. Padgett also discussed use of an improper prior for α, with $p_\alpha(a)$ proportional to a^{-1}.

If T is distributed as (33.22), and $Y = \log T$ (i.e., $T = e^Y$), then

$$\alpha^{-1}(\beta^{-1/2} e^{Y/2} - \beta^{1/2} e^{-Y/2}) = 2\alpha^{-1} \sinh\{\tfrac{1}{2}(Y - \gamma)\},$$

where $\gamma = \log \beta$, has a unit normal distribution. This is a special case $(\sigma = 2)$ of the ***sinh-normal*** distribution [denoted $\mathrm{SN}(\alpha, \mathrm{y}, \sigma)$] introduced by Rieck and Nedelman (1991) as the distribution of Z if

$$2\alpha^{-1}\, sinh\{\tfrac{1}{2}(Z - \gamma)/\sigma\}$$

has a unit normal distribution.

The pdf of Z is

$$p_Z(z) = 2\left(\alpha\sigma\sqrt{2\pi}\right)^{-1}\cosh\left\{\frac{z-\gamma}{\sigma}\right\}\exp\left[-2\alpha^{-2}\sinh^2\left\{\frac{z-\gamma}{\sigma}\right\}\right]. \quad (33.42)$$

Rieck and Nedelman (1991) noted that $\mathrm{SN}(\alpha, \gamma, \sigma)$ distributions have the following properties:

1. The distribution is symmetric about the location parameter y.
2. The distribution is strongly unimodal for a ≤ 2 and bimodal for $\alpha > 2$.
3. The mean and variance are given by $E[Z] = \gamma$ and $\mathrm{var}(Z) = \sigma^2\omega(\alpha)$, where $\omega(\alpha)$ is the variance when $\sigma = 1$. There is no closed-form expression for $\omega(\alpha)$, but Rieck (1989) provided asymptotic approximations to it for small and large a.
4. If $Z_\alpha \sim \mathrm{SN}(\alpha, \gamma, \sigma)$, then $S_\alpha = 2(Z_\alpha - \gamma)/\sigma$ converges in distribution to the standard normal distribution as a approaches 0.

Rieck and Nedelman (1991) discussed maximum likelihood and least-squares estimation of the model

$$Y_i = \mathbf{x}_i'\boldsymbol{\theta} + Z_i, \quad i = 1, 2, \ldots, n, \quad (33.43)$$

where $Z_i \sim \mathrm{SN}(\alpha, 0, 2)$, $\mathrm{x}_i' = (x_{i1}, x_{i2}, \cdots, x_{ip})$ are p explanatory variables and 0' $= (\theta_1, \theta_2, \cdots, \theta_p)$ is a vector of unknown parameters to be estimated. The MLE of α^2 is given by

$$\hat{\alpha}^2 = \left(\frac{4}{n}\right)\sum_{i=1}^{n}\sinh^2\left[\frac{Y_i - x_i'\hat{\boldsymbol{\theta}}}{2}\right], \quad (33.44)$$

where $\hat{\boldsymbol{\theta}}$ is the MLE for the vector 0. A numerical procedure must be used to determine 6. The asymptotic variance of $\hat{\alpha}$ is $\hat{\alpha}^2/(2n)$, and the vector $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$ [where $\boldsymbol{\eta} = (\boldsymbol{\theta}, \alpha)'$] converges in distribution to a multinormal with expected value vector $\mathbf{0}$.

The bimodality of the sinh-normal density when a > 2 can cause multiple maxima of the likelihood. Rieck (1989) provided an example with three least-squares solutions to the likelihood equations; two solutions were maxima, and the third a saddle point.

Experience with aircraft-engine data indicates that the case a > 2 is unusual in practice, and Rieck (1989) showed that if a < 2 the MLE of 0 is

Table 33.1 Least-squares efficiency for values of a

		Monte Carlo Results ($p = 2$)			
		θ_1		θ_2	
α	Efficiency	$n = 10$	$n = 20$	$n = 10$	$n = 20$
0.5	0.99	1.00	1.00	0.99	1.00
1.5	0.81	0.88	0.84	0.89	0.84

unique if $X' \equiv (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ has rank p. Rieck and Nedelman (1991) also investigate a "reduced biased MLE" estimator of a given by

$$\alpha^2 = [n/\{n - pA(\hat{\alpha})\}]\hat{\alpha}^2, \tag{33.45}$$

where

$$A(\alpha) = 2(2 + \alpha^2)(4 + \alpha^2)^{-1}$$

for small a. The relative error for values of a less than 0.5 is less than 0.3%; $A(\alpha)$ approaches 2 as a approaches infinity. The authors also investigated least-squares estimation of model (33.43). They found that the least-squares estimator (LSE) of θ is not as efficient as the MLE but provides an unbiased estimator for θ and that it is highly efficient for small values of a, as indicated in Table 33.1 (for p = 2), based on simulations that compare small-sample relative efficiency of LSE and MLE. The LSE of β is given by the classical equation:

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

where Y is the column vector of the observations (Y_i's) and X is as previously defined.

Achcar and Espinosa (1993) have discussed Bayesian estimation for model (33.43) when the Z's have a Birnbaum-Saunders distribution.

Chang and Tang (1994) have discussed the construction of confidence intervals for the $100p^{\text{th}}$ percentile of the Birnbaum-Saunders distribution, and then the determination of conservative two-sided tolerance limits using these confidence limits. These authors have also described how these results could be used for the reliability evaluation when using the Birnbaum-Saunders model.

Chaudhry and Ahmad (1993) have recently introduced a distribution with the pdf

$$p_T(t) = 2\sqrt{\frac{\alpha}{\pi}}\exp\left[-\left(t\sqrt{\alpha} - t^{-1}\sqrt{\beta}\right)^2\right], \qquad \alpha, \beta > 0, t > 0. \tag{33.46}$$

The mode is at I = $(\beta/\alpha)^{1/4}$. This is, in fact, the distribution of $1/\sqrt{Y}$, when Y has an inverse Gaussian distribution.

Ahmad and Chaudhry (1993) have exploited this relationship inter alia to obtain the characterization that given $T_1, \cdots, T_n$ independent and identically distributed, such that $E[T^r]$ exists for $r = -4, -2$, and 2, and that $\{E[T^{-2}]\}^{-1}$ exists, a necessary and sufficient condition for $\sum_{i=1}^n T_i^{-2}$ and $\sum_{i=1}^n T_i^2 - n^2(\sum_{i=1}^n T_i^{-1})^{-1}$ to have distribution (33.46) is that they be independent. Compare these with characterizations of the inverse Gaussian distribution presented in Chapter 15.

4 ORDERING AND CLASSIFICATION OF DISTFUBUTIONS

4.1 Basic Definitions and Bounds

In this section we will only discuss ordering among independent random variables, although the general topic was initiated, to a large extent, in a classical paper of Lehmann (1966) (and independently in the pioneering work of van Zwet in 1964) where it was associated with dependence properties of two (or more) random variables. (This aspect will find a place in our planned volume on *multivariate* distributions.)

It should be emphasized here that the number of various classifications and orderings of distributions is legion. Many of them are only loosely coordinated and a certain amount of confusion and duplication is inevitable. A systematic ordering and classification of various orderings is very much needed, perhaps utilizing modern computer facilities. The books by Pečarić, Proschan, and Tong (1992) and Shaked and Shanthikumar (1994), mentioned in the beginning of this chapter, are useful contributions in this direction.

An original aim of development of some of these orderings was to replace overreliance on moment functions—standard deviations, skewness $(\sqrt{\beta_1})$, and kurtosis (β_2), for example—as measures to be used in comparing distributions. However, it might also be said that there has been an unnecessary proliferation of varieties of ordering, many of doubtful practical value. We have attempted to give a reasonably comprehensive overview of the relatively well-established types of ordering, but it is quite likely that more or less ingenious further kinds will have been developed by the time of publication. We have endeavored to reproduce faithfully the ideas and concepts current in this field but have often not provided thorough critical evaluation.

There is an immense (and occasionally overlapping) literature on ordering, so, once again, we are compelled to be selective in our discussion of references here (we give an additional few at the end of the chapter) and even of basic concepts.

Ordering of life distributions with respect to aging properties has been a popular and fruitful area of investigation during the last 30 years. The discussion in Section 7.2 of Chapter 33 of the first edition of this volume ("Distributions Classified by Hazard Rate") dealt only with IHR (DHR) and IHRA (DHRA) criteria of aging. In the more recent literature it has been fashionable to use slightly different terminology. For completeness, we will provide eight classes of life distributions based on aging concepts and related orderings. Although these definitions are valid for distributions on the whole real line, they are usually restricted to life distributions with $F_X(0) = 0$.

As indicated in Section 2 of this chapter [see Eq. (33.1)] the *failure rate* (referred therein as hazard rate) function is

$$r_{(F)}(x) = -\frac{d}{dx}\log\{1 - F_X(x)\} = p_X(x)\{1 - F_X(x)\}^{-1} = \frac{p_X(x)}{\bar{F}_X(x)}, \tag{33.47}$$

where $\bar{F}_X(x) = 1 - F_X(x)$ is the *survivor function.*

The definitions of orderings also use the *mean residual lifetime*

$$\mu_{(F)}(x) = E[X - x|X > x] = \frac{\int_x^\infty \bar{F}_X(t)\,dt}{\bar{F}_X(x)}. \tag{33.48}$$

In the definitions below the subscript X is omitted. The subscript X will also often be omitted later.

