MARGINAL TAIL PROBABILITIES AND INFERENCE FOR REAL PARAMETERS

by

Thomas J. DiCiccio, Christopher A. Field, and Donald A.S. Fraser

Technical Report No. 256
March 1989

Prepared under the Auspices of
National Science Foundation
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
MARGINAL TAIL PROBABILITIES AND INERENCE FOR REAL PARAMETERS

by

Thomas J. DiCiccio and Christopher A. Field
Stanford University

and

Donald A. S. Fraser
York University

Technical Report No. 256
March 1989

Prepared Under the Auspices
of
National Science Foundation Grant DMS 87-08083

(Also prepared under National Science Foundation Grant DMS 86-00235 and issued as Technical Report No. 317, Department of Statistics, Stanford University.)

Department of Statistics
Stanford University
Stanford, California
MARGINAL TAIL PROBABILITIES AND INFERENCE FOR REAL PARAMETERS

by

Thomas J. DiCiccio and Christopher A. Field
Department of Statistics, Stanford University
Stanford, California 94305, U.S.A.

and

Donald A. S. Fraser
Department of Mathematics, York University
Toronto, M3T 1P3, Canada

Summary

In many situations, inference for a scalar parameter in the presence of nuisance parameters requires integration of a joint density of pivotal quantities or a joint posterior density. For such inferences, accurate marginal tail probability approximations are useful to avoid high-dimensional integrals. Two approximations are developed in this article. Numerical results are given for conditional inference in regression settings, which show the approximations to be generally accurate even for small sample sizes.

Some key words: Bayesian inference; Conditional inference; Lugannani–Rice formula; Multiple regression; Saddlepoint approximation; Signed root likelihood ratio statistic; Type II censoring.
1. Introduction

Calculation of marginal tail probabilities is central to constructing confidence intervals and testing hypotheses for a scalar parameter. For parametric situations, by conditioning or sufficiency arguments, inference about unknown parameters is often achieved through exact or approximate pivots whose joint density is known to a high order of accuracy except, perhaps, for a normalizing constant. Inference about a scalar parameter is then based on the marginal distribution function of the appropriate pivot, and exact calculation of this distribution function usually involves a multidimensional integral which can be difficult to compute in practice. In the Bayesian context, determination of marginal posterior tail probabilities can present similar difficulties.

This article derives two marginal tail probability approximations suitable for such situations. For the univariate case, the approximations coincide and they are similar to approximations given by Fraser (1988) and Barndorff-Nielsen (1988). The errors in both approximations are of order $n^{-3/2}$, where $n$ is the sample size. The main approximation is simple to compute, requiring only first- and second-order derivatives of the log density. The second approximation is more difficult to implement, but it is found to give more accurate answers in extreme cases where the sample size is small and the dimension of the joint distribution is large. The approximations are illustrated in location-scale and linear regression models for which conditional inference is appropriate.

Let $X = (X^1, \ldots, X^p)$ be a vector variable that is $O_p(n^{-1/2})$ as some parameter $n$ increases indefinitely. In applications, $n$ usually represents a number of observations. Now suppose that the density of $X$ is given by

$$f(x) \propto \exp\{\ell(x)\}, \quad x = (x^1, \ldots, x^p),$$

where the function $\ell$ is known. It is assumed that $\ell(x)$ is $O_p(n)$ for each fixed $x$ and that $\ell(x)$ attains its maximum value at $x = 0$. This article concerns approximations to the marginal distribution function of a single component of $X$, say $X^1$.

A simple example of the situation described in the preceding paragraph is obtained by taking $X = X^1 \sim n^{-1/2}t(n)$, where $t(n)$ has Student’s $t$-distribution with $n$ degrees of freedom. In this case,

$$\ell(x) = -\frac{1}{2}(n + 1)\log(1 + x^2), \quad -\infty < x < \infty.$$ 

Another instance for which $p = 1$ is provided by Fisher’s (1973, §6.9) example of the Nile problem. In this example, $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ are independent samples from
exponential distributions having means $\tau$ and $\tau^{-1}$, respectively. The maximum likelihood estimator of $\tau$ is $\hat{\tau} = (\Sigma Y_j / \Sigma Z_j)^{1/2}$, and the distribution of $A = (\Sigma Y_j \Sigma Z_j)^{1/2}$ does not depend on $\tau$. Taking $\theta = \log \tau$ and $\hat{\theta} = \log \hat{\tau}$, the conditional density of $X = X^1 = \theta - \hat{\theta}$ given $A$ is

$$f(x) \propto \exp\{-2A \cosh x\}, \quad -\infty < x < \infty.$$ 

More complicated examples in this framework arise from conditional inference for the parameters of location–scale or linear regression models. A detailed discussion of such examples is given in Sections 3 and 4.

The approximations to the distribution function of $X^1$ discussed here involve the function

$$w^1(x^1) = 2\{\ell(0) - \ell(\hat{x}(x^1))\},$$

and its signed square root

$$r^1(x^1) = \text{sgn} (x^1) \{w(x^1)\}^{1/2},$$

where $\hat{x}(x^1)$ is the point at which $\ell(x)$ is maximized subject to the constraint that the first component of $x$ equals the specified value $x^1$. In particular, normal approximations to the distribution of $R^1 = r^1(X^1)$ play a central role. In the examples considered in Sections 3 and 4 concerning maximum likelihood estimation for location–scale and linear regression models, $W^1 = w^1(X^1)$ corresponds to a likelihood ratio statistic, and $R^1$ is its signed root.

It is argued in Section 2 that the standard normal approximation to the distribution of $R^1$ has error of order $O_p(n^{-1/2})$. Thus, provided $x^1$ is $O_p(n^{-1/2})$,

$$\text{pr}(X^1 \leq x^1) = \text{pr}(R^1 \leq r^1) = \Phi(r^1) + O_p(n^{-1/2}), \quad (1)$$

where $r^1 = r^1(x^1)$. It is also argued that by accounting for the mean of $R^1$ the error in the normal approximation can be reduced to order $O_p(n^{-1})$, and that by accounting for both the mean and variance of $R^1$ the error can be reduced further to order $O_p(n^{-3/2})$. Thus,

$$\text{pr}(X^1 \leq x^1) = \Phi\{(r^1 - \mu)/\sigma\} + O_p(n^{-3/2}), \quad (2)$$

where $\mu$ and $\sigma^2$ are the mean and variance of $R^1$, respectively. Mean and variance adjustments that improve the standard normal approximation to the distribution of the signed root of the likelihood ratio statistic have been discussed by Barndorff–Nielsen (1986), DiCiccio (1984, 1988) and McCullagh (1984; 1987, §6.2.6, 7.4.5), among others.
The exact values of $\mu$ and $\sigma^2$ are typically not available for use in (2). However, in principle, they can be sufficiently well approximated so that (2) remains valid. Expansions for $\mu$ and $\sigma^2$ are given in expression (9). For the univariate case $p = 1$, $x = x^1$ and it suffices to take

$$
\mu = \frac{1}{3} \frac{\ell^{(3)}(0)}{-\ell^{(2)}(0)}^{3/2}, \quad \sigma^2 = 1 + \frac{1}{4} \frac{\ell^{(4)}(0)}{-\ell^{(2)}(0)}^2 + \frac{11}{36} \left(\frac{\ell^{(3)}(0)}{-\ell^{(2)}(0)}\right)^2, \quad (3)
$$

where $\ell^{(j)}(x) = d^j \ell(x) / (dx)^j$, $(j = 1, \ldots, 4)$. Unfortunately, for the more interesting cases that have $p > 1$, the expansions for $\mu$ and $\sigma^2$ are quite complicated, involving third-, fourth-, and sixth-order sums of second-, third-, and fourth-order partial derivatives of $\ell$. The use of (2) in such cases can be intractable.

A procedure is presented in Section 2 whereby the mean and variance adjustments in (2) can be achieved through a simple formula that involves only first- and second-order partial derivatives. The general form of the approximation is given in expression (12); for the case $p = 1$, it reduces to

$$
\Pr(X \leq x) = \Phi(r) + \phi(r) \left\{ \frac{1}{r} + \frac{-\ell^{(2)}(0)1^{1/2}}{\ell^{(1)}(x)} \right\} + O_p(n^{-3/2}), \quad (4)
$$

where $r = r(x) = \text{sgn}(x) [2\{\ell(0) - \ell(x)\}]^{1/2}$ and $\Phi$ is the standard normal density. Approximations closely related to (4) have recently been given by Barndorff-Nielsen (1988) and Fraser (1988).

Table 1 shows the results obtained from using approximations (1), (2), and (4) in Fisher's example of the Nile problem when $A = \frac{1}{2}$. Here, (3) yields $\mu = 0$ and $\sigma^2 = 1 - (8A)^{-1}$. In this problem, although (1) is already fairly accurate, the use of (4) offers some improvement, particularly in the extreme tails. In the examples considered in Sections 3 and 4, the improvement afforded by (12) over (1) is more remarkable.

It should be noted that by definition $r^1(0) = 0$, and thus approximation (12) breaks down when $x^1 = 0$. In the examples, this property is shown not to be a problem, however, since the approximation yields quite accurate results for values of $x^1$ very near 0. For instance, from Table 1, (4) works well for $x = -0.01$.

