Approximations of marginal tail probabilities and inference for scalar parameters

BY THOMAS J. DiCICCIO AND CHRISTOPHER A. FIELD

Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.

AND D. A. S. FRASER

Department of Mathematics, York University, North York, Ontario, M3J 1P3, Canada

Summary

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

Some key words: Bayesian inference; Conditional inference; Linear regression model; Location-scale model; Lugannani–Rice formula; 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. Exact calculation of this distribution function usually involves multidimensional numerical integration which can be difficult to implement in practice. In the Bayesian context, determination of marginal posterior tail probabilities can present similar difficulties.

This paper contains two approximations of marginal tail probabilities that are applicable in such situations. For the univariate case, the approximations coincide and they are similar to approximations given by Fraser (1990) and Barndorff-Nielsen (1988). The errors in both approximations are of order $O(n^{-3/2})$, where $n$ is the sample size. The first approximation is simple to compute, requiring only first- and second-order partial derivatives of the log joint 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 $X = (X^1, \ldots, X^n)$ be a vector variable that is $O_p(n^{-1})$ 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

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

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

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

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

Another example for which $p = 1$ is provided by an instance of Fisher's (1973, §6.9) problem of the Nile. In this, $Y_1, \ldots, Y_n$ 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 $\hat{\tau} = (\frac{1}{n} \sum Y_j / \sum Z_j)\bar{y}$, and the distribution of $A = (\sum Y_j / \sum Z_j)^\delta$ does not depend on $\tau$. With $\theta = \log \tau$ 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 about parameters of location-scale or linear regression models. A detailed discussion of such examples is given in §§3 and 4.

The approximations developed here for the distribution of $X^1$ involve the function

$$w^1(x^1) = 2[r(0) - l(\tilde{x}(x^1))],$$

and its signed square root

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

where $\tilde{x}(x^1)$ is the point at which $l(x)$ 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 §§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 §2 that the standard normal approximation to the distribution of $R^1$ has error of order $O(n^{-1})$. Thus provided $x^1$ is $O(n^{-1})$,

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

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 $O(n^{-1})$, and that by accounting for both the mean and variance of $R^1$ the error can be reduced further to order $O(n^{-3/2})$. Thus

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

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 distributions of signed roots of likelihood ratio statistics have been discussed by Barndorff-Nielsen (1986), DiCiccio (1984, 1988), Efron (1985) 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
Approximations of marginal tail probabilities. 

for $\mu$ and $\sigma^2$ are given in expression (9). For the univariate case $p = 1$, $x = x^1$ and it suffices to take

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

where $I^{(j)}(x) = d^j I(x)/dx^j$ ($j = 1, \ldots, 4$). 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 $l$. The use of (2) in such cases can be intractable.

A procedure is presented in §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 (12); for the case $p = 1$, it reduces to

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

where $r = r(x) = \text{sgn}(x) [2[I(0)-l(x)]^{3/2}]$ and $\phi$ is the standard normal density. Approximations that are equivalent to (4) have recently been given by Barndorff-Nielsen (1988) and Fraser (1990).

Approximation (4) is seen to be intuitively reasonable for small $x$, since in that case $r$ and $-I^{(1)}(x)/-I^{(2)}(0)$ are nearly equal and the tail probability $\Pr(X \geq x)$ is expected to behave like $1 - \Phi(r)$. For large deviations, $1 - \Phi(r)$ is cancelled by its asymptotic form $\phi(r)/r$, and that tail probability behaves like $\phi(r)/[-I^{(2)}(0)]^{3/2}/[-I^{(1)}(x)]$.

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 $\sigma^2 = 1 - (8A)^{-1}$, which are used for approximation (2) in Table 1. For this problem, although (1) is already fairly accurate, the use of (4) offers some improvement, particularly in the extreme tails. For the examples considered in §§3 and 4, the improvement afforded by (12) over (1) is more remarkable.

Table 1. Approximations to $\Pr(X \leq x | A)$ for Fisher’s Nile problem; $A = \frac{1}{2}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>(1)</th>
<th>(2)</th>
<th>(4)</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-3.0$</td>
<td>0.0103</td>
<td>0.00678</td>
<td>0.0938</td>
<td>5.002</td>
</tr>
<tr>
<td>$-2.5$</td>
<td>0.0103</td>
<td>0.0108</td>
<td>0.0332</td>
<td>2.878</td>
</tr>
<tr>
<td>$-2.0$</td>
<td>0.0108</td>
<td>0.0332</td>
<td>0.0563</td>
<td>3.574</td>
</tr>
<tr>
<td>$-1.5$</td>
<td>0.0374</td>
<td>0.0615</td>
<td>3.831</td>
<td>12.878</td>
</tr>
<tr>
<td>$-1.0$</td>
<td>0.0374</td>
<td>0.0615</td>
<td>3.831</td>
<td>12.878</td>
</tr>
<tr>
<td>$-0.01$</td>
<td>0.0374</td>
<td>0.0615</td>
<td>3.831</td>
<td>12.878</td>
</tr>
</tbody>
</table>

Table entries are percentages.

Since $l(x)$ is assumed to be maximized at $x = 0$, it follows that $\hat{x}(0) = 0$, and hence $r^1(0) = 0$. Thus approximation (12) breaks down when $x^1 = 0$. In the examples, however, this property appears not to be a problem since the approximation yields quite accurate results for values of $x^1$ very near 0. For instance, in Table 1, (4) works well for $x = -0.01$.

In some situations, where $p$ is large and $n$ is small, approximation (12) can be poor; indeed, it is possible for the approximation to be negative. Such cases are illustrated in Examples 3 and 4 of §4. The nature of this failure is appreciated by observing that the leading component of (12), $\Phi(r^1)$, takes $x^1 = 0$ as the median of $X^1$; note that when
$x^1 = 0$, $\Phi(r^*) = \Phi(0) = 0.5$. If the marginal density of $X^1$ is actually centered at a point substantially away from 0, then $\Phi(r^*)$ can be a very poor initial approximation to $\Pr(X^1 \leq x^1)$, and the correction term in (12) is required to make a large adjustment which it cannot accurately achieve. An alternative tail probability approximation that seems to avoid 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 compute than (19). A case where (19) provides substantial improvement over (12) is discussed in Example 4 of § 4.