1. $F(x)$ is said to be an *increasing failure rate* (IFR) distribution if $-\log \bar{F}(x)$ is convex. If the density exists, this is equivalent to saying that $r_{(F)}(x)$ is nondecreasing.
2. $F(x)$ is said to be an *increasing failure rate average* (IFRA) distribution if $-\log \bar{F}(x)$ is a star-shaped function, namely, if $-\log \bar{F}(\lambda x) \le -\lambda \log \bar{F}(x)$ for $0 \le \lambda \le 1$ and $x \ge 0$. When the failure rate exists, this is equivalent to saying that $\nu(x) = \int_0^x r_{(F)}(t)\,dt/x$ is nondecreasing.
3. $F(x)$ is said to be a *new better than used* (NBU) distribution if $-\log \bar{F}(x)$ is superadditive, namely, if $-\log \bar{F}(x + y) \ge -\log \bar{F}(x) - \log \bar{F}(y)$, $x, y \ge 0$. This is equivalent to the statement $\Pr[X > x + y | X > x] \le \Pr[X > y]$
4. $F(x)$ is said to be a *decreasing mean residual life* (DMRL) distribution if $\mu_{(F)}(x)$ is nonincreasing.
5. $F(x)$ is said to be a *new better than used in expectation* (NBUE) distribution if $\mu_{(F)}(x) \le \mu_{(F)}(0)$, $x \ge 0$.
6. $F(x)$ is said to be a *harmonic new better than used in expectation* (HNBUE) distribution if $\int_x^\infty \bar{F}(t)\,dt \le \mu_{(F)}(x)\exp\{-x/\mu_{(F)}(x)\}$, $x \ge 0$.
7. For absolutely continuous $F(x)$, with failure rate $r_{(F)}(x)$, we say that $F(x)$ is a *new better than used in failure rate (NBUFR)* distribution if $r_{(F)}(x) \ge r_{(F)}(0)$, $x \ge 0$.
8. For absolutely continuous $F(x)$, we say that $F(x)$ is a *new better than used in failure rate average (NBUFRA)* distribution if

$$r_{(F)}(0) \le \frac{1}{x}\int_0^x r_{(F)}(t)\,dt = \frac{-\log \bar{F}(x)}{x}.$$

Further classes can be defined by replacing B (better) by W (worse), with appropriate changes in the signs of inequalities. Hazard (H) can replace failure (F) in the definitions.

The NBU class [3 above] was extended by Hollander, Park, and Proschan (1986) to a NBU–t_0 class, defined as follows: A life distribution $F(x)$ is NBU of age t_0 (NBU–t_0) if

$$\bar{F}(x + t_0) \leq \bar{F}(x)\bar{F}(t_0) \qquad \text{for all } x \geq 0. \tag{33.49}$$

The dual NWU–t_0 class is defined analogously, reversing the inequality (33.49).

The NBU–t_0 includes the NBU class, and the NWU–t_0 includes the NWU class. There is a detailed discussion in Hollander et al. (1986). For more discussion of IFRA, HNBUE (HNWUE), NBUE (NWUE), and related classes, we refer the reader to Alzaid, Ahmed, and Al-Osh (1987), Block and Savits (1976), Bondesson (1983), Cao and Wang (1990), Klefsjö (1982a, b), and Mehrotra (1981), and there are many others.

As indicated in 1, distributions with increasing hazard rate (IHR distributions) and with decreasing hazard rate (DHR distributions) are those distributions for which $r_{(F)}(x)$ is an increasing or decreasing function of x, respectively. The exponential distribution, with a constant hazard rate, is a natural boundary between these two classes. The half-normal (chi with one degree of freedom) distribution is a DHR distribution. The Weibull distribution [see (21.1), Chapter 21] is IHR if $c > 1$, and DHR if $c < 1$.

Barlow and Marshall (1964, 1965, 1967) gave bounds on the cdfs for IHR and DHR distributions. Some of these bounds are provided in Table 33.2. These bounds can be compared with the well-known Chebyshev-type bounds, which are summarized for convenience in Table 33.3. A few references

Table 33.2 Bounds on distributions classified by hazard rates

Conditions	Values of x	Limits on $1 - F(x)$
IHR	$x \geq \mu_r'^{1/r}$	$1 - F(x) \leq w(x)$, *where*
		$\mu'_r = rx^r \int_0^1 t^{r-1}[w(x)]^t\,dt$,
		i.e., $\lambda_r = (-x/\log w(x))^r(\Gamma_{-\log w(x)}(r)/\Gamma(r))$
	$x \leq \mu_r'^{1/r}$	$1 - F(x) \geq \exp(-x/\lambda_r^{1/r})$
DHR	$x \leq r\lambda_r^{1/r}$	$1 - F(x) \geq \exp(-x/\lambda_r^{1/r})$
	$x \geq r\lambda_r^{1/r}$	$1 - F(x) \geq (rx/e)^r\lambda_r$
IHRA	$x < \mu_r'^{1/r}$	$1 - F(x) \geq \exp(-bx)$,
		where b satisfies
		$\mu'_r = x^r(1 - e^{-bx}) + b\int_x^\infty t^r e^{-bt}\,dt$

Note: $\lambda_r = \mu'_r/\Gamma(r+1)$.

Barlow and Marshall (1967) gave limits for $F(x_2) - F(x_1)$ for any $x_2 > x_1 > 0$ for each of the above cases. For the case where $r = 1$, and when both the first and second moments are known, better, but considerably more complicated, inequalities are available in Barlow and Marshall (1964, 1965).

4.2 Reliability Classification of Orderings

From the classification described in the previous section, we have the following orderings (which overlap with other orderings introduced from different viewpoints by various authors):

1. $F \overset{\text{IFR}}{<} G$ iff $G^{-1}\{F(x)\}$ is convex [$F(0) = G(0) = 0$ and G is strictly increasing on its support which is a single interval.] This is also equivalent to concex ordering denoted by $F \overset{c}{<} G$. When the densities exist, an equivalent formulation is

$$\frac{r_{(F)}(F^{-1}(u))}{r_{(G)}(G^{-1}(u))} \quad \text{is nondecreasing for u in } [0,1].$$

2. $F \overset{\text{IFRA}}{<} G$ iff $G^{-1}\{F(x)\}$ is star-shaped. This is equivalent to star-ordering and is also denoted $F \overset{*}{<} G$.
3. F is superadditice (*subadditive*) with respect to G [denoted $F \overset{su}{<} G$ $(F \overset{su}{<} G)$] if

$$G^{-1}[F(x+y)] \geq (\leq) G^{-1}[F(x)] + G^{-1}[F(y)]$$

for all x and y in the support of F.

4. $F \overset{\text{NBU}}{<} G$ iff $G^{-1} * F(x)$ is superadditive. This is equivalent to *superadditice* ordering (see 3 above) and is also denoted as $F \overset{su}{<} G$. The relationships among these orderings are expressed by the following scheme of implications:

$$F \overset{\text{IFR(c)}}{<} G \Rightarrow F \overset{\text{IFRA}(*)}{<} G \Rightarrow F \overset{\text{NBU(su)}}{<} G.$$

Kochar and Wiens (1987) and Kochar (1989) define a *(M)DMRL* (more decreasing in mean residual *life*) ordering [for distribution functions F on $[0, \infty)$ with $F(0) = 0$] as

$$F \overset{\text{DMRL}}{<} G \quad \text{if and only if}$$

$$\frac{\mu_{(F)}(F^{-1}(u))}{\mu_{(G)}(G^{-1}(u))} \quad \text{is nonincreasing for u in } [0,1].$$

If $\bar{G}(x) = e^{-x}$, then $F \overset{\text{DMRL}}{<} G$ iff F is a DMRL distribution (see Definition 4 on page 664). Exponential distributions thus act as a reference point for DMRL relationship. The relation between IFR (convex) ordering and DMRL ordering is

$$F \overset{\text{IFR}}{<} G \Rightarrow F \overset{\text{DMRL}}{<} G.$$

5. Kochar and Wiens also introduce ***NBUE*** and ***HNBUE orderings*** (cf. definitions 5 and 6 in the preceding section), using the concept of the ***equilibrium survival function,*** defined as

$$\bar{F}_e(x) = \int_x^\infty \frac{\bar{F}(t)}{\mu_{(F)}(t)}\,dt. \tag{33.51}$$

The hazard rate of $\bar{F}_e(x)$ is

$$\bar{r}_{(F_e)}(x) = \frac{\bar{F}(x)/\mu_{(F)}(x)}{\bar{F}_e(x)} \tag{33.52}$$

