(2)

ANALYTICAL APPROXIMATIONS TO CONDITIONAL DISTRIBUTION FUNCTIONS

BY
THOMAS J. DICICCIO
MICHAEL A. MARTIN
G. ALASTAIR YOUNG

TECHNICAL REPORT NO. 452
APRIL 20, 1992

PREPARED UNDER CONTRACT
N00014-92-J-1254 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ANALYTICAL APPROXIMATIONS TO CONDITIONAL DISTRIBUTION FUNCTIONS

BY

THOMAS J. DICICCIO

MICHAEL A. MARTIN

G. ALASTAIR YOUNG

TECHNICAL REPORT NO. 452

APRIL 20, 1992

Prepared under contract

N00014-92-J-1254 (NR-042-267)

For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Analytical approximations to conditional distribution functions

BY THOMAS J. DICICCIO, MICHAEL A. MARTIN
AND G. ALASTAIR YOUNG

SUMMARY. Conditional inference plays a central role in statistics, but determination of relevant conditional distributions is often difficult. We develop analytical procedures that are accurate and easy to apply for approximating conditional distribution functions. For a continuous random vector $X = (X^1, \ldots, X^p)$, we estimate conditional tail probabilities

$$\text{pr}(Y^1 \leq a^1 | Y^2 = a^2, \ldots, Y^k = a^k), \qquad k \leq p,$$

where $Y^i = g^i(X^1, \ldots X^p)$ $(i = 1, \ldots, k)$ and $g^1, \ldots, g^k$ are smooth functions of X. Previous approaches have dealt with the cases where the variable whose conditional distribution is sought is a linear function of means, and where there are $p - 1$ conditioning variables. However, in many practical circumstances the statistic of interest is a nonlinear function of means and it is advantageous to condition on a lower-dimensional ancillary statistic. Our procedure first involves approximating the marginal density function $f_{Y^1,\ldots,Y^k}(y^1, \ldots, y^k)$ for $Y^1, \ldots, Y^k$, by an approach of Phillips (1983) and Tierney, Kass and Kadane (1989). An accurate approximation to the required conditional probability is then obtained by applying a marginal tail probability approximation of DiCiccio and Martin (1991) to the conditional density of Y^1 given $Y^2, \ldots, Y^k$ which satisfies

$$f_{Y^1|Y^2,\ldots,Y^k}(y^1 | Y^2 = a^2, \ldots, Y^k = a^k) \propto f_{Y^1,\ldots,Y^k}(y^1, a^2, \ldots, a^k).$$

Our method is illustrated in several examples, including one which uses a saddlepoint approximation for the density of X, and the method is applied for conditional bootstrap inference.

Some key words: Ancillary statistic; Conditional bootstrap; Laplace's method; Marginal density; Saddlepoint approximation; Tail probability approximation.

1. Introduction

Conditional distributions play a key role in many inference problems, largely through the use of the conditionality principle and notions such as ancillarity. Unfortunately, it is often difficult or impossible to compute exact conditional distributions, and standard approximation methods often fail to work or are difficult to adapt to the situation at hand. For example, Edgeworth expansions can yield negative probability estimates in the tails of a distribution, and saddlepoint methods based on cumulant generating functions are only easily applied when the variables of interest are means.

Skovgaard (1987) investigated the use of saddlepoint methods in the case of a bivariate mean to develop analytical approximations to the conditional distribution of one mean given the other. He extended his method to the case of p means, approximating the conditional distribution of a linear function of the means given a $(p-1)$-dimensional linear function of them. Wang (1991) extended Skovgaard's results further to include the case of approximating the conditional distribution of a mean given $p-1$ nonlinear functions of the means. Davison and Hinkley (1988) applied Skovgaard's approach in a conditional bootstrap context. They extended Skovgaard's results to include the conditional distribution of certain functions of the means given a $(p-1)$-dimensional linear function of them, where the functions in question lead to statistics which are solutions of linear estimating equations. The techniques of Skovgaard and Wang have several elements in common that may limit their applicability. First, because they are based on saddlepoint approximations, their methods require knowledge of the cumulant generating function of the entire random vector of interest. Second, their technique restricts the variable whose conditional distribution is sought to be a linear function of means, or at least to be a function of means identified with a linear estimating equation. This restriction can be quite severe in practice. Finally, and most importantly, the number of conditioning variables is necessarily $p-1$. However, in many cases of practical interest, an ancillary exists that is of lower dimension than $p-1$, and interest centers on distributions conditional on that ancillary.