Section 5 contains derivations of approximations (12) and (19) that are different than those given in § 2. Comparisons between (4) and the Lugannani & Rice (1980) approximation for tail probabilities are also discussed in § 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 some additional notation. Let $I_{ij}(x) = \partial l(x) / \partial x^i \partial x^j$, etc., and set

$$I_{ij} = -I_{ji}(0), \quad I_{ijk} = I_{ijk}(0), \quad I_{ijkl} = I_{ijkl}(0), \quad (i, j, k, l = 1, \ldots, p).$$

Furthermore, take $I = (I_{ij})$, $I^{-1} = (I^i_j)$, and

$$F_{3,1} = I_{ijk} I^{j1} I^{k1} (I^{11})^{-3/2}, \quad F_{3,2} = I_{ijk} I^j I^{k1} (I^{11})^{-3},
F_{4,1} = I_{ijkl} I^{ij} I^{kl} (I^{11})^{-2}, \quad F_{4,2} = I_{ijkl} I^j I^{k1} (I^{11})^{-1}, \quad F_{4,3} = I_{ijkl} I^j I^{kl},
F_{6,1} = I_{imnn} I^{i1} I^{m1} I^{n1} (I^{11})^{-3}, \quad F_{6,2} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-2},
F_{6,3} = I_{imnn} I^{i1} I^{m1} I^{n1} (I^{11})^{-2}, \quad F_{6,4} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-1},
F_{6,5} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-1}, \quad F_{6,6} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-1},
F_{6,7} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-1}, \quad F_{6,8} = I_{imnn} I^i I^{m1} I^{n1} (I^{11})^{-1}.$$

In these and subsequent expressions, the usual notational convention is followed whereby summation over the range 1, \ldots, $p$ is understood for each index that appears 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 $\tilde{x}(x^1) = \{\tilde{x}(x^1), \ldots, \tilde{x}(x^1)\}$, it follows that $\tilde{x}(0) = 0$, $\tilde{x}(x^1) = x^1$ and $l_a(\tilde{x}(x^1)) = 0$ ($a = 2, \ldots, p$). Differentiation of $l_a(\tilde{x}(x^1)) = 0$ with respect to $x^1$ shows that

$$\left[ \frac{d\tilde{x}(x^1)}{dx^1} \right]_{x^1 = 0} = I^{i1}(I^{11})^{-1},$$

$$\left[ \frac{d^2\tilde{x}(x^1)}{(dx^1)^2} \right]_{x^1 = 0} = \{I^i_j - I^{i1}(I^{11})^{-1}I_{ijkl} I^{k1} (I^{11})^{-2},
$$

and hence, provided $x^1$ is of order $O(n^{-1})$,

$$l(\tilde{x}(x^1)) = l(0) - \frac{1}{2} \{x^1(I^{11})^{-1}\}^2 + \frac{1}{6} \{x^1(I^{11})^{-1}\}^3 F_3 + \frac{1}{24} \{x^1(I^{11})^{-1}\}^4 F_4 + O(n^{-3/2}),$$

where $F_3 = F_{3,1}$ and $F_4 = F_{4,1} + 3 F_{6,3} - 3 F_{6,1}$. Expression (6) yields

$$w^1(x^1) = x^1(I^{11})^{-1}\{1 - 6 x^1(I^{11})^{-1}\} F_3 - \frac{1}{6} \{x^1(I^{11})^{-1}\}^3 F_4 + O(n^{-3/2}),$$

$$r^1(x^1) = x^1(I^{11})^{-1}\{1 - 6 x^1(I^{11})^{-1}\} F_3 - \frac{1}{6} \{x^1(I^{11})^{-1}\}^3 F_4 + O(n^{-3/2}).$$
In these expressions, \( F_3 \) is \( O(n^{-1}) \), while \( I^{11}, F_4 \) and \( F_5^{2} = F_{3,1} = F_{6,1} \) are all \( O(n^{-1}) \). Related expansions have been given by Sprott (1980).

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

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

(8)

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

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

(9)

where

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

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

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

The approximation to the distribution function of \( X^1 \) that arises from the adjusted normal approximation for \( R^1 \) is

\[
\Pr(X^1 \leq x^1) = \Pr(R^1 \leq r^1) = \Phi\{ (r^1 - a)(1 + b - a^2)^{-1} \} + O(n^{-3/2})
\]

\[
= \Phi(r^1) - \phi(r^1) (a + \frac{1}{2} r^1 b) + O(n^{-3/2})
\]

\[
= \Phi(r^1) - \phi(r^1) (a + \frac{1}{2} x^1 (I^{11})^{-1} b) + O(n^{-3/2}),
\]

(10)

where \( x^1 \) is \( O(n^{-1}) \), \( r^1 = r^1(x^1) \), 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 for problems where \( p \) is large. A straightforward approximation to the term \( -\{a + \frac{1}{2} x^1 (I^{11})^{-1} b\} \) is helpful in applications.

A simple approximation to that term, verified below, is provided by the formula

\[
-\{a + \frac{1}{2} x^1 (I^{11})^{-1} b\} = \frac{1}{r^1 + \frac{1}{(I^{11})^{1/2} l_1(x^1)} \{ \frac{|I_{ab}|}{[-l_{ab}(x^1)]} \}^{1/3}} + O(n^{-3/2}),
\]

(11)

where \( \alpha \) and \( \beta \) vary over the range \( 2, \ldots, p \); that is, \( (I_{ab}) \) is the \( (p-1) \times (p-1) \) submatrix of \( I \) corresponding to \( x^2, \ldots, x^p \), and similarly for \( [-l_{ab}(x^1)] \). Approximation (11) allows the mean and variance of \( R^1 \) to be taken into account using only first- and second-order partial derivatives of \( l \). Since \( |I| = (I^{11})^{-1} |I_{ab}| \), it follows from (10) and (11) that the marginal distribution function of \( X^1 \) can be approximated by

\[
\Pr(X^1 \leq x^1) = \Phi(r^1) + \phi(r^1) \left( \frac{1}{r^1 + \frac{1}{l_1(x^1)} \{ |I^{11}| [-l_{ab}(x^1)] \}^{1/3}} \right) + O(n^{-3/2}),
\]