[Note that $\bar{r}_{(F)}(x) = \{\mu_{(F)}(x)\}^{-1}$.] Then

$$F \overset{\text{NBUE}}{<} G \quad \text{if and only if} \quad \frac{\bar{r}_{(F)}(F^{-1}(u))}{\bar{r}_{(G)}(G^{-1}(u))} \geq \frac{\mu_{(G)}(u)}{\mu_{(F)}(u)} \tag{33.53}$$

for all u in $[0, 1]$. Equivalently,

$$F \overset{\text{NBUE}}{<} G \quad \text{if and only if} \quad G_e^{-1} * F_e(x) \geq G^{-1} * F(x) \qquad \text{for all } x \geq 0, \tag{33.53$'$}$$

where as above $*$ denotes the convolution operator. Also

$$F \overset{\text{HNBUE}}{<} G \quad \text{if and only if} \quad \frac{G_e^{-1} * F_e(x)}{x} \geq \frac{d}{dx} G_e^{-1} * F_e(x)\Big|_{x=0}$$

$$\left(\equiv \frac{\mu_{(G)}(x)}{\mu_{(F)}(x)}\right) \qquad \text{for all } x \geq 0, \tag{33.54}$$

or equivalently

$$F \overset{\text{HNBUE}}{<} G \quad \text{if and only if} \quad \bar{F}_e(x\mu_{(F)}(x)) \geq \bar{G}_e(x\mu_G(x)) \qquad \text{for all } x \geq 0 \tag{33.54$'$}$$

In general, $F \overset{\text{NBU}}{<} G$ does not imply $F \overset{\text{NBUE}}{<} G$ [see Kochar and Wiens (1987) for a counterexample]. However,

$$\text{if} \quad \bar{G}(x) = e^{-x}, \quad \text{then} \quad F \overset{\text{NBU}}{<} G \Rightarrow F \overset{\text{NBUE}}{<} G$$

$$\text{and also} \quad F \overset{\text{NBUE}}{<} G \Rightarrow F \overset{\text{HNBUE}}{<} G.$$

Relations among DMRL, NBUE, and dispersive (tail) orderings (see Section 4.3) can be expressed in terms of equilibrium distribution functions. They are not particularly revealing [see Kochar (1989) for details].

Mitra and Basu (1994) introduced a family of life distributions, called the NWBUE (NBWUE) family, which includes the IDMRL class of distributions as well as all BFR distributions (see the definition on page 643). Their definition of this family of distributions is as follows: A life distribution F having support on $[0, \infty)$ and finite mean is said to be ***new worse then better than used in expectation*** (NWBUE) ***(new better then worse than used in expectation,*** NBWUE) if there exists a point $x_0 \geq 0$ such that the mean residual life time

$$\mu_{(F)}(x) \begin{cases} \geq (\leq) \mu_{(F)}(0) & \text{for } x < x_0 \\ \leq (\geq) \mu_{(F)}(0) & \text{for } x \geq x_0. \end{cases}$$

Point x_0, referred to as ***change point*** of the distribution function F, need not be unique. However, for a continuous distribution function F, the collection of all change points of an NWBUE (or NBWUE) distribution (say, $C_{(F)}$) is either a singleton or a closed interval. It may be noted that a NWBUE (NBWUE) distribution is NBUE (NWUE) if $0 \in C_{(F)}$ while it is NWUE (NBUE) if $\infty \in C_{(F)}$.

Mitra and Basu (1994) have established some inequalities and used them to derive bounds for moments of a NWBUE distribution.

A close relation between HNBUE and Lorenz ordering [Chapter 12, Eq. (12.8)] is worth noting. Recall that the Lorenz curve of $F(x)$ is

$$L_F(u) = \mu_{(F)}(u) \int_0^u F^{-1}(s)\, ds, \qquad 0 \leq u \leq 1, \tag{33.55}$$

and the Lorenz ordering is

$$F \leq_L G \quad \text{if and only if} \quad L_F(u) \geq L_G(u) \qquad \text{for all } u \text{ in } (0, 1).$$

Kochar (1989) shows that

$$F \overset{\text{HNBUE}}{<} G \quad \text{if and only if} \quad F \leq_L G,$$

namely, the conditions

$$\overline{F}_e(x\mu_{(F)}(x)) \geq \overline{G}_e(x\mu_{(G)}(x))$$

and

$$\mu_{(F)}(u) \int_0^u F^{-1}(s)\, ds \geq \mu_{(G)}(u) \int_0^u G^{-1}(s)\, ds$$

are equivalent.

Another revealing property of HNBUE ordering is that if $F \overset{\text{HNBUE}}{<} G$, then the coefficient of variation of F is less than that of G. (Of course the converse is not universally true.) Arnold and Villasefior (1985) show that Lorenz ordering is preserved under monotonic transformation $g(x)$ if and only if

$$g(x) = ax, \qquad a > 0, x \geq 0,$$

or

$$g(x) = b, \qquad b > 0, x \geq 0,$$

or

$$g(x) = \begin{cases} 0, & x = 0, \\ c, & c > 0, x > 0. \end{cases}$$

Arnold (1991) shows that the only functions $g(X)$ for which

$$g(X) \leq_{\mathrm{L}} X \qquad \text{for all } X$$

are defined by

$$g(x) > 0 \qquad \text{for all } x > 0,$$

and are monotonically nondecreasing for $x \geq 0$, while $g(x)/x$ are monotonically nonincreasing for $x > 0$. Analogously

$$X \leq_{\mathrm{L}} g(X) \qquad \text{for all } X$$

is equivalent to

$$g(x) > 0 \quad \text{for all } x > 0 \text{ and is monotonically nondecreasing}$$

and

$$g(x)/x \text{ is monotonically nondecreasing for } x > 0.$$

Arnold and Villaseñor (1991) study Lorenz orderings among order statistics $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ from random samples of size n for several specific population distributions. For standard uniform distributions (see Chapter 26) they show that

$$X_{i+1:n} \leq_L X_{i:n} \leq_{\mathrm{L}} X_{i:n-1} \tag{33.56a}$$

and also that

$$X_{n+2:2n+3} \leq_{\mathrm{L}} X_{n+1:2n+1}. \tag{33.56b}$$

[Here we are modifying the notation used heretofore. $X_{i:n}$, for example, represents the cdf of $X_{i:n}$.]

Relationship (33.56b) also applies for ***any*** symmetric population distribution with finite range 0 to c (> 0). These results also hold for power-function distributions (see Chapter 25) with the cdf

$$F_X(x) = \left(\frac{x}{c}\right)^a, \qquad 0 \le x \le c,\ a > 0.$$

However, for Pareto distributions with the cdf

$$F_X(x) = 1 - \left(\frac{x}{c}\right)^{-a}, \qquad c \le x,\ a > 1,$$

we have

$$X_{i:n} \le_{\mathrm{L}} X_{i+1:n}, \tag{33.57a}$$

reversing part of (33.56a); we also have

$$X_{i:n} \le_{\mathrm{L}} X_{i+1:n+1}. \tag{33.57b}$$

4.3 Alternative Stochastic Classification of Orderings

An alternative classification of orderings (not specifically geared to life distributions) is presented below, starting with the natural but very rigid ***stochastic*** ordering—which turns out to be related to other types of ordering popularized in the literature on reliability [see, e.g., Barlow and Proschan (1975, 1981)]:

1. It is hard to ascertain when and where the concept of stochastic ordering was introduced into probability theory. One of the earliest references in statistical literature is van Zwet (1964). A natural definition for "F is stochastically less than G" (denoted $F \le_{\mathrm{st}} G$) would be

   $$F(x) \ge G(x) \qquad \text{for all } x.$$

   However, a conventional definition is

   $$E[h(X)|F_X(x) = F(x)] \le E[h(X)|F_X(x) = G(x)] \qquad \text{for all } x, \tag{33.58}$$

   for all nondecreasing functions $h(\cdot)$.

2. F is less than (or equal to) G in hazard-rate sense $(F \leq_{\text{har}} G)$ if

$$\frac{\bar{F}(x)}{\bar{G}(x)} \quad \text{is nonincreasing for x in} \quad [0, G^{-1}(1)].$$

This is also called positioe uniform stochastic ordering by Keilson and Sumita (1992). There are several equivalent definitions:

a. F I,,, G $\Leftrightarrow r_{(F)}(x) \geq r_{(G)}(x)$ (if F and G are absolutely continuous). Note the reversed sign for the hazard rate inequality.
b. F I,,, G $\Leftrightarrow \bar{F} \geq_{TP_2} \bar{G}$, where TP_2 is total *positiuity* of order 2 ordering, defined by

$$\begin{vmatrix} \bar{F}(x) & \bar{F}(y) \\ \bar{G}(x) & \bar{G}(y) \end{vmatrix} \geq 0 \qquad \text{for all } 0 \leq x \leq y.$$

c. F I,,, G $\Leftrightarrow \bar{F}(x|x > y) \leq \bar{G}(x|x > y)$ for all 0 I x I y, where $\bar{F}(x|x > y) = \bar{F}(x + y)/\bar{F}(y)$ and $\bar{G}(x|x > y) = \bar{G}(x + y)/\bar{G}(y)$ are conditional survival functions.

Note that F I,,, G $\Rightarrow$ F $\leq_{st}$ G. See Keilson and Sumita (1992) for details.

3. When the densities $f(x) = F'(x)$ and $g(x) = G'(x)$ exist, we can have likelihood ratio (LR) ordering with F $\leq_{LR}$ G if $f(t)/g(t)$ is a nonincreasing function of t for t $\geq$ 0. Since F $\leq_{LR}$ G implies that for $t_2 > t_1$,

$$f(t_2) \leq \frac{g(t_2) f(t_1)}{g(t_1)},$$

we have

$$r_{(F)}(t) = \frac{f(t)}{\int_t^\infty f(x)\,dx} \geq \frac{f(t)}{\{f(t)/g(t)\}\int_t^\infty g(x)\,dx}$$

$$= \frac{g(t)}{\int_t^\infty g(x)\,dx} = r_{(G)}(t). \qquad (33.59)$$

Therefore

$$\text{F I, G} \Rightarrow F \leq_{\text{har}} G.$$

The converse is not valid.