In this paper, we develop an analytical approximation to conditional tail probabilities

for a smooth function of a random vector $X = (X^1, \ldots, X^p)$ given $k-1$ other smooth functions of X, where $k \leq p$. The vector X is not restricted to a vector of means; we assume only that its joint density function is of the form $cb(x)\exp\{\ell(x)\}$. Also, the variable whose conditional distribution is sought may be a smooth, non-linear function of X, giving our method considerable generality. Moreover, our method allows the dimension of the conditioning variable to be less than $p-1$, so that a lower-dimensional ancillary statistic may be conditioned on if it exists. Our technique produces accurate approximate conditional tail probabilities, and is based on applying DiCiccio and Martin's (1991) tail probability approximation to a marginal density approximation given by Phillips (1983) and Tierney, Kass and Kadane (1989), by noting that the required conditional density is proportional to the marginal density for fixed values of the conditioning variables. A secondary, theoretical contribution of the paper is to show that the approaches of Phillips (1983) and Tierney, Kass and Kadane (1989) for developing marginal density approximations are equivalent.

An important feature of our approximation is that it avoids costly numerical integration. An obvious alternative approach to use of our method is numerical integration of a renormalized version of the conditional density approximation that arises in developing our technique. The first obstacle to implementing this approach is that renormalization requires the computation of a second numerical integral. However, both numerical integration steps are practically infeasible because each density function evaluation requires a potentially costly constrained maximization step. In contrast, application of our method requires only four function evaluations.

Section 2 of the paper describes our theoretical results. In section 3, we discuss computational aspects of our method and provide a formal algorithm for its use. Examples of the use of our method are given in Section 4, including an application to the conditional bootstrap.

2. Conditional Tail Probability Approximation

Consider a continuous random vector $X = (X^1, \ldots, X^p)$ having probability density

function of the form

$$f_X(x) = cb(x)\exp\{\ell(x)\}, \qquad x = (x^1, \ldots, x^p),$$

and let $\hat{x} = (\hat{x}^1, \ldots, \hat{x}^p)$ be the point maximizing $\ell(x)$, and suppose that $X - \hat{x}$ is $O_p(n^{-\frac{1}{2}})$ as $n \to \infty$, where n is sample size. For each fixed x, assume that $\ell(x)$ and its partial derivatives are $O(n)$. We are interested in approximating conditional tail probabilities

$$\mathrm{pr}(Y^1 \leq a^1 \mid Y^2 = a^2, \ldots, Y^k = a^k), \qquad k \leq p,$$

where $a^2, \ldots, a^k$ are fixed constants and $Y^i = g^i(X^1, \ldots, X^p)$ $(i = 1, \ldots, k)$ for functions $g^1, \ldots, g^k$ which are assumed to have continuous gradients that do not vanish in an $n^{-\frac{1}{2}}$-neighbourhood of $\hat{x}$.

In order to study the conditional distribution of Y^1 given $Y^2, \ldots, Y^k$, we first consider an approximation to the marginal density of $Y^1, \ldots, Y^k$. Two approaches to estimating this marginal density are given by Phillips (1983) and Tierney, Kass and Kadane (1989). Both approaches utilize Laplace's method of approximating integrals to avoid the need for high-dimensional integration, and it is shown here that they yield the same marginal density approximation; see the Appendix. We will use elements of both approaches to describe our method, so we now briefly describe each approach.

Phillips (1983) assumes a 1–1 transformation

$$Y = (Y^1, \ldots, Y^p) = \{g^1(X^1, \ldots, X^p), \ldots, g^p(X^1, \ldots, X^p)\}$$

of X, where the variables of interest are $Y^1, \ldots, Y^k$, and the functions $g^{k+1}, \ldots, g^p$ are smooth and have non-zero gradients in an $n^{-\frac{1}{2}}$-neighbourhood of $\hat{x}$. Let $J\{x(y)\}$ denote the Jacobian of this transformation. Then the probability density function of Y is of the form

$$f_Y(y) = c\bar{b}(y)\exp\{\bar{\ell}(y)\}, \qquad y = (y^1, \ldots, y^p),$$

where $\bar{b}(y) = b\{x(y)\}/\det[J\{x(y)\}]$ and $\bar{\ell}(y) = \ell\{x(y)\}$. The marginal density of $Y^1, \ldots, Y^k$ is then

$$f_{Y^1,\ldots,Y^k}(y^1, \ldots, y^k) = c\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \bar{b}(y)\exp\{\bar{\ell}(y)\}\, dy^{k+1} \ldots dy^p$$

$$= \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \bar{b}(y) \exp\{\bar{\ell}(y)\}\, dy^{k+1} \ldots dy^p}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \bar{b}(y) \exp\{\bar{\ell}(y)\}\, dy^1 \ldots dy^p}. \qquad (1)$$