(12)

provided \( x^1 \) is \( O(n^{-1}) \). For the univariate case \( p = 1 \), expression (12) reduces to (4).
To verify (11), first note that
\[(r^1)^{-1} = \{x^1(I^{11})^{-1}\}^{-1}[1 + \frac{1}{2}x^1(I^{11})^{-1}F_3 + \frac{1}{2}\{x^1(I^{11})^{-1}\}^2(F_4 + F_3) + O(n^{-3/2})]. \tag{13}\]
Differentiation of \(l[\tilde{x}(x^1)]\) at (6) with respect to \(x^1\) shows that
\[l_1[\tilde{x}(x^1)] = -x^1(I^{11})^{-1}\{1 - \frac{1}{2}x^1(I^{11})^{-1}F_3 - \frac{1}{6}\{x^1(I^{11})^{-1}\}^2F_4 + O(n^{-3/2})\}, \tag{14}\]
and hence
\[\left[(I^{11})^ll_1[\tilde{x}(x^1)]\right]^{-1} = -x^1(I^{11})^{-1}\{1 + \frac{1}{2}x^1(I^{11})^{-1}F_3 + \frac{1}{6}\{x^1(I^{11})^{-1}\}^2(F_4 + \frac{3}{2}F_3) + O(n^{-3/2})\}. \tag{15}\]
It follows from (5) that
\[-l_{\alpha\beta}[\tilde{x}(x^1)] = I_{\alpha\beta} - x^1I_{\alpha\beta}I^{11}(I^{11})^{-1} - \frac{1}{2}\{x^1\}^2\{I_{\alpha\beta}\}I^{11}I^{11}(I^{11})^{-2} + I_{\alpha\beta}I_{\gamma\delta}I^{\gamma\delta}I^{11}(I^{11})^{-2} - I_{\alpha\beta}I_{\gamma\delta}I^{11}I^{11}I^{11}(I^{11})^{-3} + O(\frac{1}{n}). \]
Furthermore, application of a lemma of Barndorff-Nielsen (1986) concerning determinants gives
\[\left(\frac{|(I_{\alpha\beta})|}{|[-l_{\alpha\beta}[\tilde{x}(x^1)]]|}\right)^{\frac{1}{2}} = 1 + x^1(I^{11})^{-1}\{1 - \frac{\frac{1}{4}F_{3,1} + \frac{1}{2}F_{3,2}}{\frac{1}{2}F_{3,1} + \frac{1}{4}F_{3,2}} + \frac{\{x^1\}^2\{I_{\alpha\beta}\}I^{11}I^{11}(I^{11})^{-3} + O(n^{-3/2})}{\frac{3}{2}F_{6,3} + \frac{1}{2}F_{6,4} + \frac{1}{2}F_{6,5} + \frac{1}{2}F_{6,6} + O(n^{-3/2})}. \tag{16}\]
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 \(c\exp\{l(x)\}\) only approximates \(f(x)\) with relative error of order \(O(n^{-3/2})\), that is, if
\[f(x) = c\exp\{l(x)\}[1 + O(n^{-3/2})],\]
where \(c\exp\{l(x)\}\) integrates to 1.

Formula (12) can easily be generalized to provide approximations of marginal tail probabilities in cases where the joint density of the variables is maximized at a point other than 0. Consider now a variable \(Y = (Y_1, \ldots, Y_p)\) such that \(Y - \delta\) is \(O_p(n^{-1})\) for a vector \(\delta = (\delta^1, \ldots, \delta^p)\). Suppose that the probability density function of \(Y\) is
\[g(y) = d\exp\{m(y)\}[1 + O(n^{-3/2})], \quad y = (y^1, \ldots, y^p),\]
where \(m(y)\) is \(O(n)\) 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. As in the previous notation, let
\[m_i(y) = \delta m(y)/\partial y^i, \quad m_{ij}(y) = \delta^2 m(y)/\partial y^i\partial y^j, \quad (i, j = 1, \ldots, p),\]
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 \(y^1 - \delta^1\) is \(O(n^{-1})\),
\[pr (Y^1 \leqslant y^1) = \Phi(r^1) + \phi(r^1)\left(\frac{1}{r^1} + \frac{[\{-m_{ij}(\delta)\}]^1}{m_i(\tilde{y}(y^1))[-m_{\alpha\beta}[\tilde{y}(y^1)]]^1}\right) + O(n^{-3/2}), \tag{16}\]
where \(r^1\) is now defined as
\[r^1 = r^1(y^1) = \text{sgn}(y^1 - \delta^1)(2[m(\delta) - m(\tilde{y}(y^1))]).\]
i and j vary over 1, \ldots, p, and \( \alpha \) and \( \beta \) vary over 2, \ldots, p. For the univariate case \( p = 1 \), in obvious notation, (16) reduces to

\[
\Pr (Y \leq y) = \Phi(r) + \phi(r) \left[ 1 + \frac{\{-m_r^{(2)}(\delta)\}}{m_r^{(1)}(y)} \right] + O(n^{-3/2}),
\]

(17)

where \( r \) is now defined as \( r = r(y) = \text{sgn}(y - \delta)[2(m(\delta) - m(y))]^{\frac{1}{2}} \).

An alternative to approximation (12) can be developed by applying (17) directly to an approximation of the marginal probability density function of \( X^1 \). The first four cumulants of \( Z^1 = X^1(T^{11})^{-\frac{1}{2}} \) are

\[
\begin{align*}
\kappa_1(Z^1) &= \frac{1}{2} F_{3,2} + O(n^{-3/2}), \\
\kappa_2(Z^1) &= 1 + \frac{1}{2} F_{4,2} + \frac{1}{2} F_{6,5} + \frac{1}{2} F_{6,6} + O(n^{-3/2}), \\
\kappa_3(Z^1) &= F_{3,1} + O(n^{-3/2}), \\
\kappa_4(Z^1) &= F_{4,1} + 3 F_{6,3} + O(n^{-3/2}),
\end{align*}
\]

and the higher-order cumulants are \( O(n^{-3/2}) \) or smaller. These cumulants can be used to obtain an Edgeworth expansion for the marginal probability density function \( f_1 \) of \( X^1 \). By using this expansion in conjunction with (6) and (15), it can be shown that

\[
f_1(x^1) = c_1 \left( \frac{[l_{\alpha\beta}(x^1)]}{[-l_{\alpha\beta}(\tilde{x}(x^1))]^{\frac{1}{2}}} \right) \exp \left\{ l(\tilde{x}(x^1)) - l(0) \right\} \{1 + O(n^{-3/2})\},
\]

(18)

where \( c_1 \) is a normalizing constant. Formula (18) is also given by Tierney & Kadane (1986), who derive it by application of Laplace’s approximation for integrals. Arguments given by Tierney & Kadane establish the relative error in (18).

A second approximation to the distribution function of \( X^1 \) is obtained from (18) by applying (17) with

\[
m(x^1) = -\frac{1}{2} \log \left\{ [-l_{\alpha\beta}(\tilde{x}(x^1))] + l(\tilde{x}(x^1)) \right\}.
\]

Thus

\[
\Pr (X^1 \leq x^1) = \Phi(r^*) + \phi(r^*) \left[ 1 + \frac{\{-m_r^{(2)}(\delta)\}}{m_r^{(1)}(x^1)} \right] + O(n^{-3/2}),
\]

(19)

where \( \delta \) is the point at which \( m \) attains its maximum value,

\[
r^* = r^*(x^1) = \text{sgn}(x^1 - \delta)[2(m(\delta) - m(x^1))]^{\frac{1}{2}},
\]

and \( m^{(1)} \) and \( m^{(2)} \) are the first and second derivatives of \( m \).