4. F is stochastically less (more) *variable* than G [denoted F $\le_V$ $(\ge_V)G$] if

$$E\left[h(X)|F_X(x)=F(x)\right] \le E\left[h(X)|F_X(x)=G(x)\right] \quad (33.60)$$

for all nonincreasing (nondecreasing) *convex* integrable functions $h(\cdot)$ with

$$E[X|F_X(x)=F(x)] \le (\ge) E[X|F_X(x)=G(x)] \qquad [\text{cf. } (33.58)].$$

If $E[X|F_X(x)=F(x)] = E[X|F_X(x)=G(x)]$, then (33.60) holds for all convex integrable functions $h(\cdot)$ [Metzger and Riischendorf (1991)].

Equivalently F I, G if and only if

$$\int_x^\infty \overline{F}(t)\,dt \le \int_x^\infty \overline{G}(t)\,dt \qquad \text{for all } x>0. \qquad (33.61)$$

If F $\le_V$ G, it implies that F gives less weight to extreme values than does G. As Metzger and Riischendorf (1991) have pointed out, variability ordering is in fact "a combination of a 'variability' ordering and of the stochastic ordering $\le_{st}$", and "a pure variability ordering" can be obtained by requiring

$$F(x-E[X]|F_X(x)=F(x)) \le_V G(x-[X]|G_X(x)=G(x)). \qquad (33.62)$$

5. *Dispersive* ordering (and related orderings) have been studied by several authors, including Bickel and Lehmann (1976, 1979), Lewis and Thompson (1981), Oja (1981), Shaked (1982), Deshpande and Kochar (1983), Bartoszewicz (1985, 1986), Droste and Wefelmeyer (1985), and Marzec and Marzec (1991a, b, 1993).

As pointed out by D. J. Saunders (1984), dispersive ordering can be described loosely by saying that the graph of one cdf is always steeper than that of the other, where the definition of "steeper" is in terms of the inverse (quantile) function. The following definition is implicit in Lewis and Thompson (1981): G is more dispersed than F ($F \le_{disp} G$) if

$$G(G^{-1}(\alpha)+a) \le F(F^{-1}(\alpha)+a) \qquad (33.63)$$

for all a > 0 and 0 < a < 1. [F^{-1}(a) = inf{t: $F(t) \ge$ a)]. Deshpande and Kochar (1983) showed that this definition is equivalent to

$$G^{-1}(\beta)-G^{-1}(\alpha) \ge F^{-1}(\beta)-F^{-1}(\alpha) \qquad \text{for } 0 < \text{a} < \beta < 1, \qquad (33.63)'$$

an ordering considered by Saunders and Moran (1978) and Saunders (1978).

Lehmann *(1966)* defined F to have a *lighter tail* than G if

$$\frac{F^{-1}(\beta) - F^{-1}(\alpha)}{G^{-1}(\beta) - G^{-1}(\alpha)} \le M \tag{33.64}$$

for some $M > 0$ and all $0 < \alpha < \beta < 1$. An equivalent definition, introduced earlier by Fraser *(1957)* in his pioneering book on *Nonparametric Methods* is that $\{G^{-1}(F(x)) - x\}$ is a nondecreasing function of x. This can be established by putting $\alpha = F(x)$ and $\beta = F(y)$ with $x \le y$ in *(33.63)'*. It was referred to and used by Doksum (1969), in connection with studies on the power of rank tests, who termed it *tail-ordering,* and studied by Yanagimoto and Sibuya (1976), who described it as *G is stochastically more spread than F.* This definition is implicit in Shaked (1982), who dealt with the absolutely continuous case. Shaked's definition is given in terms of sign changes of a function. It was extended to a more general case by Lynch, Mimmack, and Proschan (1983), in the following way: Define $S(x_1, \dots, x_m)$ as the number of sign changes of the sequence $x_1, \dots, x_m$, discarding zero terms, and $S(f)$ as the number of sign changes of the function f on $(-\infty, +\infty)$. Specifically,

$$S(f) = \sup S[f(t_1), \cdots, f(t_m)]$$

over all $t_1 < t_2 < \dots < t_m$ $(m = 2, 3, \dots)$. Also denote $F(x - c)$ as F_c, and $S_c \equiv S(F_c - G)$ for distribution functions F and G. Then

$$F \le_{\text{disp}} G \Leftrightarrow \text{(for each real } c) \tag{33.65}$$

a. $S(F_c - G) \le 1$,
b. if $S_c = 1$, then $F_c - G$ changes sign from $-$ to $+$.

The family of gamma (α, β) distributions (see Chapter 17), with constant β, is dispersively ordered by the shape parameter α. The larger the value of the shape parameter α, the more dispersed is the distribution.

Barlow and Proschan (1975) describe the following orderings for life distribution cdfs F and G [with $F(0) = G(0) = 0$] and G strictly increasing on its support, which is a single interval:

a. F is ***convex-ordered*** with respect to G (denoted $F \overset{c}{<} G$) if $G^{-1}[F(x)]$ is convex.
b. F is *star-ordered* with respect to G (denoted $F \overset{*}{<} G$) if $G^{-1}[F(x)]$ is star-shaped, namely $G^{-1}[F(x)]/x$ increases with x for all x in the support of F.

There is a simple relation between convex- and star-ordering, namely

$$F \overset{c}{<} G \Rightarrow F \overset{*}{<} G. \tag{33.66}$$

Also, $F \overset{*}{<} G \Rightarrow F <_{\text{disp}} G$ under the single condition that

$$\lim_{x \to 0+} \left\{ \frac{G^{-1}[F(x)]}{x} \right\} \geq 1 \tag{33.67}$$

[Sathe (1984)]. Earlier versions of this result [Doksum (1969); Deshpande and Kochar (1983)] require that F and G be absolutely continuous and $F'(0) \geq G'(0) > 0$.

We note the relations

$$F \leq_{\text{st}} G \quad \text{and} \quad F \overset{*}{<} G \Rightarrow F <_{\text{disp}} G \qquad \text{[Sathe (1984)]}$$

and

$$F \leq_{\text{st}} G \quad \text{and} \quad F \overset{c}{<} G \Rightarrow F <_{\text{disp}} G \qquad \text{[Bartoszewicz (1985)]}$$

Bartoszewicz *(1986)* provided relations between hazard rate ordering and dispersive ordering. If F and G are absolutely continuous over their support $t \geq 0$ and $F(0) = G(0) = 0$, then the conditions $r_{(G)}(t) \leq r_{(F)}(t)$ for all $t \geq 0$ and

$$\bar{F}(x|t) = \frac{\bar{F}(t+x)}{\bar{F}(t)} \quad \text{is a decreasing function of } x \text{ for all } t \text{ L } 0$$

imply that $F <_{\text{disp}} G$. Similarly, under the same conditions,

$$F <_{\text{disp}} G \quad \text{and} \quad F \text{ (or } G\text{) is } IFR \Rightarrow r_{(G)}(t) \leq r_{(F)}(t)$$

for all $t > 0$.

Barlow and Proschan *(1975)* obtained the classical result

$$F \overset{c}{<} G \Rightarrow F \overset{*}{<} G \Rightarrow F \overset{su}{<} G. \tag{33.68}$$

Also $F \overset{su}{<} G$ and $F <_{\text{st}} G \Rightarrow F <_{\text{disp}} G$. This connection between superadditive ordering and dispersive ordering is established in Ahmed et al. (1986). [This is actually a combined version of three letters submitted by Alzaid and Ahmed, Bartoszewicz, and Kochar.]

The condition $F <_{\text{st}} G$ is essential, as can be seen from the well known counterexample

$$F(x) = 1 - \exp(-x^2),$$

$$G(x) = 1 - \exp(-x), \qquad x > 0,$$

in which $G^{-1}[F(x)] = x^2$ satisfies the superadditivity condition, while $G^{-1}[F(x)] - x = x(x-1)$ is neither nondecreasing nor nonincreasing in x for all $x \geq 0$. The same example shows that the condition $r_{(F)}(t) \geq r_{(G)}(t)$ does not, in general, imply that $F <_{\text{disp}} G$. However, the condition $F <_{\text{st}} G$ can be replaced by Sathe's condition

$$\lim_{x \to 0+} \left\{ \frac{G^{-1}[F(x)]}{x} \right\} \geq 1 \qquad [\text{Ahmed et al. (1986)}]. \qquad (33.67)'$$

Bagai and Kochar (1986) investigated the relation between TP_2 (or equivalently hazard rate) ordering (see above—the beginning of Section 4.3) and dispersive (tail-) ordering. They showed that if F or G is IFR, then

$$F <_{\text{disp}} G \Rightarrow \bar{F} <_{TP_2} \bar{G} \left(\equiv r_{(G)}(x) \leq r_{(F)}(x)\right).$$

Similarly, if F or G is DFR, then

$$\bar{F} <_{TP_2} \bar{G} \Rightarrow F <_{\text{disp}} G.$$

In particular, if both F and G are exponential distributions (and so both IFR and DFR), then

$$\bar{F} <_{\text{har}} \bar{G} \Leftrightarrow F <_{\text{disp}} G.$$

We note, again, the pivotal position of the exponential distribution in ordering schemes.