Let $\hat{y}$ be the value of y maximizing $\bar{\ell}(y)$, and let $\tilde{y} = \tilde{y}(y^1, \ldots, y^k)$ be the value of y maximizing $\bar{\ell}(y)$ subject to the first k components of y being held fixed at the values $y^1, \ldots, y^k$. Let $\bar{\ell}_i(y) = \partial \bar{\ell}(y)/\partial y^i$, $\bar{\ell}_{ij}(y) = \partial^2 \bar{\ell}(y)/\partial y^i \partial y^j$ $(i, j = 1, \ldots, p)$. Applying Laplace's method to the numerator and denominator of (1), an approximation to the marginal density of $Y^1, \ldots, Y^k$ is

$$f_{Y^1,\ldots,Y^k}(y^1, \ldots, y^k) \approx (2\pi)^{-k/2} \left[\frac{\det\{\Omega(\hat{y})\}}{\det\{\Omega'(\tilde{y})\}}\right]^{\frac{1}{2}} \frac{\bar{b}(\tilde{y})}{\bar{b}(\hat{y})} \exp\{\bar{\ell}(\tilde{y}) - \bar{\ell}(\hat{y})\}, \qquad (2)$$

where $\Omega(y)$ is the $p \times p$ matrix whose (i,j)th element is $-\bar{\ell}_{ij}(y)$, and $\Omega'(y)$ is the $(p-k) \times (p-k)$ submatrix of $\Omega(y)$ corresponding to pairs (i,j) where $i, j = k+1, \ldots, p$. An apparent problem with approximation (2) is that it requires specification of $p-k$ new functions $g^{k+1}, \ldots, g^p$ of X, the choice of which may affect the accuracy of (2).

Tierney, Kass and Kadane (1989) provide a formula for the approximate marginal density of $Y^1, \ldots, Y^k$ that does not require specification of $g^{k+1}, \ldots, g^p$. To describe their formula, we first need additional notation. Let $\tilde{x}$ be the value of x maximizing $\ell(x)$ subject to the constraints $g^1(x^1, \ldots, x^p) = y^1, \cdots, g^k(x^1, \ldots, x^p) = y^k$, and let $H(x)$ be the Lagrangian for this constrained maximization,

$$H(x) = \ell(x) + \lambda_\alpha \{g^\alpha(x) - y^\alpha\},$$

where $\lambda_\alpha = \lambda_\alpha(y^1, \ldots, y^k)$ $(\alpha = 1, \ldots, k)$ and the usual summation convention applies whereby summation is assumed over indices appearing as both a subscript and a superscript. Let $\ell_i(x) = \partial \ell(x)/\partial x^i$, $\ell_{ij}(x) = \partial^2 \ell(x)/\partial x^i \partial x^j$, $H_{ij}(x) = \partial^2 H(x)/\partial x^i \partial x^j$, $g^\alpha_i(x) = \partial g^\alpha(x)/\partial x^i$, $g^\alpha_{ij}(x) = \partial^2 g^\alpha(x)/\partial x^i \partial x^j$ $(i, j = 1, \ldots, p;\ \alpha = 1, \ldots, k)$ denote the partial derivatives of ℓ, H and g^α, respectively. Define the $p \times p$ matrices $\Lambda(x) = \{-\ell^{ij}(x)\}$, the inverse of the matrix whose (i,j)th element is $-\ell_{ij}(x)$, and $\bar{\Lambda}(x) = \{-H^{ij}(x)\}$, the inverse of the matrix whose (i,j)th element is $-H_{ij}(x)$ $(i, j = 1, \ldots, p)$, and the $k \times k$ matrix $\Theta(x)$ whose (α, β)th element is $-H^{ij}(x) g^\alpha_i(x) g^\beta_j(x)$ $(\alpha, \beta = 1, \ldots, k)$. Then, Tierney,

Kass and Kadane's approximation to the marginal density of $Y^1, \ldots, Y^k$ is

$$f_{Y^1,\ldots,Y^k}(y^1,\ldots,y^k) \approx (2\pi)^{-k/2}\left[\frac{\det\{\bar{\Lambda}(\tilde{x})\}}{\det\{\Lambda(\hat{x})\}\det\{\Theta(\tilde{x})\}}\right]^{\frac{1}{2}}\frac{b(\tilde{x})}{b(\hat{x})}\exp\{\ell(\tilde{x})-\ell(\hat{x})\}. \quad (3)$$

Proposition 1. Approximations (2) and (3) to the marginal density of $Y^1, \ldots, Y^k$ are equivalent.