Note that

\[
\Pr (X^1 \leq x^1) = \Phi(r^*) + O(n^{-\frac{1}{2}}),
\]

(20)

and hence the approximation to the distribution of \( X^1 \) based on the standard normal approximation to the distribution of \( R^* = r^*(X^1) \) has the same order of error as \( (1) \). In practice, however, (20) is likely to perform better than (1). Recall that \( r^1 = 0 \) when \( x^1 = 0 \), so approximation (1) gives \( \Pr (X^1 \leq 0) = 0.5 \). In situations where \( n \) is small and \( p \) is large, the marginal distribution of \( X^1 \) can be centred at a point quite far from 0, and thus (1) performs poorly. Situations where 0 is a fairly extreme quantile of \( X^1 \) are illustrated in Examples 3 and 4 of § 4. On the other hand, \( r^* = 0 \) when \( x^1 = \delta \), so approximation (20) gives \( \Pr (X^1 \leq \delta) = 0.5 \). By definition, \( \delta \) is close to the mode of \( X^1 \), and accordingly (20) can be expected to be more accurate than (1). This behaviour is encountered in the examples of § 4. Since the leading component of (19) is likely to be a better initial approximation than the leading component of (12), it is also reasonable to expect (19) to be more accurate than (12). Note that (19) is more difficult to implement as it requires
calculation of the maximizing point $\delta$ and determination of the first two derivatives of $m$. Of course, in the univariate case $p = 1$, approximations (1) and (20) coincide, as do (12) and (19).

Tierney & 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. D. A. S. Fraser, H. S. Lee and N. Reid, in as yet unpublished work, consider tail area approximations based on direct integration of (18) and thus derive approximations closely related to (19).

It should be stressed that the derivations in this section are nonrigorous; the asymptotic expansions of cumulants given here do not provide strict proofs of the distributional approximations. Moreover, regularity conditions are necessary for strict proofs. In particular, the discussion is restricted to continuous random variates. Discussions of regularity conditions are given by Barndorff-Nielsen & Cox (1979) and Durbin (1980).

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 $Y_{(1)} \leq \ldots \leq Y_{(m)}$ 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

$$A_i = (Y_{(i)} - \hat{\mu}) e^{-\hat{\theta}} \quad (i = 1, \ldots, m).$$  \hspace{1cm} (21)

Moreover, if $l$ is defined by

$$l(x^1, x^2) = \log L(x^1, x^2; A) = \log L(\hat{\mu} + x^1 e^{\hat{\theta}}, \hat{\theta} + x^2; Y) + m\hat{\theta},$$  \hspace{1cm} (22)

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

$$f_{x^1, x^2|A}(x^1, x^2) \propto \exp \{l(x^1, x^2)\}. \hspace{1cm} (23)$$

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 $e_p = F^{-1}(p)$ is the known $p$th quantile of $Z_i$, then $y_p = \mu + \sigma e_p$. 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 $Z_i = Z_i - e_p$ has density $f(z_i + e_p)$. By applying (22) in this context, it can be seen that the joint conditional density of the pivots

$$X^1 = (y_p - \hat{\mu}) e^{-\hat{\theta}} = ((\mu - \hat{\mu}) + (e^\theta - e^{\hat{\theta}}) e_p) e^{-\hat{\theta}}, \quad X^2 = \theta - \hat{\theta}$$

is given by (23), where now

$$l(x^1, x^2) = -\left\{ mx^2 + \sum_{i=1}^{m} g(P_i + e_p) + G(P_m + e_p) \right\}, \hspace{1cm} (24)$$

where $g(x) = \ln(\frac{1}{1 - F(x)})$, $G(x) = \ln(\frac{1}{F(x)})$ and $F(x)$ is the distribution function.
Approximations of marginal tail probabilities

\[ g(z) = -\log f(z), \quad G(z) = -(n - m) \log \{1 - F(z)\}, \]

\[ P_1 = \{A_i - (x^1 + \varepsilon_p)\} e^{-x^2} \quad (i = 1, \ldots, m), \]

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

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

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

where \( g^{(1)} \) and \( G^{(1)} \) are the first derivatives of \( g \) and \( G \), and \( \tilde{P}_i = \{A_i - (x^1 + \varepsilon_p)\} e^{-\tilde{x}^2} \).

From (12), the approximation to the conditional distribution function of \( X^1 = (y_p - \tilde{y}_p) e^{-\tilde{\theta}} \) is

\[ \text{pr} (X^1 \leq x^1 | A) = \Phi(r^1) + \phi(r^1) \left[ \frac{1}{r^1} + \frac{(I_{11}I_{22} - I_{12}^2)^{\frac{1}{2}}}{I_2(x^1, \tilde{x}^2)(-I_{22}(x^1, \tilde{x}^2))^{\frac{1}{2}}} \right], \]

where

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

\[ l_1(x^1, \tilde{x}^2) = e^{-\tilde{x}^2} \left\{ \sum_{i=1}^{m} g^{(1)}(\tilde{P}_i + \varepsilon_p) + G^{(1)}(\tilde{P}_m + \varepsilon_p) \right\}, \]

\[ -l_{22}(x^1, \tilde{x}^2) = m + \sum_{i=1}^{m} g^{(2)}(\tilde{P}_i + \varepsilon_p) \tilde{P}_i^2 + G^{(2)}(\tilde{P}_m + \varepsilon_p) \tilde{P}_m^2, \]

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

\[ I_{22} = m + \sum_{i=1}^{m} g^{(2)}(A_i)A_i^2 + G^{(2)}(A_m)A_m^2, \]

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 signed root of the usual likelihood ratio statistic for \( y_p \).