Droste and Wefelmeyer (1985) discuss the dispersive ordering and its alternative versions for strongly unimodal distributions [i.e., for the distributions such that their convolution with any unimodal distribution is unirnodal; Ibragimov (1956)].

6. Oja (1981) introduces ordering based on convexity of order k. [A function $f(x)$ is concex of order k, $k = 2, 3, \ldots,$ iff

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_k \\ & & & \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_k^{k-1} \\ f(x_1) & f(x_2) & \cdots & f(x_k) \end{vmatrix} \geq 0 \qquad \text{for all } x_1 < x_2 < \ldots < x_k.$$

If the sign of the inequality is reversed, the function is *concave* of order k. If the derivative exists, $f(x)$ is convex (concave) of order k if and only if $f^{(k)}(x) \geq (\leq)\, 0$ for all x.]

The ordering $F \leq_3 G$ applies if and only if

$$\Delta(x) = G^{-1}[F(x)] - x$$

is convex of order k. The ordering $F \leq_1 G$ is, of course equivalent to $F <_{\text{disp}} G$, while $F \leq_2 G$ is the usual *convex* ordering, introduced by van Zwet (1964). Oja (1981) used the terminology that $F \leq_3 G$ means that F is "not more skewed to the right than G," and showed that

$$F \leq_2 G \Leftrightarrow \frac{R(x_4) - R(x_3)}{R(x_2) - R(x_1)} \geq \frac{x_4 - x_3}{x_2 - x_1}$$

for all $x_1 < (x_2, x_3) \leq x_4$ in the support of F, where

$$R(x) = G^{-1}[F(x)].$$

The $\leq_3$ ordering is a natural extension of IFR ordering for random variables that are not necessarily positive. Indeed the condition "$G^{-1}F$ is convex (concave)" is an alternative definition for IFR (DFR) ordering (cf. 1 in Section 4.2). Oja (1981) asserted that $\leq_3$ ordering is an alternative to ordering with respect to kurtosis, and he provided a motivation based on the convexity of the function

$$\frac{F'(t)}{G'(G^{-1}[F(t)])}. \tag{33.69}$$

7. Another ordering, focusing on tail behavior, is the more recently formulated Parzen ordering. This was suggested explicitly by Alzaid and Al-Osh (1989), motivated by the original paper of Parzen (1979) and its extension and refinement by Schuster (1984). It is based on the concept of the density-*quan*tile function

$$F'[F^{-1}(u)] \qquad \text{for } u \text{ in } [0, 1]. \tag{33.70}$$

Parzen (1979) observed that, as $u \to 1$,

$$F'[F^{-1}(u)] \sim (1 - u)^{\alpha_0}. \tag{33.71}$$

He called the exponent α_0 the right-tail exponent. It is defined by

$$\alpha_0 = \lim_{u \to 1^-} \frac{(1 - u)J(u)}{F'[F^{-1}(u)]}, \tag{33.72a}$$

where

$$J(u) = \frac{-F''[F^{-1}(u)]}{F'[F^{-1}(u)]}.$$

Similarly, as $u \to 0$,

$$F'[F^{-1}(u)] \sim u^{\alpha_1}, \qquad (33.72b)$$

where

$$\alpha_1 = \lim_{u \to 0^+} \frac{uJ(u)}{F'[F^{-1}(u)]}. \qquad (33.72c)$$

Alzaid and Al-Osh (1989) defined an ordering based on the tail-exponent function

$$\alpha_{(F)}(1-u) = \frac{(1-u)J(u)}{F'[F^{-1}(u)]}. \qquad (33.73)$$

For exponential distributions, $\alpha_{(F)}(\cdot) \equiv 1$. Values for some other distributions are

Pareto (Chapter 19)	$p_X(x) = (\beta x^{1+\beta^{-1}})^{-1}$	$\alpha_{(F)}(u) = 1 + \beta$
Logistic (Chapter 23)	$p_X(x) = e^{-x}(1+e^{-x})^{-2}$	$\alpha_{(F)}(u) = 1 - 2u$
Power function (Chapter 25)	$p_X(x) = \beta(1-x)^{\beta-1}$	$\alpha_{(F)}(u) = 1 - \beta$

In view of the relationships

a. $F \overset{c}{<} G \Leftrightarrow \alpha_{(F)}(u) \le \alpha_{(G)}(u)$,
b. F is IFR (DFR) if and only if $\alpha_{(F)}(u) \le (\ge)\ 1$ for all u in [0, 1],

Alzaid and Al-Osh (1989) defined the *Parzen* ordering $F <_P G$ if and only if

$$\lim_{u \to 1^-} \alpha_{(F)}(1-u) < \lim_{u \to 1^-} \alpha_{(G)}(1-u).$$

This ordering is preserved for the order statistics—namely $F <_P G$ if and only if $F_{j:n} <_P G_{j:n}$, where $F_{j:n}, G_{j:n}$ are the cdfs of the jth-order statistic in random samples of size n from F, G, respectively.

We have

$$\text{Uniform} <_P \text{Power}(\beta) <_P \text{Exponential} <_P \text{Pareto}\ (a < 1)$$
$$<_P \text{Cauchy} <_P \text{Pareto}\ (a > 1), \qquad (33.74)$$

an ordering in terms of tail behavior that has some intuitive appeal.

8. We finally note that Metzger and Riischendorf (1991) have introduced a more general concept of ***conditional*** ordering, in which the inequality(s) defining the ordering need apply only on a restricted subset of values of the variable. They applied this for the variability ordering [see (33.60) above] but the technique can be used for most other orderings.

Rojo (1992, 1993) introduced two "pure-tail" orderings:

D-ordering is defined by

$$\begin{cases} F \leq_{\mathrm{D}} G & \text{if } \limsup_{x \to \infty} \dfrac{\bar{F}(x)}{\bar{G}(x)} < \infty, \\ F <_{\mathrm{D}} G & \text{if F ■ } G \text{ but } G \nleq_{\mathrm{D}} \text{F}, \\ F \sim_{\mathrm{D}} G & \text{if } F \text{ ■ } G \text{ and } G \text{ ■ } F. \end{cases} \tag{33.75}$$

q-ordering is defined by

$$\begin{cases} \text{F} \leq_{\mathrm{q}} G & \text{if } \limsup_{u \to 1} \dfrac{F^{-1}(u)}{G^{-1}(u)} < \infty, \\ F <_{\mathrm{q}} G & \text{if } F \text{ ■ } G \text{ but } G \nleq_{\mathrm{q}} \text{F}, \\ F \sim_{\mathrm{q}} G & \text{if F ■ } G \text{ and } G \leq_{\mathrm{q}} \text{F}. \end{cases} \tag{33.76}$$

Although closely related, these two orderings are not identical. In fact q-ordering is location and scale invariant, but D-ordering is not. There is a detailed discussion of relationships between the two orderings in Rojo (1992).

D-ordering is related to hazard rate orderings (definition 2 at the start of this section) in that both depend on the behavior of the function $\bar{F}(x)/\bar{G}(x)$. Rojo (1988) discusses the effects on D-ordering and q-ordering of operations of convolution, finite mixture, and (strictly increasing) transformations. The "weak" orderings $\leq_{\mathrm{D}}$, ■ (also $\sim_{\mathrm{D}}$, $\sim_{\mathrm{q}}$) are preserved under these operations. Although this is often also true for the "strong" orderings $<_{\mathrm{D}}$, $<_{\mathrm{q}}$, some restrictions are needed in convolutions for $<_{\mathrm{D}}$ and in transformations for $<_{\mathrm{q}}$. See Rojo (1993) for details.

By considering two distributions F and G with respective densities f and g, Jorgensen (1994) defined a strict partial ordering by taking F ***has a heauier upper tail than G*** to mean there exists $l > 0$ such that for every a and every $b > 0$ there is an x^* for which

$$l + \frac{f(x)}{1 - F(x)} - \frac{g(a + bx)}{1 - G(a + bx)} < 0 \qquad \text{for every } x > x^*$$

Jorgensen similarly defined the ordering F ***has a heacier lower tail than G.***

REFERENCES

Abouammoh, A. M., and Ahmed, A. N. (1988). The new better than used failure rate class of distributions, ***Advances in Applied Probability,*** **20,** 237–240. (Correction: **20,** 917.)

Abouammoh, A. M., and Ahmed, A. N. (1992). On renewal failure rate classes of life distributions, ***Statistics & Probability Letters,*** **14,** 211–217.

Achcar, J. A. (1993). Inference for the Birnbaum-Saunders fatigue life model using Bayesian methods, ***Computational Statistics & Data Analysis,*** **15,** 367–380.

Achcar, J. A., and Espinosa, M. M. (1993). Bayesian methods in accelerated life tests considering a log-linear model for the Birnbaum-Saunders distribution, ***Manuscript.***

Ahmad, I. A. (1988). Jackknife estimation for a family of life distributions, ***Journal of Statistical Computation and Simulation,*** **29,** 211–223.

Ahmad, M., and Chaudhry, M. A. (1993). Characterizations of a probability function useful in size modeling, ***Pakistan Journal of Statistics, Series A,*** **9,** 75–77.