In some situations, where $p$ is large and $n$ is small, approximation (12) can give poor results; indeed, it is possible for the approximation to be negative. Such cases are illustrated in Examples 3 and 4 of Section 4. The reason for this failure is that the leading component of (12), $\Phi(r^1)$, takes $x^1 = 0$ as the median of $X^1$. If the marginal density of
$X^1$ is actually centered at a point substantially away from 0, then $\Phi(r^1)$ can be a very poor approximation and the correction term in (12) is required to make a large adjustment which it cannot accurately achieve. An alternative tail probability approximation that avoids this difficulty is given in (19). The leading component of (19) is $\Phi(r^*)$, where $r^*$ is an adjusted version of $r^1$ that takes the value 0 when $x^1$ is close to the mode of $X^1$. In practice, however, approximation (12) is much more simple to use than (19). A case where (19) provides substantial improvement over (12) is discussed in Example 4 of Section 4.

Section 5 provides derivations of approximations (12) and (19) that are different than those given in Section 2. Comparisons between (4) and the Lugannani–Rice (1980) approximation for tail probabilities are also discussed in Section 5.

2. Approximations of marginal tail probabilities

To present a fuller account of normal approximations to the distribution of $R^1$, it is necessary to introduce additional notation. Let $\ell_i(x) = \partial \ell(x)/\partial x^i$, $\ell_{ij}(x) = \partial^2 \ell(x)/\partial x^i \partial x^j$, etc., and set $I_{ij} = -\ell_{ij}(0)$, $I_{ijk} = \ell_{ijk}(0)$, and $I_{ijk\ell} = \ell_{ijk\ell}(0)$, $(i, j, k, \ell = 1, \ldots, p)$. Furthermore, take $I = (I_{ij})$, $I^{-1} = (I^{ij})$, and

$$F_{3,1} = I_{ijk} I^{ij} I^{kl} (I^{11})^{-3/2}, \quad F_{3,2} = I_{ijkl} I^{ij} I^{kl} (I^{11})^{-1/2},$$
$$F_{4,1} = I_{ijkl} I^{ij} I^{kl} (I^{11})^{-2}, \quad F_{4,2} = I_{ijkl} I^{ij} I^{kl} (I^{11})^{-1}, \quad F_{4,3} = I_{ijkl} I^{ij} I^{kl},$$
$$F_{6,1} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-3}, \quad F_{6,2} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-2}, \quad F_{6,3} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-1},$$
$$F_{6,4} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-2}, \quad F_{6,5} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-1}, \quad F_{6,6} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-1},$$
$$F_{6,7} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-1}, \quad F_{6,8} = I_{ijklm} I^{ij} I^{kl} I^{lm} (I^{11})^{-1}.$$

The usual notational convention is followed here in which summation over the range 1, \ldots, $p$ is understood for each index that appears in an expression as both a subscript and a superscript.

The expansion of $r^1(x^1)$ required to derive asymptotic formulae for the moments of $R^1$ is readily obtainable through straightforward calculations. From the definition of $\bar{x}(x^1) = (\bar{x}^1(x^1), \ldots, \bar{x}^p(x^1))$, it follows that $\bar{x}(0) = 0$, $\bar{x}^1(x^1) = x^1$, and $\ell_\alpha(\bar{x}(x^1)) = 0$ ($\alpha = 2, \ldots, p$). Differentiation of $\ell_\alpha(\bar{x}(x^1)) = 0$ with respect to $x^1$ shows

$$\frac{d\bar{x}^i(x^1)}{dx^1} \bigg|_{x^1=0} = I^{11} (I^{11})^{-1}, \quad \frac{d^2\bar{x}^i(x^1)}{(dx^1)^2} \bigg|_{x^1=0} = (I^{ij} - I^{11} I^{ij} (I^{11})^{-1}) (I_{ij \ell} I^{kl} (I^{11})^{-2}),$$

(5)
and moreover, provided \( x^1 \) is of order \( O_p(n^{-1/2}) \),

\[
\ell(\tilde{x}(x^1)) = \ell(0) - \frac{1}{2} \left\{ x^1(I^{11})^{-1/2} \right\}^2 + \frac{1}{6} \left\{ x^1(I^{11})^{-1/2} \right\}^3 F_3 \\
+ \frac{1}{24} \left\{ x^1(I^{11})^{-1/2} \right\}^4 F_4 + O_p(n^{-3/2}),
\]

(6)

where \( F_3 = F_{3,1} \) and \( F_4 = F_{4,1} + 3F_{6,3} - 3F_{6,1} \). Therefore,

\[
w^1(x^1) = \left\{ x^1(I^{11})^{-1/2} \right\}^2 - \frac{1}{3} \left\{ x^1(I^{11})^{-1/2} \right\}^3 F_3 - \frac{1}{12} \left\{ x^1(I^{11})^{-1/2} \right\}^4 F_4 + O_p(n^{-3/2}),
\]

and

\[
r^1(x^1) = x^1(I^{11})^{-1/2} \left[ 1 - \frac{1}{6} \left\{ x^1(I^{11})^{-1/2} \right\} F_3 - \frac{1}{24} \left\{ x^1(I^{11})^{-1/2} \right\}^2 (F_4 + \frac{1}{3} F_3^2) \right] + O_p(n^{-3/2}).
\]

(7)

In these expressions, \( F_3 \) is \( O_p(n^{-1/2}) \), while \( I^{11}, \ F_4, \), and \( F_3^2 = F_{3,1}^2 = F_{6,1} \) are all of order \( O_p(n^{-1}) \). Related expansions have been given by Sprott (1980).

Calculations similar to ones described by Hinkley (1978) show that the normalizing constant \( c \) defined by \( f(x) = c \exp \{ \ell(x) \} \) satisfies

\[
c = \exp \{ \ell(0) \} (2\pi)^{-p/2} |I|^{1/2} \left\{ 1 - \frac{1}{8} F_{4,3} - \frac{1}{8} F_{6,7} - \frac{1}{12} F_{6,8} + O_p(n^{-3/2}) \right\}.
\]

(8)

It follows from expansions (7) and (8) that

\[
E(R^1) = a + O_p(n^{-3/2}), \quad E\{(R^1)^2\} = 1 + b + O_p(n^{-3/2}),
\]

(9)

where

\[
a = -\frac{1}{6} F_{3,1} + \frac{1}{2} F_{3,2},
\]

\[
b = -\frac{1}{4} F_{4,1} + \frac{1}{2} F_{4,2} + \frac{5}{12} F_{6,1} - \frac{1}{2} F_{6,2} - \frac{3}{4} F_{6,3} + \frac{1}{4} F_{6,4} + \frac{1}{2} F_{6,5} + \frac{1}{2} F_{6,6};
\]

third- and higher-order cumulants of \( R^1 \) are \( O_p(n^{-3/2}) \). Note that \( a \) is of order \( O_p(n^{-1/2}) \), \( b \) is of order \( O_p(n^{-1}) \), and \( \text{var}(R^1) = 1 + (b - a^2) + O_p(n^{-3/2}) \).

Thus, the standard normal approximations to the distributions of \( R^1 \), \( R^1 - a \), and \( (R^1 - a)\{1 + (b - a^2)\}^{-1/2} \) have errors of order \( O_p(n^{-1/2}), \ O_p(n^{-1}), \) and \( O_p(n^{-3/2}) \), respectively. In addition, the approximations to the distributions of \( W^1 \) and \( W^1(1 + b)^{-1} \) by the chi-squared distribution with a single degree of freedom have errors of orders \( O_p(n^{-1}) \) and \( O_p(n^{-2}) \), respectively. Multiplicative adjustments that improve chi-squared
approximations to the distributions of likelihood ratio statistics have been discussed by various authors, including Barndorff–Nielsen and Cox (1984), Barndorff–Nielsen and Hall (1988), Lawley (1956), and McCullagh (1987, §7.4.4).

The approximation to the distribution function of $X^1$ arising from the adjusted normal approximation for $R^1$ is

$$
\text{pr}(X^1 \leq x^1) = \Phi(r^1) - \Phi(r^1) \left[ a + \frac{1}{2}r^1 b \right] + O_p(n^{-3/2})
$$

where $x^1$ is $O_p(n^{-1/2})$, $r^1 = r^1(x^1)$, and $\Phi$ and $\phi$ are the standard normal density and distribution functions. Use of approximation (10) suffers from difficulties encountered in the calculation of $a$ and $b$, particularly in problems where $p$ is large. A straightforward approximation to the term $[-a + \frac{1}{2}(x^1(I^{11})^{-1/2})b] + O_p(n^{-3/2})$ is helpful for applications.