For approximate conditional inference about the log-scale parameter \( \theta \), it is convenient to consider (24) in the case \( \varepsilon_p = 0 \), so that

\[ P_i = (A_i - x^1) e^{-x^2} \quad (i = 1, \ldots, m). \]

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

\[ \text{pr} (X^2 \leq x^2 | A) = \Phi(r^2) + \phi(r^2) \left[ \frac{1}{r^2} + \frac{(I_{11}I_{22} - I_{12}^2)^{\frac{1}{2}}}{l_2(\tilde{x}^1, x^2)(-I_{22}(\tilde{x}^1, x^2))^{\frac{1}{2}}} \right], \]
where
\[ r^2 = r^2(x^2) = \text{sgn}(x^2) \left( 2 \left[ mx^2 + \sum_{i=1}^{m} \left\{ g(\tilde{P}_i) - g(A_i) \right\} + \left\{ G(\tilde{P}_m) - G(A_m) \right\} \right] \right), \]

\[ l_2(\tilde{x}^2, x^2) = -m + \sum_{i=1}^{m} g^{(1)}(\tilde{P}_i) \tilde{P}_i + G^{(1)}(\tilde{P}_m) \tilde{P}_m, \]

\[ -l_{11}(\tilde{x}^2, x^2) = e^{-2x^2} \left\{ \sum_{i=1}^{m} g^{(2)}(\tilde{P}_i) + G^{(2)}(\tilde{P}_m) \right\}, \]

and \( I_{11}, I_{12} \) 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 \( m = n, \epsilon_p = \Phi^{-1}(p) \), and \( g(z) = \frac{1}{2} \log(2\pi) + \frac{1}{2} z^2 \). For this situation, the conditional and unconditional distributions of \( X^1 = (y_p - \tilde{P}_p) e^{-\tilde{P}_p} \) and \( X^2 = \theta - \tilde{P}_p \) coincide. In particular,
\[ \Pr(X^1 \leq x^1) = \Pr\{t'_{n-1}(n^1\epsilon_p) \leq (n-1)^1(x^1 + \epsilon_p)\}, \]
where \( t'_{n-1}(n^1\epsilon_p) \) has the noncentral \( t \) distribution with \( (n-1) \) degrees of freedom and noncentrality parameter \( n^1\epsilon_p \). Furthermore,
\[ \Pr(X^2 \leq x^2) = \Pr(\chi^2_{n-1} \leq n \ e^{-2x^2}), \]
where \( \chi^2_{n-1} \) has the chi-squared distribution with \( (n-1) \) degrees of freedom.

Solving (25), we have
\[ 2 \ e^{x^2} = -\epsilon_p(x^1 + \epsilon_p) + \{4 + (4 + \epsilon_p^2)(x^1 + \epsilon_p)^2\}^{\frac{1}{2}}, \]
and the components required in (26) are
\[ r^1 = \text{sgn}(x^1) [n\{2\tilde{x}^2 + \epsilon_p^2 - \epsilon_p(x^1 + \epsilon_p) e^{-\tilde{P}_p}\}]^{\frac{1}{2}}, \]
\[ l_2(x^1, \tilde{x}^2) = n e^{-\tilde{x}^2} \{\epsilon_p - (x^1 + \epsilon_p) e^{-\tilde{P}_p}\}, \]
\[ -l_{11}(x^1, \tilde{x}^2) = n[1 + \{1 + (x^1 + \epsilon_p)^2\} e^{-2\tilde{x}^2}], \quad (I_{11}I_{22} - I_{12}^2)^{\frac{1}{4}} = \sqrt{2n}. \]

Table 2 compares the approximations obtained from (1) and (26) with exact results for \( p = 0.001, 0.05, 0.5 \) when \( n = 3 \). In the case \( p = 0.5 \), \( X^1 \) has a symmetric distribution, and only negative values of \( x^1 \) are considered for Table 2.

Equation (27) gives \( \tilde{x}^1 = 0 \), and the components required in (28) are
\[ r^2 = \text{sgn}(x^2) [n\{2x^2 - 1 + e^{-2x^2}\}]^{\frac{1}{2}}, \]
\[ l_2(\tilde{x}^1, x^2) = n\{1 + e^{-2x^2}\}, \quad -l_{11}(\tilde{x}^1, x^2) = n \ e^{-2x^2}. \]

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
\[ \frac{\beta}{\alpha} \left( \frac{t}{\alpha} \right)^{\beta-1} \exp\left\{ -\left( \frac{t}{\alpha} \right)^{\beta} \right\} \quad (t > 0), \]
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 conditional inference for the Weibull distribution is given by Lawless (1978). In this case,
\[ g(z) = -z + e^z, \quad G(z) = (n - m) \ e^z. \]
Approximations of marginal tail probabilities

Table 2. Approximations to \( \Pr (X^1 \leq x^1) \) and \( \Pr (X^2 \leq x^2) \) for Example 1; \( n = 3 \)

(a) \( X^1, \; p = 0.001 \)

<table>
<thead>
<tr>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-200)</td>
<td>(-25)</td>
<td>(-10)</td>
<td>(0.01^*)</td>
<td>(0.75^*)</td>
<td>(1.5^*)</td>
<td>(2.5^*)</td>
</tr>
<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>
</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>
</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>
</tr>
</tbody>
</table>

(b) \( X^1, \; p = 0.05 \)

<table>
<thead>
<tr>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-100)</td>
<td>(-15)</td>
<td>(-5)</td>
<td>(0.01^*)</td>
<td>(0.5^*)</td>
<td>(1.25^*)</td>
<td>(3^*)</td>
</tr>
<tr>
<td>(1)</td>
<td>0.0242</td>
<td>0.0745</td>
<td>1.398</td>
<td>49.548</td>
<td>25.277</td>
<td>4.878</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>
</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>
</tr>
</tbody>
</table>

(c) \( X^1, \; p = 0.5 \)

<table>
<thead>
<tr>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-30)</td>
<td>(-5)</td>
<td>(-3)</td>
<td>(-2)</td>
<td>(-1.5)</td>
<td>(-0.5)</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>(1)</td>
<td>0.0312</td>
<td>0.0885</td>
<td>0.429</td>
<td>1.400</td>
<td>3.003</td>
<td>20.663</td>
</tr>
<tr>
<td>(26)</td>
<td>0.0556</td>
<td>0.896</td>
<td>2.377</td>
<td>4.926</td>
<td>7.899</td>
<td>26.976</td>
</tr>
<tr>
<td>Exact</td>
<td>0.078</td>
<td>0.971</td>
<td>2.566</td>
<td>5.279</td>
<td>8.397</td>
<td>27.639</td>
</tr>
</tbody>
</table>

(d) \( X^2 \)

<table>
<thead>
<tr>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-0.9)</td>
<td>(-0.5)</td>
<td>(-0.25)</td>
<td>(0.01)</td>
<td>(1.25^*)</td>
<td>(2.50^*)</td>
<td>(5^*)</td>
</tr>
<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>
</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>
</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>0.368</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>
</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>
</tr>
</tbody>
</table>

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

Table 3. Approximations to \( \Pr (X^1 \leq x^1) \) and \( \Pr (X^2 \leq x^2) \) for Example 2; \( n = 5, m = 3 \)

(a) \( X^1, \; p = 0.01 \)

<table>
<thead>
<tr>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
<th>( x^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-289)</td>
<td>(-59)</td>
<td>(-15)</td>
<td>(-1^*)</td>
<td>(1^*)</td>
<td>(2.5^*)</td>
</tr>
<tr>
<td>(1)</td>
<td>0.0249</td>
<td>0.0300</td>
<td>1.190</td>
<td>64.993</td>
<td>30.235</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>
</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>
</tr>
</tbody>
</table>