Ahmed, A. N. (1988). Preservation properties for the mean residual life ordering, ***Statistische Hefte,*** **29,** 143–150.

Ahmed, A. N., Alzaid, A. A., Bartoszewicz, J., and Kochar, S. C. (1986). Dispersive and superadditive ordering, ***Advances in Applied Probability,*** **18,** 1019–1027.

Alzaid, A. A., Ahmed, A. N., and Al-Osh, M. (1987). On the NBAFR (NWAFR) class of life distributions, ***Statistische Hefte,*** **28,** 203–216.

Alzaid, A. A., and Al-Osh, M. (1989). Ordering probability distributions by tail behavior, ***Statistics & Probability Letters,*** **8,** 185–188.

Armero, C., and Bayarri, M. J. (1993). A Bayesian analysis of a queueing system with unlimited service, ***Technical Report No. 93-50,*** Department of Statistics, Purdue University, West Lafayette, IN.

Arnold, B. C. (1987). ***Majorization and the Lorenz Order: A Brief Introduction,*** Lecture Notes in Statistics, vol. **43,** Berlin: Springer-Verlag.

Arnold, B. C. (1991). Preservation and attenuation of inequality as measured by Lorenz order, In ***Stochastic Orders and Decisions under Risk,*** IMS Lecture Notes—Monograph Series, 19, pp. 25–37.

Arnold, B. C., and Groeneveld, R. A. (1993). Skewness and kurtosis orderings: An introduction, In ***Stochastic Inequalities,*** IMS Lecture Notes—Monograph Series, **22,** pp. 17–24.

Arnold, B. C., and Villaseñor, J. A. (1991). Lorenz ordering of order statistics, In ***Stochastic Orders and Decisions under Risk,*** IMS Lecture Notes—Monograph Series, **19,** pp. 38–47.

Averous, J., and Heste, M. (1989). Tailweight and life distributions, ***Statistics & Probability Letters,*** **8,** 381–387.

Bagai, I., and Kochar, S. C. (1986). On tail-ordering and comparison of failure rates, ***Communications in Statistics—Theory and Methods,*** **15,** 1377–1388.

Bain, L. J. (1974). Analysis for the linear failure-rate life-testing distribution, ***Technometrics. 16.*** 551–559.

Bain, L. J., and Engelhardt, M. (1991). *Statistical Analysis of Reliability and Life-testing Models,* 2nd ed., New York: Dekker.

Barlow, R. E. (1968). Some recent developments in reliability theory, *Operations Research Center, University of California,* Berkeley, Report *ORC* 68-19.

Barlow, R. E., and Marshall, A. W. (1964). Bounds for distributions with monotone hazard rate, I, *II, Annals of Mathematical Statistics, 35,* 1234–1257, 1258–1274.

Barlow, R. E., and Marshall, A. W. (1965). Tables of bounds for distributions with monotone hazard rate, *Journal of the American Statistical Association, 60,* 872–890.

Barlow, R. E., and Marshall, A. W. (1967). Bounds on interval probabilities for restricted families of distributions, *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 3,* 229–257.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate, *Annals of Mathematical Statistics, 34,* 375–389.

Barlow, R. E., and Mendel, M. B. (1992). De Finetti-type representations for life distributions, *Journal of the American Statistical Association,* **87,** 1116–1122.

Barlow, R. E., and Proschan, F. (1965). *Mathematical Theory of Reliability,* New York: Wiley.

Barlow, R. E., and Proschan, F. (1966). Tolerance and confidence limits for classes of distributions based on failure rate, *Annals of Mathematical Statistics,* **37,** 1593–1601.

Barlow, R. E., and Proschan, F. (1967). Exponential life test procedures when the distribution has monotone failure rate, *Journal of the American Statistical Association,* **62,** 548–560.

Barlow, R. E., and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models,* New York: Holt, Reinhardt and Winston.

Barlow, R. E., and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing: Probability Models,* 2nd ed., Silver Spring, MD: To Begin With.

Bartoszewicz, J. (1985). Dispersive ordering and monotone failure rate distributions, *Advances in Applied Probability,* **17,** 472–474.

Bartoszewicz, J. (1986). Dispersive ordering and the total time on test transform, *Statistics & Probability Letters, 4,* 285–288.

Bennett, G. (1968). A one-sided probability inequality on the sum of independent, bounded random variables, *Biometrika,* **55,** 565–569.

Bhattacharjee, M. C. (1991). Some generalized variability orderings among the life distributions with reliability applications, *Journal of Applied Probability,* **28,** 374–383.

Bhattacharjee, M. C., and Sethuraman, J. (1990). Families of life distributions characterized by two moments, *Journal of Applied Probability,* **27,** 720–725.

Bhattacharyya, G. K., and Fries, A. (1982). Fatigue failure models—Birnbaum-Saunders via inverse Gaussian, IEEE *Transactions on Reliability, 31,* 439–441.

Bickel, P. J., and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion, *Annals of Statistics, 4,* 1139–1158.

Bickel, P. J., and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models. IV. Spread, In *Contributions to Statistics*, (ed., J. Jureckova), Dordrecht: Reidel, pp. 33–40.

Birnbaum, Z. W., Esary, J. D., and Marshall, A. W. (1966). A stochastic characteristic of wear-out for components and systems, *Annals of Mathematical Statistics*, **37**, 816–825.

Birnbaum, Z. W., and Saunders, S. C. (1968). A probabilistic interpretation of Miner's rule, *SIAM Journal of Applied Mathematics*, *16*, 637–652.

Birnbaum, Z. W., and Saunders, S. C. (1969). A new family of life distributions, *Journal of Applied Probability*, *6*, 319–327.

Black, S. M., Durham, S. D., Lynch, J. D., and Padgett, W. J. (1989). A new probability distribution for the strength of brittle fibers, *Fiber-Tex* 1989—3rd Conference on Advanced Engineering Fibers and Textile Structures for Composites, *NASA Conference Publication No.* 3082, pp. 363–374.

Blackstone, E. H., Naftel, D. C., and Turner, M. E. (1980). The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information, *Journal of the American Statistical Association*, *81*, 615–624.

Block, H. W., and Savits, T. H. (1976). The IFRA closure problem, *Annals of Probability*, *4*, 1030–1042.

Bondesson, L. (1983). On preservation of classes of life distributions under reliability operations: complementary results, *Naval Research Logistics Quarterly*, **30**, 443–447.

Buckland, W. R. (1964). *Statistical Assessment of Life Characteristic*, New York: Hafner.

Calabria, R., and Pulcini, G. (1990). Asymptotic approximations to the normal hazard functions and mean residual life function, *Statistics*, **50**, 581–584.

Cao, J. H., and Wang, Y. D. (1990). New kinds of multidimensional IFR distributions, *Advances in Applied Probability*, 22, 251–253.

Cao, J. H., and Wang, Y. D. (1991). The NBUC and NWUC classes of life distributions, *Journal of Applied Probability*, *23*, 473–479.

Chang, D. S., and Tang, L. C. (1994a). Percentile bounds and tolerance limits for the Birnbaum-Saunders distribution, *Communications in Statistics—Theory and Methods*, **23**, 2853–2863.

Chang, D. S., and Tang, L. C. (1994b). Random number generator for the Birnbaum-Saunders distribution, *Proceedings of the 16th International Conference on Computers & Industrial Engineering*, pp. 727–730.

Chaudhry, M. A., and Ahmad, M. (1993). On a probability function useful in size modeling, *Canadian Journal of Forest Research*, **23**, 1679–1683.

Colvert, R. E., and Boardman, T. J. (1976). Estimation in the piecewise constant hazard rate model, *Communications in Statistics*, **A5**, 1013–1029.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

Cramér, H., and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*, New York: Wiley.

Creedy, J., and Martin, V. L. (1994). A model of the distribution of prices, ***Oxford Bulletin of Economics and Statistics*, 56, 67–76.**

Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweating, T. J. **(1991).** ***Statistical Analysis of Reliability Data*,** London: Chapman and Hall.

D'Agostino, R. B., and Stephens, M. A. (eds.). (1986). ***Goodness-of-fit Techniques*,** New York: Dekker.

Davis, H. T., and Feldstein, M. I. (1979). The generalized Pareto law as model for progressively censored survival data, ***Biometrika*, 66, 299–306.**

Deshpande, J. V., and Kochar, S. C. (1983). Dispersive ordering is the same as tail-ordering, ***Advances in Applied Probability*, 15, 686–687.**

Desmond, A. F. **(1985).** Stochastic models of failure in random environments, ***Canadian Joumal of Statistics*, 13, 171–183.**

Desmond, A. F. **(1986).** On the relationship between two fatigue-life models, ***IEEE Transactions on Reliability*, 35, 167–169.**

Devroye, L. **(1986).** The analysis of some algorithms for generating random variables with a given hazard rate, ***Naval Research Logistics Quarterly*, 33, 281–292.**

Dhillon, B. S. (1981). Life distributions, ***IEEE Transactions on Reliability*, 30,457–459.**

Dimitrov, ***B.*,** Chukova, S., and Green, D. (1993). Probability distribution in periodic random environment and its application, ***Proceedings of the Section on Quality and Productivity of ASA*,** pp. **251–256.**

Doksum, K. A. **(1969).** Starshaped transformations and the power of rank tests, ***Annals of Mathematical Statistics*, *40*, 1169–1176.**

Droste, M., and Wefelmeyer, W. (1985). A note on strong modality and dispersivity, ***Journal of Applied Probability*, 22, 235–239.**

Emad-Eldin, A. A. A., and Kochar, S. C. (1993). On hazard rate ordering of dependent variables, ***Advances in Applied Probability*, 25, 477–482.**

Engelhardt, M., Bain, L. J., and Wright, F. T. **(1981).** Inferences on parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation, ***Technometrics*, 23, 251–256.**

Evans, **M.,** and Swartz, T. **(1994).** Distribution theory and inference for polynomial-normal densities, ***Communications in Statistics—Theory and Methods*, 23, 1123–1148.**

Faguioli, E., and Pellerey, F. **(1993).** New partial orderings and applications, ***Naval Research Logistics*, 40, 829–842.**

Fisher, N. I., Lewis, T., and Embleton, B. J. **(1987).** ***Statistical Analysis of Spherical Data*,** Cambridge: Cambridge University Press.