A simple approximation to that term is provided by the formula

$$
-\left[ a + \frac{1}{2}(x^1(I^{11})^{-1/2})b \right] = \frac{1}{r^1} + \frac{1}{(I^{11})^{1/2} \ell_1(\tilde{x}(x^1)) \left| \frac{\partial I_{\alpha\beta}}{\partial \ell_{\alpha\beta}(\tilde{x}(x^1))} \right|^{1/2}} + O_p(n^{-3/2}),
$$

where $\alpha$ and $\beta$ vary over the range $2, \ldots, p$; that is, $(I_{\alpha\beta})$ is the $(p-1) \times (p-1)$ submatrix of $I$ corresponding to $x^2, \ldots, x^p$, and similarly for $(-\ell_{\alpha\beta}(\tilde{x}(x^1)))$. By means of (11), the mean and variance of $R^1$ can be taken into account by using only first– and second–order partial derivatives of $\ell$. Combining (10) and (11), the marginal distribution function of $X^1$ can be approximated by

$$
\text{pr}(X^1 \leq x^1) = \Phi(r^1) + \phi(r^1) \left[ \frac{1}{r^1} + \frac{|I|^{1/2}}{\ell_1(\tilde{x}(x^1))((-\ell_{\alpha\beta}(\tilde{x}(x^1))))^{1/2}} \right] + O_p(n^{-3/2}),
$$

provided $x^1$ is $O_p(n^{-1/2})$, since $|I| = (I^{11})^{-1} |(I_{\alpha\beta})|$. For the univariate case $p = 1$, expression (12) reduces to formula (4) given in Section 1.

To verify (11), first note that

$$
\frac{1}{r^1} = \left[ x^1(I^{11})^{-1/2} \right]^{-1} \left[ 1 + \frac{1}{6} \left( x^1(I^{11})^{-1/2} \right) F_3 + \frac{1}{24} \left( x^1(I^{11})^{-1/2} \right)^2 (F_4 + F_3^2) + O_p(n^{-3/2}) \right].
$$

Differentiation of $\ell(\tilde{x}(x^1))$ at (6) with respect to $x^1$ shows

$$
\ell_1(\tilde{x}(x^1)) = -\left[ x^1(I^{11})^{-1} \right] \left[ 1 - \frac{1}{2} \left( x^1(I^{11})^{-1/2} \right) F_3 - \frac{1}{6} \left( x^1(I^{11})^{-1/2} \right)^2 F_4 + O_p(n^{-3/2}) \right],
$$
and hence
\[
\frac{1}{(I^{11})^{1/2} \ell_1(\tilde{x}(x^1))} = -\left\{x^1(I^{11})^{-1/2}\right\}^{-1} \left[1 + \frac{1}{2} \{x^1(I^{11})^{-1/2}\}^2 F_3 + \frac{1}{6} \{x^1(I^{11})^{-1/2}\}^2 F_4 + \frac{3}{2} F_3 + O_p(n^{-3/2})\right].
\]
(14)

Using expression (5), it can be shown that
\[
-\ell_{\alpha \beta}(\tilde{x}(x^1)) = I_{\alpha \beta} - x^1 I_{\alpha \beta i} I^{i1} (I^{11})^{-1} - \frac{1}{2} (x^1)^2 \{I_{\alpha \beta i j} I^{i1} I^{j1} (I^{11})^{-2} + I_{\alpha \beta i j k l} I^{i1} I^{j1} I^{k1} I^{l1} (I^{11})^{-3}\} + O_p(n^{-1/2}).
\]

It then follows from a lemma of Barndorff–Nielsen (1986) concerning determinants that
\[
\left\{\frac{|(I_{\alpha \beta})|}{|(-\ell_{\alpha \beta}(\tilde{x}(x^1)))|}\right\}^{1/2} = 1 + \left\{x^1(I^{11})^{-1/2}\right\} \left(\frac{1}{2} F_{3,1} + \frac{1}{2} F_{3,2}\right) + \left\{x^1(I^{11})^{-1/2}\right\}^2 \left(-\frac{1}{4} F_{4,1} + \frac{1}{4} F_{4,2} + \frac{5}{8} F_{6,1}\right)
\[ + \frac{1}{2} F_{6,2} - \frac{3}{4} F_{6,3} + \frac{1}{8} F_{6,4} + \frac{1}{4} F_{6,5} + \frac{1}{4} F_{6,6}\right) + O_p(n^{-3/2}).
\]
(15)

The validity of (11) is established by combining (13), (14), and (15).

From the preceding derivation, it is apparent that expression (12) remains valid even if the density \(c \exp\{\ell(x)\}\) only approximates \(f(x)\) with relative error of order \(O_p(n^{-3/2})\), i.e., if \(f(x) = c \exp\{\ell(x)\}\{1 + O_p(n^{-3/2})\}\), where \(c \exp\{\ell(x)\}\) integrates to 1.

Formula (12) can easily be generalized to provide marginal tail probability approximations in cases where the joint density of the variables is maximized at a point other than 0. Consider now a variable \(Y = (Y_1, \ldots, Y_n)\) such that \(Y - \delta\) is \(O_p(n^{-1/2})\) for a vector \(\delta = (\delta^1, \ldots, \delta^p)\). Suppose that the probability density function \(Y\) is
\[
g(y) = d \exp\{m(y)\}\{1 + O_p(n^{-3/2})\}, \ y = (y^1, \ldots, y^p),
\]
where \(m(y)\) is \(O_p(n)\) for each fixed \(y\), \(m(y)\) attains its maximum value at \(y = \delta\), and \(d\) is a normalizing constant so that \(d \exp\{m(y)\}\) integrates to 1. Following the previous notation, let \(m_i(y) = \partial m(y)/\partial y^i\), \(m_{ij}(y) = \partial^2 m(y)/\partial y^i \partial y^j\), \((i, j = 1, \ldots, p)\), and let \(\tilde{y}(y^1)\) be the point at which \(m(y)\) attains its maximum value subject to the constraint that the first component of \(y\) equals the specified value \(y^1\). Then, provided \(y^1 - \delta^1 = O_p(n^{-1/2})\),
\[
\Pr(Y^1 \leq y^1) = \Phi(r^1 + \phi(r^1) \left\{1 + \frac{|(-m_{ij}(\delta))|^{1/2}}{m_1(\tilde{y}(y^1))|(-m_{\alpha \beta}(\tilde{y}(y^1)))|^{1/2}}\right\} + O_p(n^{-3/2}),
\]
(16)
where \( r^1 \) is now defined as \( r^1 = r^1(y^1) = \text{sgn}(y^1 - \delta^1)[2\{m(\delta) - m(\tilde{y}(y^1))\}]^{1/2}, \) \( i \) and \( j \) vary over \( 1, \ldots, p, \) and \( \alpha \) and \( \beta \) vary over \( 2, \ldots, p. \) For the univariate case \( p = 1, \) using obvious notation, (16) reduces to
\[
\text{pr}(Y \leq y) = \Phi(r) + \phi(r) \left[ \frac{1}{r} + \frac{\{-m^{(2)}(\delta)\}^{1/2}}{m^{(1)}(y)} \right] + O_p(n^{-3/2}),
\]
where \( r \) is now defined as \( r = r(y) = \text{sgn}(y - \delta)[2\{m(\delta) - m(y)\}]^{1/2}. \)

An alternative approximation to (12) can be developed by applying (17) to the marginal probability density function of \( X^1. \) The first four cumulants of \( X^1(I_{11})^{-1/2} \)
are
\[
\kappa_1\{X^1(I_{11})^{-1/2}\} = \frac{1}{2} F_{3,2} + O_p(n^{-3/2}),
\kappa_2\{X^1(I_{11})^{-1/2}\} = 1 + \frac{1}{2} F_{4,2} + \frac{1}{2} F_{6,5} + \frac{1}{2} F_{6,6} + O_p(n^{-3/2}),
\kappa_3\{X^1(I_{11})^{-1/2}\} = F_{3,1} + O_p(n^{-3/2}),
\kappa_4\{X^1(I_{11})^{-1/2}\} = F_{4,1} + 3F_{6,3} + O_p(n^{-3/2}),
\]
and the higher-order cumulants are \( O_p(n^{-3/2}) \) or smaller. These cumulants can be used to obtain an Edgeworth expansion for the marginal probability density function \( f_1 \) of \( X^1. \) Using this expansion in conjunction with (6) and (15), it can be shown that
\[
f_1(x^1) = c_1 \left\{ \frac{\| (I_{\alpha\beta} \|}{\|(-\ell_{\alpha\beta}(\tilde{x}(x^1)))\|} \right\}^{1/2} \exp \{ \ell(\tilde{x}(x^1)) - \ell(0) \} \{1 + O_p(n^{-3/2})\},
\]
where \( c_1 \) is a normalizing constant. Formula (18) is also given by Tierney and Kadane (1986), who derive it by application of Laplace's approximation for integrals.

A second approximation to the distribution function of \( X^1 \) is obtained then by applying (17) with
\[
m(x^1) = -\frac{1}{2} \log \left| (-\ell_{\alpha\beta}(\tilde{x}(x^1))) \right| + \ell(\tilde{x}(x^1)).
\]
Thus,
\[
\text{pr}(X^1 \leq x^1) = \Phi(r^*) + \phi(r^*) \left[ \frac{1}{r^*} + \frac{\{-m^{(2)}(\delta)\}^{1/2}}{m^{(1)}(x^1)} \right] + O_p(n^{-3/2}),
\]
where \( \delta \) is the point at which \( m(x^1) \) attains its maximum value, \( r^* = \text{sgn}(x^1 - \delta)[2\{m(\delta) - m(x^1)\}]^{1/2}, \) and \( m^{(1)} \) and \( m^{(2)} \) are the first and second derivatives of \( m. \) Although in practice approximation (19) can be expected to give more accurate results than (12), (19) is more difficult to implement since it requires calculation of the maximizing point \( \delta \) and determination of the first two derivatives of \( m. \) Note that
\[
\text{pr}(X^1 \leq x^1) = \Phi(r^*) + O_p(n^{-1/2}),
\]
and approximation (20) can be expected to give more accurate results than (1). Of course, in the univariate case \( p = 1 \), approximations (12) and (19) coincide, as do (1) and (20).