(b) \( X^2 \)

<table>
<thead>
<tr>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
<th>( x^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-1.1)</td>
<td>(-0.7)</td>
<td>(-0.3)</td>
<td>(0.3)</td>
<td>(1.5^*)</td>
<td>(2.7^*)</td>
</tr>
<tr>
<td>(1)</td>
<td>0.415</td>
<td>5.887</td>
<td>26.650</td>
<td>71.122</td>
<td>1.212</td>
</tr>
<tr>
<td>(28)</td>
<td>0.067</td>
<td>1.296</td>
<td>8.926</td>
<td>39.575</td>
<td>10.575</td>
</tr>
<tr>
<td>Exact</td>
<td>0.056</td>
<td>1.258</td>
<td>8.701</td>
<td>39.082</td>
<td>10.711</td>
</tr>
</tbody>
</table>

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

The solution \( \hat{x}^2 \) of (25) 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

\[
e^{\hat{x}^1} = \left[ \sum_{i=1}^{m} \exp (A_i e^{-x^2}) + (n - m) \exp (A_m e^{-x^2}) \right] / m \]

For \( n = 5, m = 3 \), 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 results obtained from (1) and (26) for $X^1$ when $p = 0.01$ and with those from (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. Linear 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 $\mu_i = \Sigma u_{is} \beta^s$ and $Z_1, \ldots, Z_n$ are independent and identically distributed with known probability density function $f$. The vectors $u_i = (u_{i1}, \ldots, u_{in})^T \; (i = 1, \ldots, n)$ of covariate values are assumed to be known, and the vector $\beta = (\beta^1, \ldots, \beta^q)$ 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, Appen. G) and Verhagen (1961), among others, that inference about $\beta$ and $\theta$ should be based on the joint conditional distribution of the maximum likelihood estimators $\hat{\beta} = (\hat{\beta}^1, \ldots, \hat{\beta}^q)$ and $\hat{\theta}$ given the sample configuration $A = (A_1, \ldots, A_n)$, where $A_i = (Y_i - \beta u_i) e^{-\theta}$. If $l$ is defined by

$$l(x^1, \ldots, x^q, x^{q+1}) = \log L(x^1, \ldots, x^q, x^{q+1}; A)$$

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

then the joint conditional density of the pivots

$$X^1 = (\beta^1 - \hat{\beta}^1) e^{-\hat{\theta}}, \ldots, X^q = (\beta^q - \hat{\beta}^q) e^{-\hat{\theta}}, X^{q+1} = \theta - \hat{\theta}$$

satisfies

$$f_{X^1, \ldots, X^{q+1}}(x) \propto \exp \{ l(x) \}, \quad x = (x^1, \ldots, x^{q+1}).$$

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

$$l_s(x) = e^{-x^{q+1}} \sum_{i=1}^{n} u_{is} g^{(1)}(P_i), \quad l_{q+1}(x) = -n + \sum_{i=1}^{n} g^{(1)}(P_i) P_i,$$

$$-l_{s,t}(x) = e^{-2x^{q+1}} \sum_{i=1}^{n} u_{is} u_{it} g^{(2)}(P_i),$$

$$-l_{s,s+1}(x) = e^{-x^{q+1}} \sum_{i=1}^{n} u_{is} \{ g^{(1)}(P_i) + g^{(2)}(P_i) P_i \},$$

$$-l_{q+1,q+1}(x) = \sum_{i=1}^{n} \{ g^{(1)}(P_i) P_i + g^{(2)}(P_i) P_i^2 \} \quad (s, t = 1, \ldots, q),$$

where $g(z) = -\log f(z)$, $g^{(1)}$ and $g^{(2)}$ are the first and second derivatives of $g$, and

$$P_i = \{ A_i - (u_{i1} x^1 + \ldots + u_{iq} x^q) \} \exp (-x^{q+1}) \quad (i = 1, \ldots, n).$$
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 $s = 1, \ldots, q$ and for $\theta$ when $s = q + 1$. Just as in the location-scale case, approximate inference about the $p$th quantile of an observation having 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 for any specific data set it is difficult to obtain exact conditional results by integrating (30), a simulation of size 1000 was performed. This investigation used $n = 10$, $q = 6$, the matrix of covariate values

$$
U^T = (u_{is})^T
$$

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 $(1 - \alpha)$ conditional confidence limits for $\beta^1$, $\beta^3$ and $\theta$ constructed from (1) and (12) actually covered the true parameter values. Table 4 reports the overall rates of noncoverage of the one-sided confidence intervals observed in the 1000 simulations. Thus, the approximations are assessed in terms of unconditional rather than conditional coverage.

### Table 4. Simulated rates of noncoverage for $\beta^1$, $\beta^3$ and $\theta$ in Example 3;

$n = 10$, $q = 6$, 1000 trials

(a) Upper confidence limits

<table>
<thead>
<tr>
<th>$100\alpha$</th>
<th>$\beta^1$</th>
<th>$\beta^3$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (12)</td>
<td>(1) (12)</td>
<td>(1) (12)</td>
</tr>
<tr>
<td>0.5</td>
<td>4.9</td>
<td>5.8</td>
<td>36.0</td>
</tr>
<tr>
<td>1.0</td>
<td>6.7</td>
<td>7.4</td>
<td>42.7</td>
</tr>
<tr>
<td>2.5</td>
<td>10.9</td>
<td>10.3</td>
<td>55.3</td>
</tr>
<tr>
<td>5.0</td>
<td>14.1</td>
<td>14.2</td>
<td>65.2</td>
</tr>
<tr>
<td>10.0</td>
<td>19.2</td>
<td>18.6</td>
<td>73.8</td>
</tr>
<tr>
<td>25.0</td>
<td>32.0</td>
<td>30.2</td>
<td>88.3</td>
</tr>
</tbody>
</table>

(b) Lower confidence limits

<table>
<thead>
<tr>
<th>$100\alpha$</th>
<th>$\beta^1$</th>
<th>$\beta^3$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (12)</td>
<td>(1) (12)</td>
<td>(1) (12)</td>
</tr>
<tr>
<td>0.5</td>
<td>7.7</td>
<td>7.6</td>
<td>0.0</td>
</tr>
<tr>
<td>1.0</td>
<td>9.4</td>
<td>10.2</td>
<td>0.0</td>
</tr>
<tr>
<td>2.5</td>
<td>12.8</td>
<td>15.2</td>
<td>0.0</td>
</tr>
<tr>
<td>5.0</td>
<td>17.7</td>
<td>19.8</td>
<td>0.0</td>
</tr>
<tr>
<td>10.0</td>
<td>23.7</td>
<td>26.4</td>
<td>0.0</td>
</tr>
<tr>
<td>25.0</td>
<td>33.5</td>
<td>37.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table entries are percentages.
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 $\theta$. The discussion following (20) suggests that use of approximation (19) would yield better results for $\theta$; however, 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 $X^1, \ldots, X^{q+1}$ coincide; in particular,

$$\text{pr} \left( X^{q+1} \leq x^{q+1} \right) = \text{pr} \left( \chi^2_{(n-q)} \geq n e^{-2x^{q+1}} \right).$$