Flehinger, B. J., and Lewis, P. A. **(1959).** Two-parameter lifetime distributions for reliability studies of renewal processes, ***IBM Journal of Research*, 3, 58–73.**

Fraser, D. A. S. **(1957).** ***Nonparametric Methods of Statistics*,** New York: Wiley.

Fraser, D. A. S. **(1968).** ***The Structure of Inference*,** New York: Wiley.

Fréchet, M. **(1950).** ***Generalités Sur les Probabilités. Elements Aléatoires***, 2nd ed., Paris: Gauthier Villars.

Freudenthal, A. M., and Shinozuka, M. **(1961).** Structural safety under conditions of ultimate load failure and fatigue, ***Technical Report WADD-TR-61-77*,** Wright Air Development Division, Wright Air Force Base, Dayton, OH.

Gaver, D. P., and Acar, M. (1979). Analytical hazard representations for use in reliability, mortality and simulation studies, *Communications in Statistics—Simulation and Computation,* **8, 91–111.**

Glaser, R. E. (1980). Bathtub and related failure rate characterizations, *Journal of the American Statistical Association, 75,* **667–672.**

Godwin, H. J. **(1964).** *Inequalities on Distribution Functions,* New York: Hafner.

Greenwich, M. **(1992).** A unimodal hazard rate function and its failure distribution, *Statistische Hefte,* **33, 187–202.**

Guess, F., Hollander, M., and Proschan, F. (1986). Testing exponentiality versus a trend change in mean residual life, *Annals of Statistics, 14,* **1388–1398.**

Gulati, S., Durham, S. D., and Padgett, W. J. (1993). On bounds for moments of the brittle fracture distribution, *Journal of Statistical Computation and Simulation,* **47, 105–108.**

Gupta, R. C., Kirmani, S. N. U. A.. and Launer, R. L. (1987). On life distributions having monotone residual variance, *Probability in the Engineering and Informational Sciences,* **1, 299–307.**

Hazelrig, J. **B.,** Turner, M. E., and Blackstone, E. H. **(1982).** Parametric survival analysis combining longitudinal and cross-sectional censored and interval-censored data with concomitant information, *Biometrics*, 38, **1–15.**

Hjorth, U. (1977). A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rate, *Report LITH-MAT-R-77-14,* Matematiska Institutionen, Tekniska Hogskolan i Linkoping, Linkoping, Finland.

Hjorth, U. **(1980).** A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates, *Technometrics,* 22, **99–107.**

Hoeffding, **W. (1963).** Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, *58,* **13–30.**

Hollander, **M.,** Park, D. H., and Proschan, F. (1986). A class of life distributions for aging, *Journal of the American Statistical* **Association, 81, 91–95.**

Hsieh, J. J. (1977). A formulation of the lifetime distribution and the existence of its moments, *Journal of Mathematical Biology, 5,* **75–86.**

Ibragimov, I. A. **(1956).** On the composition of unimodal distributions, *Theory of Probability and Its Applications, 1,* **255–260.**

Joe, H. (1985). Characterizations of life distributions for percentile residual life times, *Annals of the Institute of Statistical Mathematics, 37,* **165–172.**

Joe, H., and Proschan, F. **(1983).** Comparison of two life distributions on the basis of their percentile residual life functions, *Canadian Journal of Statistics, 12,* **91–97.**

Jørgensen, B., Seshadri, V., and Whitmore, G. A. (1991). On the mixture of the inverse Gaussian distribution with its complementary reciprocal, *Scandinavian Journal of Statistics*, **18, 77–89.**

Jorgensen, M. A. (1994). Tail functions and iterative weights in binary regression, *The American Statistician*, **48, 230–234.**

Keilson, J., and Sumita, U. **(1982).** Uniform stochastic ordering and related inequalities, *Canadian Journal of Statistics, 10,* **181–188.**

Klefsjö, B. (1982a). The HNBUE and HNWUE classes of life distributions, *Naval Research Logistics Quarterly, 29,* **331–344.**

Klefsjö, B. (1982b). NBU and NBUE survival under the Marshall-Olkin shock model, *IAPQR Transactions,* **7,** 87–96.

Kochar, S. C. (1989). On extensions of DMRL and related partial orderings of life distributions, ***Communications in Statistics—Stochastic Models, 5,*** 235–245.

Kochar, S. C. (1990a). Some partial ordering results on record values, ***Communications in Statistics—Theory and Methods,*** **19,** 299–306.

Kochar, S. C. (1990b). On preservation of some partial orderings under shock models, ***Advances in Applied Probability,*** **22,** 508–509.

Kochar, S. C., and Wiens, D. (1987). Partial ordering of life distributions with respect to their aging properties, ***Naval Research Logistics,*** **34,** 823–829.

Kunitz, H. (1989). A new class of bathtub-shaped hazard rates and its application in a comparison of two test-statistics, ***IEEE Transactions on Reliability,*** **38,** 351–355.

Lawless, J. F. (1982). ***Statistical Models & Methods for Lifetime Data,*** New York: Wiley.

Lee, M.-L. T., and Gross, A. J. (1991). Lifetime distributions under unknown environment, ***Journal of Statistical Planning and Inference,*** **29,** 137–143.

Leemis, L. M. (1986). Lifetime distribution identities, ***IEEE Transactions on Reliability,*** **35,** 170–174.

Lehmann, E. L. (1966). Some concepts of dependence, ***Annals of Mathematical Statistics,*** **37,** 1137–1153.

Lehmann, E. L., and Rojo, J. (1992). Invariant directional orderings, ***Annals of Statistics,*** **20,** 2100–2110.

Lewis, P. A. W., and Shedler, G. S. (1976). Simulation of non-homogeneous Poisson process with log-linear rate function, ***Biometrika,*** **61,** 501–505.

Lewis, P. A. W., and Shedler, G. S. (1979). Simulation of non-homogeneous Poisson process with degree-two exponential polynomial rate function, ***Operations Research,*** **27,** 1026–1041.

Lewis, T., and Thompson, J. W. (1981). Dispersive distributions, and the connection between dispersivity and strong unimodality, ***Journal of Applied Probability,*** **19,** 310–320.

Li, H., and Shaked, M. (1993). Stochastic majorization of stochastically monotone families of random variables, ***Advances in Applied Probability, 25,*** 895–913.

Lou, W. Y. (1984). Bounds on ARE's for restricted classes of distributions defined via tail orderings, ***Annals of Probability, 12,*** 685–701.

Lye, J. N., and Martin, V. L. (1993). Robust estimation, nonnormalities, and generalized exponential distributions, ***Journal of the American Statistical Association,*** **88,** 261–267.

Lynch, J. D., Mimmack, G., and Proschan, F. (1983). Dispersive ordering results, ***Adcances in Applied Probability,*** **15,** 889–891.

MacGillivray, H. L. (1981). A note on the normalized moments of distributions with non-monotonic hazard rate, ***Journal of Applied Probability, 18,*** 530–535.

Makino, T. (1984). Mean hazard rate and its applications, ***Naval Research Logistics Quarterly,*** **31,** 1–8.

Mallows, C. L. (1956). Generalizations of Tchebysheffs inequalities, ***Journal of the Royal Statistical Society, Series B,*** **18,** 139–176.

Mallows, C. L. (1963). A generalization of the Chebychev inequalities, *Proceedings of London Mathematical Society, 3rd Series,* **13,** 385–412.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data,* New York: Wiley.

Marshall, A. W., and Proschan, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate, *Annals of Mathematical Statistics, 36,* 69–77.

Marzec, L., and Marzec, P. (1991a). Testing the dispersive equivalence of two populations, *Statistics & Probability Letters,* **12,** 233–237.

Marzec, L., and Marzec, P. (1991b). Testing the equality in dispersion of two distributions, *Biometrika,* 78, 923–925.

Marzec, L., and Marzec, P. (1993). Tests for dispersive comparison of two distributions, *Communications in Statistics—Theory and Methods,* **22,** 843–851.

McDonald, J. B. (1984). Some generalized functions for the size distribution of income, *Econometrica,* **52,** 647–663.

McDonald, J. B., and Richards, D. O. (1987). Hazard rates and generalized beta distributions, *IEEE Transactions on Reliability,* **36,** 463–466.

Mehrotra, K. G. (1981). A note on the mixture of new worse than used in expectation, *Naval Research Logistics Quarterly,* **28,** 181–184.