Tierney and Kadane (1986) use (18) to approximate marginal posterior densities. It is clear that (12) and (19) can be applied in Bayesian contexts to obtain approximations of marginal posterior distribution functions.

### 3. Location–scale models

Consider variables \( Y_1, \ldots, Y_n \) such that \( Y_i \sim \mu + \sigma Z_i \) (\( i = 1, \ldots, n \)), where \( Z_1, \ldots, Z_n \) are independent and identically distributed with known probability density function \( f \) and distribution function \( F \). Suppose that Type II censoring occurs and that only the \( m \) smallest \( Y_{(1)} \leq \cdots \leq Y_{(m)} \) of these variables are observed. The likelihood function for \( \mu \) and \( \theta = \log \sigma \) based on \( Y = (Y_{(1)}, \ldots, Y_{(m)}) \) is

\[
L(\mu, \theta; Y) = \frac{1}{e^{m\theta}} \prod_{i=1}^{m} f \left( \frac{Y_{(i)} - \mu}{e^{\theta}} \right) \left[ 1 - F \left( \frac{Y_{(m)} - \mu}{e^{\theta}} \right) \right]^{n-m}.
\]

Various authors, including Fisher (1934), Fraser (1976), Hinkley (1978), and Lawless (1978), have argued that inference about \( \mu \) and \( \theta \) should be based on the joint conditional distribution of the maximum likelihood estimators \( \hat{\mu} \) and \( \hat{\theta} \) given the sample configuration \( A = (A_1, \ldots, A_m) \), where

\[
A_i = (Y_{(i)} - \hat{\mu})e^{-\hat{\theta}} \quad (i = 1, \ldots, m).
\]  

(21)

Moreover, if \( \ell \) is defined by

\[
\ell(x^1, x^2) = \log L(x^1, x^2; A) = \log L(\hat{\mu} + x^1 e^{\hat{\theta}}, \hat{\theta} + x^2; Y) + m\hat{\theta},
\]

(22)

then the joint conditional density of the pivots \( X^1 = (\mu - \hat{\mu})e^{-\hat{\theta}} \) and \( X^2 = \theta - \hat{\theta} \) is given by

\[
f_{X^1, X^2|A}(x^1, x^2) \propto \exp \{ \ell(x^1, x^2) \}.
\]

(23)

Approximations to the marginal conditional distribution functions of \( X^1 \) and \( X^2 \) can thus be obtained from expression (12).

Formula (12) can also be applied for approximate conditional inference about the \( p \)th quantile of \( Y_i \). If the \( p \)th quantile of \( Y_i \) is denoted by \( y_p \) and \( \epsilon_p = F^{-1}(p) \) is the known \( p \)th quantile of \( Z_i \), then \( y_p = \mu + \sigma \epsilon_p \), and the model can be rewritten in a location–scale
form with $y_p$ as the location parameter. Specifically, $Y_i \sim y_p + \sigma Z_i$, where $Z_i = Z_i - \varepsilon_p$ has density $f(z_i + \varepsilon_p)$. By applying (22) in this context, it can be seen that the joint conditional density of the pivots

$$X^1 = (y_p - \hat{y}_p)e^{-\theta} = \{(\mu - \hat{\mu}) + (e^{\theta} - e^{\hat{\theta}})\varepsilon_p\}e^{-\theta}, \ X^2 = \theta - \hat{\theta}$$

is given by (23), where

$$\ell(x^1, x^2) = -\left\{ m x^2 + \sum_{i=1}^{m} g(P_i + \varepsilon_p) + G(P_m + \varepsilon_p) \right\},$$

$$g(z) = \log f(z), \ G(z) = -(n - m) \log(1 - F(z)), \ P_i = \{A_i - (x^1 + \varepsilon_p)\}e^{-z^2} (i = 1, \ldots, m),$$

and $A_1, \ldots, A_m$ are as defined in (21).

For a specified value of $x^1$, the corresponding value $\bar{x}^2 = \bar{x}^2(x^1)$ of $x^2$ that maximizes (24) satisfies

$$\sum_{i=1}^{m} g^{(1)}(P_i + \varepsilon_p)\bar{P}_i + G^{(1)}(\bar{P}_m + \varepsilon_p)\bar{P}_m = m,$$

where $g^{(1)}$ and $G^{(1)}$ are the first derivatives of $g$ and $G$, and $\bar{P}_i = \{A_i - (x^1 + \varepsilon_p)\}e^{-\bar{x}^2}$. The approximation to the conditional distribution function of $X^1 = (y_p - \hat{y}_p)e^{-\theta}$ provided by (12) is

$$\text{pr}(X^1 \leq x^1|A) \approx \Phi(r^1) + \phi(r^1) \left[ \frac{1}{r^1} + \frac{(I_{11}I_{22} - I_{12}^2)^{1/2}}{\ell_1(x^1, \bar{x}^2)\{ - \ell_22(x^1, \bar{x}^2)\}^{1/2}} \right],$$

where

$$r^1 = r^1(x^1) = \text{sgn}(x^1) \sqrt{2\left\{ m\bar{x}^2 + \sum_{i=1}^{m} \{g(\bar{P}_i + \varepsilon_p) - g(A_i)\} + \{G(\bar{P}_m + \varepsilon_p) - G(A_m)\}\right\})^{1/2},$$

$$\ell_1(x^1, \bar{x}^2) = e^{-\bar{x}^2} \left\{ \sum_{i=1}^{m} g^{(1)}(\bar{P}_i + \varepsilon_p) + G^{(1)}(\bar{P}_m + \varepsilon_p) \right\},$$

$$-\ell_22(x^1, \bar{x}^2) = m + \sum_{i=1}^{m} g^{(2)}(\bar{P}_i + \varepsilon_p)\bar{P}_i^2 + G^{(2)}(\bar{P}_m + \varepsilon_p)\bar{P}_m^2,$$

$$I_{11} = \sum_{i=1}^{m} g^{(2)}(A_i), \ I_{12} = \sum_{i=1}^{m} g^{(2)}(A_i)A_i + G^{(2)}(A_m)A_m,$$

$$I_{22} = m + \sum_{i=1}^{m} g^{(2)}(A_i)A_i^2 + G^{(2)}(A_m)A_m^2.$$
and \(g^{(2)}\) and \(G^{(2)}\) are the second derivatives of \(g\) and \(G\). It is possible to show from (22) that \(R^1 = r(X^1)\) is the usual likelihood ratio statistic for \(\theta_p\).

For approximate conditional inference about the log-scale parameter \(\theta\), it is convenient to consider (24) in the case \(\varepsilon_p = 0\), i.e., \(P_i = (A_i - x^1)e^{-x^2} (i = 1, \ldots, m)\). For a specified value of \(x^2\), the corresponding value \(\tilde{x}^1 = \tilde{x}^1(x^2)\) of \(x^1\) that maximizes (24) satisfies

\[
\sum_{i=1}^{m} g^{(1)}(\tilde{P}_i) + G^{(1)}(\tilde{P}_m) = 0,
\]

where \(\tilde{P}_i = (A_i - \tilde{x}^1)e^{-x^2}\). Note that this definition of \(\tilde{P}_i\) is different than the one used in (25). The approximation to the conditional distribution function of \(X^2 = \theta - \hat{\theta}\) provided by (12) is

\[
\Pr(X^2 \leq x^2|A) \approx \Phi(r^2) + \phi(r^2) \left[ \frac{1}{r^2} + \frac{(I_{11}I_{22} - I_{12}^2)^{1/2}}{\ell_2(\tilde{x}^1, x^2\{ - \ell_1(\tilde{x}^1, x^2)\})^{1/2}} \right],
\]

where

\[
r^2 = r^2(x^2) = \text{sgn}(x^2)(2mx^2 + \sum_{i=1}^{m} \{g(\tilde{P}_i) - g(A_i)\} + \{G(\tilde{P}_m) - G(A_m)\})^{1/2},
\]

\[
\ell_2(\tilde{x}^1, x^2) = -m + \sum_{i=1}^{m} g^{(1)}(\tilde{P}_i)\tilde{P}_i + G^{(1)}(\tilde{P}_m)\tilde{P}_m,
\]

\[
-\ell_1(\tilde{x}^1, x^2) = e^{-2x^2} \left\{ \sum_{i=1}^{m} g^{(2)}(\tilde{P}_i) + G^{(2)}(\tilde{P}_m) \right\},
\]

and \(I_{11}, I_{12}, \text{ and } I_{22}\) are as defined for (26). Again, it is possible to show from (22) that \(R^2 = r(X^2)\) is the signed root of the usual likelihood ratio statistic for \(\theta\).

The accuracy of (26) and (28) is now illustrated in two examples.