Note that the distribution of $X^{q+1}$ does not depend on the covariate values. Approximation (12) gives

$$\text{pr} \left( X^{q+1} \leq x^{q+1} \right) = \Phi(r) + \phi(r) \frac{1}{r} \left( \frac{2}{n} \right)^{\frac{1}{2}} \frac{e^{qx^{q+1}}}{(e^{-2x^{q+1}} - 1)},$$

where

$$r = \text{sgn} \left( x^{q+1} \right) \left[ n(2x^{q+1} - 1 + e^{-2x^{q+1}}) \right]^{\frac{1}{2}}.$$

To implement (19), it suffices to take

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

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

$$\text{pr} \left( X^{q+1} \leq x^{q+1} \right) = \Phi(r^*) + \phi(r^*) \left[ \frac{1}{r^*} \frac{(2(n-q))^{\frac{1}{2}}}{n(e^{-2x^{q+1}} - e^{-2\delta})} \right],$$

where

$$r^* = \text{sgn} \left( x^{q+1} - \delta \right) \left[ 2(n-q)(x^{q+1} - \delta) + n(e^{-2x^{q+1}} - e^{-2\delta}) \right]^{\frac{1}{2}}.$$

Table 5 compares the use of (1), (20), (31) and (32) in the case $n = 10, q = 6$ which was considered for Example 3. In particular, Table 5 shows the true probabilities with which approximate upper and lower $(1 - \alpha)$ confidence limits fail to cover $\theta$. Approximation

<table>
<thead>
<tr>
<th>$100\alpha$</th>
<th>25</th>
<th>10</th>
<th>5</th>
<th>2.5</th>
<th>1</th>
<th>0.5</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<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>
<td>7.324</td>
</tr>
<tr>
<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>
<td>0.0141</td>
</tr>
<tr>
<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>
<td>0.0236</td>
</tr>
<tr>
<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>
<td>0.0487</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$100\alpha$</th>
<th>25</th>
<th>10</th>
<th>5</th>
<th>2.5</th>
<th>1</th>
<th>0.5</th>
<th>0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.978</td>
<td>0.204</td>
<td>0.0699</td>
<td>0.0253</td>
<td>0.00700</td>
<td>0.00274</td>
<td>0.00819</td>
</tr>
<tr>
<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>
<td>0.0225</td>
</tr>
<tr>
<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>
<td>35.506</td>
</tr>
<tr>
<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>
<td>0.0496</td>
</tr>
</tbody>
</table>

Table entries are percentages.
(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 whenever \( x^2 \) is less than the 35.501 percentile of \( X^2 \). 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 \( q = 1 \) 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 \( X \). This method avoids the moment calculations referred to in § 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 & Rice (1980) are explored.

Consider for the moment the univariate case \( p = 1 \). As in the notation of §§ 1 and 2, suppose the density of \( X = X^1 \) is

\[
 f(x) = c \exp \{ l(x) \} \{ 1 + O(n^{-3/2}) \},
\]

where \( l(x) \) is maximized at \( x = 0 \). If \( I = -\bar{l}^{(2)}(0) \), then the density of \( Y = X I^1 \) is

\[
 g(y) = \bar{c} \exp \{ \bar{l}(y) \} \{ 1 + O(n^{-3/2}) \},
\]

where \( \bar{l}(y) = l(y I^{-1}) \) and \( \bar{c} \) is a normalizing constant. The necessary observations concerning the variable \( Y \) for subsequent calculations are that \( \bar{l}^{(1)}(0) = 0 \) and \( -\bar{l}^{(2)}(0) = 1 \); in addition

\[
 \bar{l}^{(3)}(0) = l^{(3)}(0) I^{-3/2} = F_3, \quad \bar{l}^{(4)}(0) = l^{(4)}(0) I^{-2} = F_4.
\]

It follows that

\[
 r = \text{sgn}(y)[2\{\bar{l}(0) - \bar{l}(y)\}]^{1/2} = \text{sgn}(x)[2\{l(0) - l(x)\}]^{1/2}
\]

has the expansion

\[
 r = y - \frac{1}{6} y^2 F_3 - \frac{1}{24} y^3 (F_4 + \frac{1}{2} F_3^2) + O(n^{-3/2}),
\]

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

\[
 y = r + \frac{1}{6} r^2 F_3 + \frac{1}{24} r^3 (F_4 + \frac{1}{2} F_3^2) + O(n^{-3/2});
\]

hence,

\[
 dy/dr = \exp \{ \frac{1}{3} r F_3 + \frac{1}{24} r^2 (3 F_4 + \frac{1}{2} F_3^2) + O(n^{-3/2}) \}.
\]

Use of (33) and (35) in conjunction with the expansion

\[
 g(y) = \bar{c} \exp \{ \bar{l}(0) - \frac{1}{2} y^2 + \frac{1}{6} y^3 F_3 + \frac{1}{24} y^4 F_4 + O(n^{-3/2}) \},
\]

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

\[
 \exp \{ -\frac{1}{2} r^2 + \frac{1}{6} r F_3 + \frac{1}{24} r^2 (3 F_4 + \frac{1}{2} F_3^2) + O(n^{-3/2}) \},
\]

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

\[
 h = 1 + \frac{1}{2}(F_4 + \frac{2}{3} F_3^2) + O(n^{-3/2}).
\]

If \( \zeta \) is defined by \( \zeta = -\bar{l}^{(1)}(y) \), then \( \zeta = -I^{-1} l^{(1)}(x) \); moreover,

\[
 \zeta = \frac{1}{2} \frac{d}{dy} \left( \frac{r^2}{dy} \right) = r \left( \frac{dr}{dy} \right).
\]
On combining (35), (36) and (37), we have that the probability density function of $R$ is

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

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

$$\text{pr}(X \leq x) = \text{pr}(Y \leq y) = h\int_{-\infty}^{r} \phi(r)\frac{r}{\zeta} dr,$$

and integration by parts yields

$$\text{pr}(X \leq x) = h\left\{ -\frac{\phi(r)}{\zeta} + \int_{-\infty}^{r} \phi(r)\left(\frac{d\xi^{-1}}{dr}\right) dr \right\}.$$  