Metzger, C., and Rüschendorf, L. (1991). Conditional ordering of distributions, In *Stochastic Models and Their Applications* (eds., *F. J.* Radermacher and U. Rieder) [vol. **32** of *Annals of Operations Research* (ed., P. *L.* Hammer)], Basel: Baltzer, pp. 127–140.

Michael, J. R., Schucany, W. R., and Haas, R. W. (1976). Generating random variates using transformations with multiple roots, *The American Statistician, 30,* 88–90.

Mitra, M., and Basu, S. K. (1994). On a nonparametric family of life distributions and its dual, *Journal of Statistical Planning and Inference, 39,* 385–397.

Mosler, K., and Scarsini, M. (1993). *Stochastic Orders and Applications. A Classified Bibliography,* Berlin: Springer-Verlag.

Nelson, W. (1982). *Applied Life Data Analysis,* New York: Wiley.

Nelson, W. (1990). *Accelerated Testing, Statistical Models, Test Plans and Data Analysis,* New York: Wiley.

Nelson, W., and Doganoksoy, N. (1994). Statistical analysis of life or strength data from specimens of various sizes, using the power-(log)normal model, In *Recent Advances in Life-testing and Reliability* (ed., N. Balakrishnan), Boca Raton, FL: CRC Press.

Oja, H. (1981). On location, scale, skewness and kurtosis of univariate distributions, *Scandinavian Journal of Statistics,* **8,** 154–168.

Padgett, W. J. (1982). On the Bayes estimation of reliability for the Birnbaum-Saunders fatigue life model, *IEEE Transactions on Reliability, 31,* 436–438.

Parzen, E. (1979). Nonparametric statistical data modeling (with discussion), *Journal of the American Statistical Association,* **74,** 105–131.

Pečarić, J. E., Proschan, F., and Tong, Y. L. (1992). *Convex Functions, Partial Orderings, and Statistical Applications,* San Dicgo, CA: Academic Press.

Pellerey, F. (1993). Partial orderings under cumulative damage shock models, *Advances in Applied Probability,* **25,** 939–946.

Prairie, R. R., and Ostle, B. (1961). An analysis of some relay failure data from a composite exponential population, ***Technometrics,*** 3, 423–428.

Rao, B. R., and Demaraju, C. V. (1992). New better than used and other concepts for a class of life distributions, ***Biometrical Journal,*** 34, 919–935.

Rao, B. R., and Talwalker, S. (1990). Setting the clock back to zero property of a family of life distributions, ***Journal of Statistical Planning and Inference, 24,*** 347–352.

Ratnasabapathi, D., Thavaneswaran, A., and Singh, J. (1992). Optimal estimation of polynomial hazard function, ***Communications in Statistics—Theory and Methods, 21,*** 2309–2316.

Rieck, J. R. (1989). Statistical theory of the Birnbaum-Saunders fatigue life distribution, ***Ph.D. dissertation,*** Department of Mathematical Sciences, Clemson University, Clemson, SC.

Rieck, J. R., and Nedelman, J. R. (1991). A log-linear model for the Birnbaum-Saunders distribution, ***Technometrics,*** 33, 51–60.

Rojo, J. (1992). A pure tail ordering based on ratio of quantile functions, ***Annals of Statistics, 20,*** 570–579.

Rojo, J. (1993). On the preservation of some pure-tail ordering by reliability operations, ***Statistics & Probability Letters, 17,*** 189–198.

Ross, S. M. (1983). ***Stochastic Processes,*** New York: Wiley.

Sathe, Y. S. (1984). Comment on Deshpande and Kochar (1983), ***Adcances in Applied Probability, 16,*** 692.

Saunders, D. J. (1984). Dispersive ordering of distributions, ***Advances in Applied Probability, 16,*** 693–694.

Saunders, I. W. (1978). Locating bright spots in a point process, ***Adcances in Applied Probability, 10,*** 557–562.

Saunders, I. W., and Moran, P. A. P. (1978). On the quantiles of the gamma and F distributions, ***Journal of Applied Probability, 15,*** 426–432.

Saunders, S. C., and Birnbaum, Z. W. (1969). Estimation for a family of life distributions with applications to fatigue, ***Journal of Applied Probability, 6,*** 328–347.

Schuster, F. F. (1984). Classification of probability laws by tail behavior, ***Journal of the American Statistical Association, 74,*** 936–939.

Setō, S., Iwase, K., and Oohara, M. (1993). Characteristics of rainfall for a single event, ***Technical Report No. 93-02,*** Department of Applied Mathematics, Hiroshima University, Hiroshima, Japan.

Shaked, M. (1977). Statistical inference for a class of life distributions, ***Communications in Statistics, A6,*** 1323–1329.

Shaked, M. (1982). Dispersive ordering of distributions, ***Journal of Applied Probability, 19,*** 310–320.

Shaked, M., and Shanthikumar, J. G. (1994). ***Stochastic Orders and Their Applications,*** San Diego, CA: Academic Press.

Shanthikumar, J. G., Yamazaki, G., and Sakesagawa, H. (1991). Characterization of optimal order of servers in a tandem queue with blocking, ***Operations Research Letters. 10.*** 17–22.

Sibuya, M. **(1984).** Unimodal hazard functions, *Japanese Journal of Applied Statistics*, **13, 115–124.**

Sibuya, M. (1994). The shapes of a probability density function and its hazard function, *Proceedings of the International Conference on Lifetime Data Models in Reliability and Survival Analysis.*

Sibuya, *M.*, and Hirotsu. C. **(1984).** Parametric and nonparametric models in hazard analysis, *Proceedings of the 12th International Biometrics Conference, Tokyo,* pp. **42–48.**

Smith, R. M., and Bain, L. J. (1975). An exponential-power life-testing distribution, *Communications in Statistics,* **4, 469–481.**

Sobolev, I. G. **(1994).** Interrelations between probability distributions belonging to the $q - r$ family, ***Izmeritel'naya*** *Tekhnika,* **10, 11–14.**

Sorensen, A. **(1971).** Product reliability and random fatigue, *IEEE Transactions on Reliability,* **20, 284–288.**

Tierney, L., Kass, R. E., and Kadane, J. B. **(1989).** Fully exponential Laplace approximations to expectations and variances of non-positive functions, *Journal of the American Statistical Association,* **84, 710–716.**

Tsukatani, **T.,** and Shigemitsu, K. **(1980).** Simplified Pearson distributions applied to air pollutant concentration, *Atmospheric* ***Environment***, **14,** 245–248.

van Zwet, W. R. **(1964).** Convex transformations: A new approach to skewness and kurtosis, ***Statistica*** *Neerlandica,* **18, 433–441.**

Wald, A. **(1938).** Generalization of the inequality of Markoff, *Annals of Mathematical* ***Statistics***, *9,* 244–255.

Whitt, W. **(1980).** Uniform conditional stochastic order, *Journal of Applied Probability,* **17, 112–123.**

Whitt, W. **(1985).** Uniform conditional variability ordering of probability distributions, *Journal of Applied Probability,* **22, 619–633.**

Yanagimoto, T., and Sibuya, M. **(1976).** Isotonic tests for spread and tail, *Annals of the Institute of Statistical Mathematics,* **28, 329–342.**

Zelterman, D. **(1992).** A statistical distribution with an unbounded hazard function and its application to a theory from demography, ***Biometrics***, **48, 807–818.**

Abbreviations

ABLUE	asymptotically best linear unbiased estimator
ARE	asymptotic relative efficiency
ARMA	autoregressive moving average
BFR	bathtub failure rate
BHR	bathtub hazard rate
BLIE	best linear invariant estimator
BLUE	best linear unbiased estimator
cdf	cumulative distribution function
cgf	cumulant generating function
CPU	central processing unit
DFR	decreasing failure rate
DFRA	decreasing failure rate on average
DHR	decreasing hazard rate
DHRA	decreasing hazard rate on average
DIMRL	decreasing, then increasing mean residual life
DMRL	decreasing mean residual life
G3B	generalized three-parameter beta (distribution)
G3F	"corrected" F-distributions
GB2	generalized beta of the second kind
GNB	generalized negative binomial distribution
GT	generalized t (distribution)
HNBUE	harmonic new better than used in expectation
HNWUE	harmonic new worse than used in expectation
IDMRL	increasing, then decreasing mean residual life
IFR	increasing failure rate
IFRA	increasing failure rate on average
IHR	increasing hazard rate
IHRA	increasing hazard rate average
LSE	least-squares estimator
MHR	mean hazard rate
MLE	maximum likelihood estimator

MME	method of moments estimator
MMME	modified method of moments estimator
MRL	mean residual life
MSE	mean square error
MVUE	minimum variance unbiased estimator
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
NBU	new better than used
NBUE	new better than used in expectation
NBUFR	new better than used in failure rate
NBUFRA	new better than used in failure rate on average
NWUE	new worse than used in expectation
pdf	probability density function
PERT	program evaluation and review technique
Pr[E]	probability of event E
PWM	probability-weighted moment
UMVUE	uniform(ly) minimum variance unbiased estimator

Author Index

The purpose of this Index is to provide readers with quick and easy access to the contributions (pertinent to this volume) of any individual author, and not to highlight any particular author's contribution.

Subject Index

For readers' convenience, lists of all continuous and discrete distributions that are referred to in this volume have also been included in this Index.