**Example 1.** The normal distribution. Consider the case in which \(Z_1, \ldots, Z_n\) have the standard normal distribution and no censoring occurs. Then \(m = n, \varepsilon_p = \Phi^{-1}(p), \text{ and } g(z) = \{ \log 2\pi + z^2 \}/2\). For this situation, the conditional and unconditional distributions of \(X^1 = (y_p - \hat{y}_p)e^{-\theta}\) and \(X^2 = \theta - \hat{\theta}\) coincide. In particular,

\[
\Pr(X^1 \leq x^1) = \Pr(t'_{(n-1)}(n^{1/2}\varepsilon_p) \leq (n - 1)^{1/2}(x^1 + \varepsilon_p)),
\]

where \(t'_{(n-1)}(n^{1/2}\varepsilon_p)\) has the noncentral \(t\)-distribution with \((n - 1)\) degrees of freedom and noncentrality parameter \(n^{1/2}\varepsilon_p\), and

\[
\Pr(X^2 \leq x^2) = \Pr(x_{(n-1)}^2 \geq ne^{-2x^2}),
\]

11
where $\chi^2_{n-1}$ has the chi-squared distribution with $(n - 1)$ degrees of freedom.

Solving (25) gives

$$2e^{\hat{\beta}^2} = -\varepsilon_p(x^1 + \varepsilon_p) + \left\{4 + (4 + \varepsilon_p^2)(x^1 + \varepsilon_p)^2\right\}^{1/2},$$

and the components required in (21) are

$$\begin{align*}
r^1 &= \text{sgn}(x^1)[n\{2\hat{\beta}^2 + \varepsilon_p^2 - \varepsilon_p(x^1 + \varepsilon_p)e^{-\hat{\beta}^2}\}]^{1/2}, \\
\ell_1(x^1, \hat{\beta}^2) &= -n\varepsilon_p \{\varepsilon_p + (x^1 + \varepsilon_p)e^{-\hat{\beta}^2}\}, \quad -\ell_2(x^1, \hat{\beta}^2) = n[1 + \{1 + (x^1 + \varepsilon_p)^2\}e^{-2\hat{\beta}^2}], \\
(I_{11}I_{22} - I_{12}^2)^{1/2} &= \sqrt{2}n.
\end{align*}$$

Table 2 compares the approximations obtained from (1) and (26) with exact results for $p = 0.001, 0.05, 0.5$ with $n = 3$. When $p = 0.5$, $X^1$ has a symmetric distribution.

Solving (27) gives $\hat{x}^1 = 0$, and the components required in (28) are

$$\begin{align*}
r^2 &= \text{sgn}(x^2)[n\{2\hat{x}^2 - 1 + e^{-2\hat{x}^2}\}]^{1/2}, \\
\ell_2(\hat{x}^1, x^2) &= n\{1 + e^{-2\hat{x}^2}\}, \quad -\ell_{11}(\hat{x}^1, x^2) = ne^{-2\hat{x}^2}.
\end{align*}$$

Table 2 also compares approximations obtained from (1) and (28) with exact results for $n = 3$.

**EXAMPLE 2.** The Weibull distribution. Consider a sample $T_1, \ldots, T_n$ from the Weibull distribution having probability density function

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^\beta\right\}, \quad t > 0,$$

where $\alpha > 0$ and $\beta > 0$ are the scale and shape parameters of the distribution. If $Y_i = \log T_i$ ($i = 1, \ldots, n$), then $Y_i \sim \mu + \sigma Z_i$, where $\mu = \log \alpha$, $\sigma = \beta^{-1}$, and $Z_i$ has the extreme value distribution with density $f(z) = \exp\{-z - e^z\}$. A detailed discussion of inference for the Weibull distribution from the conditional point of view is given by Lawless (1978). In this case, $g(z) = -z + e^z$ and $G(z) = (n - m)e^z$. The solution $\hat{x}^2$ of (19) has no closed-form expression, and hence it must be found iteratively. On the other hand, the solution $\hat{x}^1$ of (27) is given by

$$e^{\hat{x}^1} = \left[\left\{\sum_{i=1}^m \exp(A_i e^{-x^2}) + (n - m)\exp(A_m e^{-x^2})\right\} / m\right]^{x^2}.$$
Using $n = 5, m = 3$, and the particular set of observations $-1.07606, -1.04493, -0.26842$ generated at random, Table 3 compares exact conditional results obtained by numerical integration of (23) with approximations (1) and (26) for $X^1$ when $p = 0.01$ and with (1) and (28) for $X^2$.

As can be seen from Tables 2 and 3, (26) and (28) provide excellent approximations in these examples. Moreover approximation (1) provides much inferior results. Since computation of the correction terms in (26) and (28) involves little more effort than is required to compute $r^1$ and $r^2$, approximations (26) and (28) are practically as simple to implement as approximation (1).

4. Multiple regression models

Consider observed random variables $Y_1, \ldots, Y_n$ such that $Y_i \sim \mu_i + \sigma Z_i$ ($i = 1, \ldots, n$), where $\mu_i = \sum u_{i2} \beta^2$ and $Z_1, \ldots, Z_n$ are independent and identically distributed with known probability density function $f$. The vectors $u_i = (u_{i1}, \ldots, u_{ip})^T$ ($i = 1, \ldots, n$) of covariate values are assumed to be known, and the vector $\beta = (\beta^1, \ldots, \beta^p)$ of regression coefficients is to be estimated. Based on the sample $Y = (Y_1, \ldots, Y_n)$, the likelihood function for $\beta$ and $\theta = \log \sigma$ is

$$L(\beta, \theta; Y) = \frac{1}{e^{n\theta}} \prod_{i=1}^{n} f \left( \frac{Y_i - \beta u_i}{e^{\theta}} \right).$$

As in the location-scale case, it has been argued, by Fraser (1979, §6.1), Lawless (1982, Appendix G), and Verhagen (1961), among others, that inferences about $\beta$ and $\theta$ should be based on the joint conditional distribution of the maximum likelihood estimators $\hat{\beta} = (\hat{\beta}^1, \ldots, \hat{\beta}^p)$ and $\hat{\theta}$ given the sample configuration $A = (A_1, \ldots, A_n)$, where $A_i = (Y_i - \hat{\beta} u_i) e^{-\hat{\theta}}$. If $\ell$ is defined by

$$\ell(x^1, \ldots, x^p, x^{p+1}) = \log L(x^1, \ldots, x^p, x^{p+1}; A)
= \log L(\hat{\beta}^1 + x^1 \hat{\sigma}, \ldots, \hat{\beta}^p + x^p \hat{\sigma}, \hat{\theta} + x^{p+1}; Y) + n \hat{\theta},$$

then the joint conditional density of the pivots $X^1 = (\beta^1 - \hat{\beta}^1)e^{-\hat{\theta}}, \ldots, X^p = (\beta^p - \hat{\beta}^p)e^{-\hat{\theta}}$, $X^{p+1} = \theta - \hat{\theta}$ satisfies

$$f_{X^1, \ldots, X^{p+1}|A}(x) \propto \exp \{ \ell(x) \}, \quad x = (x^1, \ldots, x^{p+1}).$$

Approximations to the marginal conditional distribution functions of $X^1, \ldots, X^{p+1}$
can thus be obtained from expression (12). To apply the approximation, note that

\[
\ell_s(x) = e^{-x^{p+1}} \sum_{i=1}^{n} u_{is}g^{(1)}(P_i), \quad \ell_{s+1}(x) = -n + \sum_{i=1}^{n} g^{(1)}(P_i)P_i
\]

\[
-\ell_{st}(x) = e^{-2x^{p+1}} \sum_{i=1}^{n} u_{is}u_{it}g^{(2)}(P_i), \quad -\ell_{s,p+1}(x) = e^{-x^{p+1}} \sum_{i=1}^{n} u_{is}\{g^{(1)}(P_i) + g^{(2)}(P_i)P_i\},
\]

\[
-\ell_{p+1,p+1}(x) = \sum_{i=1}^{n} \{g^{(1)}(P_i)P_i + g^{(2)}(P_i)P_i^2\}, \quad (s, t = 1, \ldots, p),
\]

where \(g(z) = -\log f(z)\), \(g^{(1)}\) and \(g^{(2)}\) are the first and second derivatives of \(g\), and \(P_i = \{A_i - (u_{i1}x^1 + \cdots + u_{ip}x^p)\}e^{-x^{p+1}} (i = 1, \ldots, n)\). It can be seen from (29) that the variable \(R^s\) arising in approximation (12) is the signed root of the usual likelihood ratio statistic for \(\beta^s\) when \(s = 1, \ldots, p\) and for \(\theta\) when \(s = p + 1\). Just as in the location–scale case, approximate inferences about the \(p\)th quantile of an observation corresponding to a specified set of covariate values can also be achieved through (12).

**Example 3.** Extreme value regression. Consider the case where \(Z_1, \ldots, Z_n\) have the extreme value distribution discussed in Example 2. Weibull and extreme value regression models are considered by Lawless (1982, §6.4). Since it is difficult to obtain exact conditional results from integrating (30) for any specific data set, a simulation of size 1,000 was performed. This investigation used \(n = 10, p = 6\), the matrix of covariate values

\[
U^T = (u_{is})^T = \begin{pmatrix}
0.686 & 0.640 & 0.908 & 0.886 & 0.508 & 0.255 & 0.197 & 0.056 & 0.646 & 0.317 \\
0.566 & 0.632 & 0.130 & 0.480 & 0.669 & 0.930 & 0.869 & 0.204 & 0.961 & 0.321 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0.255 & 0.197 & 0.056 & 0.646 & 0.317 \\
0 & 0 & 0 & 0 & 0 & 0.930 & 0.869 & 0.204 & 0.961 & 0.321 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\]

and parameter values \(\beta = (20, -20, 10, -2, 2, 1), \theta = 0\). For each data set generated, it was determined whether the approximate upper and lower 100(1 − \(\alpha\))% conditional confidence limits for \(\beta^1, \beta^3\), and \(\theta\) constructed from (1) and (12) actually covered the true parameter values. Thus, the approximations are assessed in terms of unconditional rather than conditional coverage. Table 4 reports the rates of noncoverage of the one–sided confidence intervals obtained from the 1,000 simulations.