It follows from (34) and (37) that

$$\frac{d\xi^{-1}}{dr} = -r^{-2}\left\{ \frac{dy}{dr} - \frac{r^2 y}{(dr)^2} \right\} = -r^{-2}\{1 - \frac{1}{8}r^2(F_4 + \frac{5}{3}F_3)\} = -(1 + r^{-2}) + h^{-1}.$$  

Hence,

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

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

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 §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

$$\frac{dy}{dx^1} = \left\{ \frac{|L_{aB}|}{|[-L_{aB}(\tilde{x}(x^1))]|} \frac{1}{|L^{11}|} \right\}^{\frac{1}{2}} = \frac{|L|^{\frac{1}{2}}}{|[-L_{aB}(\tilde{x}(x^1))]|^{\frac{1}{2}}}.$$  

and $y(0) = 0$. Then the density of $Y$ is proportional to $\exp\{\tilde{L}(y)\{1 + O(n^{-3/2})\}\}$, where $\tilde{L}(y) = \tilde{L}(y(x^1)) = l(\tilde{x}(x^1))$, provided $y$ is $O(1)$. It can be shown from (5) that $\tilde{L}^{(1)}(0) = 0$ and $-\tilde{L}^{(2)}(0) = 1$; therefore, the conditions on the variable $Y$ required in the previous derivation of (38) are met. For this case,

$$r = \text{sgn}(y)[2\{\tilde{L}(0) - \tilde{L}(y)\}]^{\frac{1}{2}} = \text{sgn}(x^1)(2[|L(0)| - l(\tilde{x}(x^1))])^{\frac{1}{2}},$$

$$\zeta = -\tilde{L}^{(1)}(y) = l_i(\tilde{x}(x^1))\frac{dx^1}{dy} = l_i(\tilde{x}(x^1))\frac{|L|^{-\frac{1}{2}}}{|[-L_{aB}(\tilde{x}(x^1))]|^{\frac{1}{2}}}.$$  

Substitution of these expressions into (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 approximation for tail probabilities 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 Lugannani–Rice approximation in the context of an exponential family.

The Lugannani–Rice approximation is (Daniels, 1987, form. (4.9))

\[
\Pr(\bar{X} \geq x) = 1 - \Phi(r) + \phi(r)(1/\hat{z} - 1/r) + O(n^{-3/2}),
\]

where

\[
\hat{z} = \hat{\theta}(nK^{(2)}(\hat{\theta})), \quad r = \text{sgn}(\hat{\theta})(2n[\hat{\theta}x - K(\hat{\theta})])^{1/2},
\]

\( K(t) \) is the cumulant generating function of the underlying observations, and \( K^{(1)}(\hat{\theta}) = x \). The point \( \hat{\theta} \) 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

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

The derivation then proceeds by transforming the exponent to a quadratic and taking two terms in a Taylor expansion of the resulting Jacobian. Approximation (39) can be extended to the case of an exponential family. In this situation,

\[
f_\theta(x) = \exp\{\theta u(x) - K(\theta)\} + o(1),
\]

and \( K(z) - zx_0 \) is replaced by

\[
K_\theta(z) = K(\theta + z) - K(\theta) - zK^{(1)}(\hat{\theta}),
\]

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 (39) with \( \hat{z} = (\hat{\theta}_0 - \theta)(K^{(2)}(\hat{\theta}_0))^{1/2} \) and

\[
r = \text{sgn}(\hat{\theta}_0 - \theta)[2n\{K(\theta) - K(\hat{\theta}_0) + (\hat{\theta}_0 - \theta)K^{(1)}(\hat{\theta}_0)\}]^{1/2}.
\]

where \( l \) is the log likelihood function based on the sample of size \( n \).

This approximation is discussed by Fraser (1990, form. (1.5)–(1.7)) and is given in a more general form by Daniels (1983, form. (6.1)).

To compare the Lugannani–Rice approximation and (4) when applied to approximating \( \Pr(X \geq x_0) \) in a general context, it is helpful to view both approximations as being

Table 6. Approximations to \( \Pr(X \leq x) \) for extreme value distribution

<table>
<thead>
<tr>
<th>x</th>
<th>(4)</th>
<th>LR</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>-9</td>
<td>0.0132</td>
<td>0.0124</td>
<td>0.0123</td>
</tr>
<tr>
<td>-5</td>
<td>0.707</td>
<td>0.799</td>
<td>0.672</td>
</tr>
<tr>
<td>-1</td>
<td>31.043</td>
<td>28.111</td>
<td>30.780</td>
</tr>
<tr>
<td>2</td>
<td>0.0628</td>
<td>0.0776</td>
<td>0.0618</td>
</tr>
<tr>
<td>2.5</td>
<td>0.0526</td>
<td>0.0245</td>
<td>0.0512</td>
</tr>
</tbody>
</table>

Table entries are percentages. For case marked *, reported results pertain to \( \Pr(X \geq x) \).
obtained by incorporating an exponential family into the problem, and using the exponential family approximation given above. For the Lugannani–Rice approximation, the appropriate exponential family is given by

\[ c(x) \exp \{ T(x_0)(x - x_0) + \log f(x) \}, \]

where \( T(x_0) \) is the saddlepoint. Now, using the exponential family result to approximate \( \Pr \{ \hat{T} \geq T(x_0) \} \), the Lugannani–Rice 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 (4), Fraser (1990) shows that the appropriate exponential family is proportional to \( \exp \{ \lambda(\phi)s + l(x_0 - \phi) \} \), where

\[ l(x) = \log f(x), \quad \lambda(\phi) = l^{(1)}(x - \phi_0)/\{ -l^{(2)}(0) \}, \quad s = l^{(1)}(x - \phi). \]

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 \( \Pr \{ X \geq x_0 \} \). See also Fraser (1988) and Fraser & Reid (1988).

Thus, both approximations can be viewed as incorporating an exponential family into the problem and then using the Lugannani–Rice 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 Lugannani–Rice 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 Lugannani–Rice 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 \(-l^{(1)}(x)\). To see that this interpretation is plausible, consider the saddlepoint equation \( K^{(1)}(T) = x_0 \), which can be written as

\[ \int (x - x_0) \exp \{ T(x - x_0) + l(x) \} \, dx = 0. \]

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

\[ \int (x - x_0) \exp \{ -l^{(1)}(x_0)(x - x_0) + l(x) \} \, dx = O(n^{-3/2}). \]

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

Acknowledgement

Helpful comments of the referees are gratefully acknowledged.

References


Approximations of marginal tail probabilities


*Received March 1989. Revised August 1989*