As can be seen from Table 4, approximation (12) gives excellent results for the regression coefficients, but it performs poorly for the log–scale parameter \(\theta\). Presumably
approximation (19) would yield better results for $\theta$, although this was not attempted. Similar inaccuracy is found in the case of the log–scale parameter for normal regression, where (19) is easy to use. In that case, (19) provides very accurate approximations, as is demonstrated in the next example.

**Example 4.** Normal regression. Suppose that $Z_1, \ldots, Z_n$ have the normal distribution as discussed in Example 1. Here, the conditional and unconditional distributions of the pivots $X^1, \ldots, X^{p+1}$ coincide; in particular,

$$\Pr(X^{p+1} \leq x^{p+1}) = \Pr(\chi_{n-p}^2 \geq ne^{-2x^{p+1}}).$$

Approximation (12) gives

$$\Pr(X^{p+1} \leq x^{p+1}) \doteq \Phi(r) + \phi(r) \left\{ \frac{1}{r} + \left( \frac{2}{n} \right)^{1/2} \frac{e^{p+x^{p+1}}}{(e^{-2x^{p+1}} - 1)} \right\}, \quad (31)$$

where

$$r = \text{sgn}(x^{p+1}) \left\{ n \{2x^{p+1} - 1 + e^{-2x^{p+1}} \} \right\}^{1/2}.$$

To implement (19), it suffices to take

$$m(x^{p+1}) = -(n - p)x^{p+1} - (n/2)e^{-2x^{p+1}},$$

which gives $e^x = \{n/(n - p)\}^{1/2}$. Then, the alternative approximation is

$$\Pr(X^{p+1} \leq x^{p+1}) \doteq \Phi(r^*) + \phi(r^*) \left\{ \frac{1}{r^*} + \left( \frac{2(n - p)}{n(e^{-2x^{p+1}} - e^{-2\delta})} \right)^{1/2} \right\}, \quad (32)$$

where

$$r^* = \text{sgn}(x^{p+1} - \delta) \left\{ 2(n - p)(x^{p+1} - \delta) + n(e^{-2x^{p+1}} - e^{-2\delta}) \right\}^{1/2}.$$

Table 5 compares the use of (1), (20), (31), and (32) in the case $n = 10, p = 6$. In particular, Table 5 shows the true probabilities with which approximate upper and lower 100(1 $- \alpha$)% confidence limits fail to cover $\theta$. Approximation (32) gives excellent results in this situation, and (20) performs fairly well too, considering the small sample size. Both (1) and (31) perform poorly. Indeed, for this example, approximation (31) gives negative answers with probability 35.501%. To further illustrate the behaviour of (20) and (32), Table 2 shows results obtained for $X^2$ from these approximations in the location–scale case $p = 1$ with $n = 3$. In this example, $\delta$ corresponds to the usual correction for degrees of freedom.
5. Alternative derivations and the Lugannani–Rice formula

In this section, approximation (4) which applies in the univariate case \( p = 1 \) is derived by a method that involves direct integration of the density of \( X \) and thus avoids the moment calculations referred to in Section 2. Approximations (12) and (19) which pertain to the multivariate case \( p > 1 \) are also discussed from this perspective. Finally, connections between (4) and a tail probability approximation of Lugannani and Rice (1980) are explored.

Consider for the moment the univariate case \( p = 1 \). Following the notation of Sections 1 and 2, the density of \( X = X^1 \) is \( f(x) = c \exp\{\ell(x)\} \). If \( I = -\ell^{(2)}(0) \), then the density of \( Y = XI^{1/2} \) is \( g(y) = \bar{c} \exp\{\bar{\ell}(y)\} \), where \( \bar{\ell}(y) = \ell(yI^{-1/2}) \) and \( \bar{c} \) is a normalizing constant. The crucial observations are that \( \bar{\ell}^{(1)}(0) = 0 \) and \( -\bar{\ell}^{(2)}(0) = 1 \); in addition, \( \bar{\ell}^{(3)}(0) = \ell^{(3)}(0)I^{-3/2} = F_3 \) and \( \bar{\ell}^{(4)}(0) = \ell^{(4)}(0)I^{-2} = F_4 \). From (7), the function

\[
\begin{align*}
  r &= \text{sgn}(x) \left[ 2\{\ell(0) - \ell(x)\} \right]^{1/2} = \text{sgn}(y) \left[ 2\{\bar{\ell}(0) - \bar{\ell}(y)\} \right]^{1/2}
\end{align*}
\]

has the expansion

\[
\begin{align*}
  r &= y - \frac{1}{6}y^2F_3 - \frac{1}{24}y^3 \left( F_4 + \frac{1}{3}F_3^2 \right) + O_p(n^{-3/2}),
\end{align*}
\]

provided \( y \) is \( O_p(1) \). Inversion of (33) yields

\[
\begin{align*}
  y &= r + \frac{1}{6}r^2F_3 + \frac{1}{24}r^3 \left( F_4 + \frac{5}{3}F_3^2 \right) + O_p(n^{-3/2});
\end{align*}
\]

hence,

\[
\begin{align*}
  \frac{dy}{dr} &= \exp \left\{ \frac{1}{3}rF_3 + \frac{1}{24}r^2 \left( 3F_4 + \frac{11}{3}F_3^2 \right) + O_p(n^{-3/2}) \right\}. 
\end{align*}
\]

Using (33) and (35) in conjunction with the expansion

\[
\begin{align*}
  g(y) &= \bar{c} \exp \left\{ \bar{\ell}(0) - \frac{1}{2}y^2 + \frac{1}{6}y^3F_3 + \frac{1}{24}y^4F_4 + O_p(n^{-3/2}) \right\},
\end{align*}
\]

shows that the probability density function of \( R = r(X) \) is proportional to

\[
\begin{align*}
  \exp \left\{ -\frac{1}{2}r^2 + \frac{1}{3}rF_3 + \frac{1}{24}r^2 \left( 3F_4 + \frac{11}{3}F_3^2 \right) + O_p(n^{-3/2}) \right\},
\end{align*}
\]

and the normalizing constant is \( (2\pi)^{1/2} h \), where

\[
\begin{align*}
  h &= 1 + \frac{1}{8} \left( F_4 + \frac{5}{3}F_3^2 \right) + O_p(n^{-3/2}).
\end{align*}
\]
If \( \zeta \) is defined by \( \zeta = -\bar{\epsilon}^{(1)}(y) \), then \( \zeta = -I^{-1/2} \bar{\epsilon}^{(1)}(x) \); moreover,

\[
\zeta = \frac{1}{2} \frac{d(y^2)}{dy} = r \left( \frac{dr}{dy} \right),
\]

and combining (35), (36), and (37), the probability density function of \( R \) is

\[
h\phi(r) \left( \frac{dy}{dr} \right) \left\{ 1 + \mathcal{O}_p(n^{-3/2}) \right\} = h\phi(r) \frac{r}{\zeta} \left\{ 1 + \mathcal{O}_p(n^{-3/2}) \right\}.
\]

In the calculations that follow, errors of order \( \mathcal{O}_p(n^{-3/2}) \) are notationally suppressed. Now,

\[
\Pr(X \leq x) = \Pr(Y \leq y) = h \int_{-\infty}^{r} \phi(r) \frac{r}{\zeta} dr + \mathcal{O}_p(n^{-3/2}),
\]

and integration by parts yields

\[
\Pr(X \leq x) = h \left\{ -\frac{\phi(r)}{\zeta} + \int_{-\infty}^{r} \phi(r) \left( \frac{d\zeta^{-1}}{dr} \right) dr \right\}.
\]

It follows from (34) and (37) that

\[
\frac{d\zeta^{-1}}{dr} = -r^{-2} \left\{ \frac{dy}{dr} - r \frac{d^2 y}{(dr)^2} \right\} = -r^{-2} \left\{ 1 - \frac{1}{8} r^2 \left( F_4 + \frac{5}{3} F_3^2 \right) + \mathcal{O}_p(n^{-3/2}) \right\}
\]

\[
= -(1 + r^{-2}) + h^{-1}.
\]

Hence,

\[
\Pr(X \leq x) = h \left[ -\frac{\phi(r)}{\zeta} + \int_{-\infty}^{r} \phi(r) \left\{ -(1 + r^{-2}) + h^{-1} \right\} dr \right]
\]

\[
= h \left\{ -\frac{\phi(r)}{\zeta} + \frac{\phi(r)}{r} + h^{-1} \Phi(r) \right\} = \Phi(r) + \phi(r) \left\{ \frac{1}{r} - \frac{1}{\zeta} \right\},
\]

which agrees with (4)

Consider now the multivariate case. The derivation of approximation (19) from (4) by means of (17) and (18) is discussed in Section 2. In a similar manner, (12) can also be deduced from (4). Starting from the Laplace approximation (18) to the marginal density of \( X^1 \), consider a variable \( Y = y(X^1) \) such that

\[
\frac{dy}{dx^1} = \left\{ \frac{|(I_{x^1})|}{\left| -\ell_{x^1} (\tilde{x}(x^1)) \right|} \right\}^{1/2},
\]

where
and \( y(0) = 0 \). Then the density of \( Y \) is proportional to \( \exp \{ \bar{L}(y) \} \), where \( \bar{L}(y) = \bar{L}(y(x^1)) = L(\hat{x}(x^1)) \). Using (5), it can be shown that \( \bar{L}^{(1)}(0) = 0 \) and \( -\bar{L}^{(2)}(0) = 1 \); therefore, the preceding argument applies to \( Y \). In this case

\[
\zeta = -\bar{L}^{(1)}(y) = \ell_1(\hat{x}(x^1)) \frac{dx^1}{dy} = \ell_1(\hat{x}(x^1)) \frac{|I|^{-1/2}}{|-\ell_{\alpha\beta}(\hat{x}(x^1))|^{-1/2}},
\]

and (38) yields formula (12).

Thus, formulae (12) and (19) differ in the way that they account for the determinant term in the Laplace approximation (18). For approximation (12), this term is made constant by transforming \( X^1 \) to a new variable \( Y \) having a different scale. For (19), it is incorporated directly into the variable \( R^* \), upon whose asymptotic normal distribution the approximation is based.

The relationship between approximation (4) and the Lugannani–Rice (LR) approximation for tail areas of a mean is now considered. The formulae are of the same form, but they differ in some essential ways. For the case of a single observation from the extreme value distribution with density \( f(x) = \exp\{x - e^x\} \), it is possible to apply both approximations. Table 6 gives some numerical results. Both approximations perform well in this case, but they give different numerical results. In order to understand the relationship between the two, it is useful to consider the LR approximation in the context of an exponential family.

The LR approximation is (Daniels, 1987, formula 4.9)

\[
pr(\bar{X} \geq x) = 1 - \Phi(r) + \phi(r) \left\{ \frac{1}{\hat{\varepsilon}} - \frac{1}{r} \right\} + O(n^{-3/2})
\]

(39)

where \( \hat{\varepsilon} = \hat{T} \{ nK^{(2)}(\hat{T}) \}^{1/2}, r = \text{sgn}(\hat{T}) [2n \{ \hat{T} x - K(\hat{T}) \}]^{1/2}, K(t) \) is the cumulant generating function of the underlying observations, and \( K^{(1)}(\hat{T}) = x \). The point \( \hat{T} \) is referred to as the saddlepoint. The development of (39) begins by taking the Fourier inversion of the characteristic function for the mean and then integrating this density from \( x \) to \( \infty \), which leads to the integral

\[
\frac{n}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp \left\{ n \{ K(z) - zx_0 \} \right\} z^{-1} \, dz.
\]

The approximation proceeds by transforming the exponent to a quadratic and taking two terms in a Taylor expansion of the resulting jacobian. In the case of an exponential family

\[
f_\theta(x) = \exp \{ \theta u(x) - K(\theta) + d(x) \}, K(z) - zx_0 \text{ is replaced by } K_\delta(z) = K(\theta + z) - K(\theta) - zK^{(1)}(\hat{\theta}),
\]

which is the cumulant generating function of the score function \( u(X) - K^{(1)}(\theta) \).
To approximate \( \Pr(\hat{\theta} \geq \hat{\theta}_0) \), note that the saddlepoint is given by \( \hat{\theta}_0 - \theta \). The approximation follows directly, and it is given by (16) with
\[
 r = \text{sgn}(\hat{\theta}_0 - \theta) \left[ 2n \left\{ K(\theta) - K(\hat{\theta}_0) + (\hat{\theta}_0 - \theta)K^{(1)}(\hat{\theta}_0) \right\} \right]^{1/2} \\
= \text{sgn}(\hat{\theta}_0 - \theta) \left[ 2 \{ \ell(\hat{\theta}_0) - \ell(\theta) \} \right]^{1/2},
\]
where \( \ell \) is the log likelihood function based on the sample of size \( n \), and with
\[
 \hat{\varepsilon} = (\hat{\theta}_0 - \theta) K^{(2)}(\hat{\theta}_0) \right]^{1/2}.
\]

This approximation is given in Fraser (1988, formula 1.9) and it appears in a more general form in Daniels (1983, formula 6.1).

To compare the LR approximation and formula (4) when applied to approximating \( \Pr(X \geq x_0) \) in a general context, it is helpful to view both approximations as being obtained by incorporating an exponential family into the problem, and using the exponential family approximation given above. For the LR approximation, the appropriate exponential family is given by
\[
c(x) \exp \left\{ T(x_0)(x - x_0) + \log f(x) \right\},
\]
where \( T(x_0) \) is the saddlepoint. Now, using the exponential family result to approximate \( \Pr(\hat{T} \geq T(x_0)) \), the LR approximation follows directly by noting that \( \Pr(X \geq x_0) = \Pr(\hat{T} \geq T(x_0)) \) because of the monotonicity of the saddlepoint.

For formula (4), Fraser (1988) shows that the appropriate exponential family is proportional to
\[
 \exp \left\{ \lambda(\phi) s + \ell(\mu x - \phi) \right\}
\]
where \( \ell(x) = \log f(x) \), \( \lambda(\phi) = \ell^{(1)}(x - \phi_0) / \{ - \ell^{(2)}(0) \} \), and \( s = \ell^{(1)}(x - \phi) \). Again, the exponential family result for \( \Pr(\hat{\phi} \geq \phi_0) \), where \( \phi_0 = \hat{\phi}(x_0) = x_0 \) gives approximation (4) for \( \Pr(X \geq x_0) \).

Thus, both approximations can be viewed as incorporating an exponential family into the problem and then using the LR approximation for the exponential family. The exponential families used for the two approximations are different, and the approximations behave differently. For instance, (4) gives quite accurate results for Cauchy's distribution, while the LR approximation cannot be applied in that case because there is not a moment generating function that ensures the existence of a real saddlepoint. On the other hand, it is possible to construct densities where the LR formula gives reasonable results while (4) breaks down. Implicit in the development given by Fraser is an assumption that \( \hat{\phi} \) is monotone. If this monotonicity does not hold, (4) can fail to give sensible results.
To conclude the comparison, note that it is possible to view (4) from the perspective of a saddlepoint approximation. A comparison of (4) and (39) suggests that in (4), the role of saddlepoint is being played by $-\ell^{(1)}(x)$. To see that this interpretation is plausible, consider the saddlepoint equation $K^{(1)}(T) = x_0$, which can be written as

$$\int (x - x_0) \exp \{ T(x - x_0) + \ell(x) \} \, dx = 0.$$  

If we replace $T$ by $-\ell^{(1)}(x_0)$, and if $\ell(x)$ and $\ell^{(1)}(x)$ are both $O_p(n)$, then a direct application of Laplace's approximation gives

$$\int (x - x_0) \exp \{ -\ell^{(1)}(x_0)(x - x_0) + \ell(x) \} \, dx = O_p(n^{-3/2}).$$  

This suggests that (4) may be viewed as a version of the LR approximation using an approximate saddlepoint. However, the relationships between these approximations are not clear and deserves further attention.
References


Fraser, D. A. S. (1988). Tail probabilities from observed likelihoods. Submitted to *Biometrika*.


Table 1. Approximations to $\text{pr}(X \leq x)$ for Fisher’s example of the Nile problem; $A = \frac{1}{2}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>-3.0</th>
<th>-2.5</th>
<th>-2.0</th>
<th>-1.5</th>
<th>-1.0</th>
<th>-0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.02103</td>
<td>0.0678</td>
<td>0.938</td>
<td>5.002</td>
<td>14.866</td>
<td>49.601</td>
</tr>
<tr>
<td>(2)</td>
<td>0.04434</td>
<td>0.0108</td>
<td>0.332</td>
<td>2.878</td>
<td>11.441</td>
<td>49.539</td>
</tr>
<tr>
<td>(4)</td>
<td>0.03408</td>
<td>0.0332</td>
<td>0.563</td>
<td>3.574</td>
<td>12.349</td>
<td>49.551</td>
</tr>
<tr>
<td>Exact</td>
<td>0.03460</td>
<td>0.0374</td>
<td>0.615</td>
<td>3.831</td>
<td>12.878</td>
<td>49.563</td>
</tr>
</tbody>
</table>

Table entries are percentages.
Table 2. Approximations to $\text{pr}(X^1 \leq x^1)$ and $\text{pr}(X^2 \leq x^2)$ for Example 1; $n = 3$

<table>
<thead>
<tr>
<th>$X^1, p = 0.001$</th>
<th>$x^1$</th>
<th>-200</th>
<th>-25</th>
<th>-10</th>
<th>0.01*</th>
<th>0.75*</th>
<th>1.5*</th>
<th>2.5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.03149</td>
<td>0.0782</td>
<td>0.911</td>
<td>49.712</td>
<td>25.636</td>
<td>4.878</td>
<td>0.02407</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>0.0341</td>
<td>1.766</td>
<td>7.852</td>
<td>25.510</td>
<td>10.833</td>
<td>1.631</td>
<td>0.02144</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0359</td>
<td>1.859</td>
<td>8.242</td>
<td>24.271</td>
<td>10.071</td>
<td>1.461</td>
<td>0.02124</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X^1, p = 0.05$</th>
<th>$x^1$</th>
<th>-100</th>
<th>-15</th>
<th>-5</th>
<th>0.01*</th>
<th>0.5*</th>
<th>1.25*</th>
<th>3*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.03242</td>
<td>0.0745</td>
<td>1.398</td>
<td>49.548</td>
<td>25.277</td>
<td>1.878</td>
<td>0.02177</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>0.0410</td>
<td>1.511</td>
<td>8.983</td>
<td>29.958</td>
<td>14.025</td>
<td>1.222</td>
<td>0.02827</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0441</td>
<td>1.626</td>
<td>9.613</td>
<td>28.261</td>
<td>13.017</td>
<td>1.154</td>
<td>0.02878</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X^1, p = 0.5$</th>
<th>$x^1$</th>
<th>-30</th>
<th>-5</th>
<th>-3</th>
<th>-2</th>
<th>-1.5</th>
<th>-0.5</th>
<th>-0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.03312</td>
<td>0.0885</td>
<td>0.429</td>
<td>1.400</td>
<td>3.003</td>
<td>20.663</td>
<td>49.309</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>0.0256</td>
<td>0.896</td>
<td>2.377</td>
<td>4.926</td>
<td>7.899</td>
<td>26.976</td>
<td>49.482</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0278</td>
<td>0.971</td>
<td>2.566</td>
<td>5.279</td>
<td>8.397</td>
<td>27.639</td>
<td>49.500</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X^2$</th>
<th>$x^2$</th>
<th>-0.9</th>
<th>-0.5</th>
<th>-0.25</th>
<th>0.01</th>
<th>1.25*</th>
<th>2.50*</th>
<th>5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.0897</td>
<td>7.106</td>
<td>25.208</td>
<td>50.974</td>
<td>1.468</td>
<td>0.0263</td>
<td>0.04102</td>
<td></td>
</tr>
<tr>
<td>(28)</td>
<td>0.0121</td>
<td>1.768</td>
<td>8.710</td>
<td>23.537</td>
<td>11.304</td>
<td>0.978</td>
<td>0.02663</td>
<td></td>
</tr>
<tr>
<td>(20)</td>
<td>0.0306</td>
<td>3.373</td>
<td>14.333</td>
<td>34.020</td>
<td>5.931</td>
<td>0.363</td>
<td>0.02169</td>
<td></td>
</tr>
<tr>
<td>(32)</td>
<td>0.0117</td>
<td>1.709</td>
<td>8.460</td>
<td>22.973</td>
<td>11.830</td>
<td>1.055</td>
<td>0.02730</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0115</td>
<td>1.695</td>
<td>8.432</td>
<td>22.986</td>
<td>11.585</td>
<td>1.006</td>
<td>0.02681</td>
<td></td>
</tr>
</tbody>
</table>

Table entries are percentages. For cases marked *, reported results pertain to $\text{pr}(X^1 \geq x^1)$ and $\text{pr}(X^2 \geq x^2)$. 

24
### Table 3. Approximations to \( \text{pr}(X^1 \leq x^1) \) and \( \text{pr}(X^2 \leq x^2) \) for Example 2; \( n = 5, m = 3 \)

<table>
<thead>
<tr>
<th>( X^1, p = 0.01 )</th>
<th>( x^1 )</th>
<th>(-289)</th>
<th>(-59)</th>
<th>(-15)</th>
<th>(-1^*)</th>
<th>(1^*)</th>
<th>(2.5^*)</th>
<th>(4^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.03249</td>
<td>0.0300</td>
<td>1.190</td>
<td>64.993</td>
<td>30.235</td>
<td>3.658</td>
<td>0.02126</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>0.0470</td>
<td>0.971</td>
<td>9.163</td>
<td>38.355</td>
<td>13.689</td>
<td>1.283</td>
<td>0.02251</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0502</td>
<td>1.013</td>
<td>9.554</td>
<td>36.995</td>
<td>12.912</td>
<td>1.206</td>
<td>0.02258</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( X^2 )</th>
<th>( x^2 )</th>
<th>(-1.1)</th>
<th>(-0.7)</th>
<th>(-0.3)</th>
<th>(0.3)</th>
<th>(1.5^*)</th>
<th>(2.7^*)</th>
<th>(6.5^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.415</td>
<td>5.887</td>
<td>26.650</td>
<td>71.122</td>
<td>1.212</td>
<td>0.0307</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>0.0567</td>
<td>1.296</td>
<td>8.926</td>
<td>39.575</td>
<td>10.575</td>
<td>1.137</td>
<td>0.0607</td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>0.0556</td>
<td>1.258</td>
<td>8.701</td>
<td>39.082</td>
<td>10.711</td>
<td>1.149</td>
<td>0.0613</td>
<td></td>
</tr>
</tbody>
</table>

Table entries are percentages. For cases marked *, reported results pertain to \( \text{pr}(X^1 \geq x^1) \) and \( \text{pr}(X^2 \geq x^2) \).
Table 4. Simulated rates of noncoverage for Example 3; \( n = 10, p = 6, 1000 \) trials

<table>
<thead>
<tr>
<th>Upper</th>
<th>( 100\alpha )</th>
<th>( \beta^1 ) (1) (12)</th>
<th>( \beta^3 ) (1) (12)</th>
<th>( \theta ) (1) (12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>4.9 0.6</td>
<td>5.8 0.9</td>
<td>36.0 0.2</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>6.7 1.2</td>
<td>7.4 1.5</td>
<td>42.7 0.5</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>10.9 3.1</td>
<td>10.3 2.9</td>
<td>55.3 1.1</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>14.1 5.3</td>
<td>14.2 6.0</td>
<td>65.2 2.0</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>19.2 9.4</td>
<td>18.6 10.6</td>
<td>73.8 3.5</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>32.0 22.6</td>
<td>30.2 23.6</td>
<td>88.3 10.8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lower</th>
<th>( 100\alpha )</th>
<th>( \beta^1 ) (1) (12)</th>
<th>( \beta^3 ) (1) (12)</th>
<th>( \theta ) (1) (12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>7.7 0.6</td>
<td>7.6 0.7</td>
<td>0.0 49.6</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>9.4 0.9</td>
<td>10.2 0.9</td>
<td>0.0 50.1</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>12.8 1.9</td>
<td>15.2 2.2</td>
<td>0.0 50.7</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>17.7 5.3</td>
<td>19.8 4.6</td>
<td>0.0 52.5</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>23.7 11.3</td>
<td>26.4 11.4</td>
<td>0.0 55.9</td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td>33.5 28.6</td>
<td>37.9 27.1</td>
<td>1.0 64.8</td>
<td></td>
</tr>
</tbody>
</table>

Tables entries are percentages.
<table>
<thead>
<tr>
<th>100α</th>
<th>25</th>
<th>10</th>
<th>5</th>
<th>2.5</th>
<th>1</th>
<th>0.5</th>
<th>0.005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>(1)</td>
<td>87.814</td>
<td>74.280</td>
<td>63.631</td>
<td>53.508</td>
<td>41.614</td>
<td>33.945</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>33.425</td>
<td>15.183</td>
<td>8.248</td>
<td>4.444</td>
<td>1.943</td>
<td>1.032</td>
</tr>
<tr>
<td></td>
<td>(31)</td>
<td>12.853</td>
<td>4.953</td>
<td>2.438</td>
<td>1.207</td>
<td>0.479</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>(32)</td>
<td>24.955</td>
<td>9.956</td>
<td>4.967</td>
<td>2.478</td>
<td>0.988</td>
<td>0.493</td>
</tr>
<tr>
<td>Lower</td>
<td>(1)</td>
<td>0.978</td>
<td>0.204</td>
<td>0.0699</td>
<td>0.0253</td>
<td>0.02700</td>
<td>0.02274</td>
</tr>
<tr>
<td></td>
<td>(20)</td>
<td>18.315</td>
<td>6.649</td>
<td>3.137</td>
<td>1.492</td>
<td>0.564</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td>(31)</td>
<td>56.337</td>
<td>44.733</td>
<td>40.338</td>
<td>37.987</td>
<td>36.514</td>
<td>36.011</td>
</tr>
<tr>
<td></td>
<td>(32)</td>
<td>24.988</td>
<td>9.984</td>
<td>4.988</td>
<td>2.492</td>
<td>0.996</td>
<td>0.498</td>
</tr>
</tbody>
</table>

Table entries are percentages.
TABLE 6. Approximations to $\Pr(X \leq x)$ for the extreme value distribution

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2*</th>
<th>2.5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4)</td>
<td>0.0132</td>
<td>0.707</td>
<td>31.043</td>
<td>0.0628</td>
<td>0.03526</td>
</tr>
<tr>
<td>LR</td>
<td>0.0124</td>
<td>0.799</td>
<td>28.111</td>
<td>0.0776</td>
<td>0.03245</td>
</tr>
<tr>
<td>Exact</td>
<td>0.0123</td>
<td>0.672</td>
<td>30.780</td>
<td>0.0618</td>
<td>0.03512</td>
</tr>
</tbody>
</table>

Table entries are percentages. For case marked *, reported results pertain to $\Pr(X \geq x)$. 