Proposition 1 is proved in the Appendix. In particular, it is shown there that

$$\left[\frac{\det\{\Omega(\hat{y})\}}{\det\{\Omega'(\tilde{y})\}}\right]^{\frac{1}{2}}\frac{\bar{b}(\tilde{y})}{\bar{b}(\hat{y})} = \left[\frac{\det\{\bar{\Lambda}(\tilde{x})\}}{\det\{\Lambda(\hat{x})\}\det\{\Theta(\tilde{x})\}}\right]^{\frac{1}{2}}\frac{b(\tilde{x})}{b(\hat{x})}, \quad (4)$$

and

$$\bar{\ell}(\tilde{y}) - \bar{\ell}(\hat{y}) = \ell(\tilde{x}) - \ell(\hat{x}). \quad (5)$$

Our rationale in what follows is that

$$f_{Y^1|Y^2,\ldots,Y^k}(y^1 \mid Y^2 = a^2,\ldots,Y^k = a^k) \propto f_{Y^1,\ldots,Y^k}(y^1, a^2,\ldots,a^k),$$

so that we may write

$$f_{Y^1|Y^2,\ldots,Y^k}(y^1 \mid Y^2 = a^2,\ldots,Y^k = a^k) \propto b^*(y^1)\exp\{\ell^*(y^1)\}, \quad (6)$$

for suitably defined functions ℓ^* and b^*. We apply the DiCiccio and Martin (1991) tail probability formula to obtain approximations to conditional tail probabilities $\mathrm{pr}(Y^1 \leq a^1 \mid Y^2 = a^2,\ldots,Y^k = a^k)$. Fix the values of $y^2,\ldots,y^k$ in the preceding discussion at their conditioned values $a^2,\ldots,a^k$, respectively. Then $\tilde{y} = \tilde{y}(y^1, a^2,\ldots,a^k)$ is a function of y^1 alone, and is the value of y maximizing $\bar{\ell}(y)$ subject to the first k components of y being fixed at the values $y^1, a^2,\ldots,a^k$, respectively. Analogously, $\tilde{x} = \tilde{x}(y^1, a^2,\ldots,a^k)$ is a function of y^1, and is the value of x maximizing $\ell(x)$ subject to the constraints $g^1(x) = y^1, g^2(x) = a^2,\ldots,g^k(x) = a^k$. Then $b^*(y^1)$ in (2) is given by

$$\begin{aligned} b^*(y^1) &= \left(\frac{\det\{\Omega(\hat{y})\}}{\det[\Omega'\{\tilde{y}(y^1, a^2,\ldots,a^k)\}]}\right)^{\frac{1}{2}}\frac{\bar{b}\{\tilde{y}(y^1, a^2,\ldots,a^k)\}}{\bar{b}(\hat{y})} \\ &= \left(\frac{\det[\bar{\Lambda}\{\tilde{x}(y^1, a^2,\ldots,a^k)\}]}{\det\{\Lambda(\hat{x})\}\det[\Theta\{\tilde{x}(y^1, a^2,\ldots,a^k)\}]}\right)^{\frac{1}{2}}\frac{b\{\tilde{x}(y^1, a^2,\ldots,a^k)\}}{b(\hat{x})}, \end{aligned} \quad (7)$$

and $\ell^*(y^1)$ is given by

$$\ell^*(y^1) = \bar{\ell}\{\tilde{y}(y^1, a^2, \ldots, a^k)\} - \bar{\ell}(\hat{y}) = \ell\{\tilde{x}(y^1, a^2, \ldots, a^k)\} - \ell(\hat{x});$$

see (4) and (5), respectively. DiCiccio and Martin's (1991) tail probability approximation for densities of the form (6) is

$$\text{pr}(Y^1 \leq a^1 \mid Y^2 = a^2, \ldots, Y^k = a^k) \approx \Phi(r) + \phi(r)\left\{\frac{1}{r} + \frac{\{-\ell^{*(2)}(\bar{y}^1)\}^{\frac{1}{2}}}{\ell^{*(1)}(a^1)} \frac{b^*(a^1)}{b^*(\bar{y}^1)}\right\}, \quad (8)$$

where $\bar{y}^1$ maximizes $\ell^*(y^1)$, $r = \text{sgn}(a^1 - \bar{y}^1)[2\{\ell^*(\bar{y}^1) - \ell^*(a^1)\}]^{\frac{1}{2}}$, $\ell^{*(1)}(y^1) = d\ell^*(y^1)/dy^1$, $\ell^{*(2)}(y^1) = d^2\ell^*(y^1)/d(y^1)^2$ denote the first two derivatives of $\ell^*(y^1)$, and Φ and ϕ denote standard normal distribution and density functions, respectively. A simpler approximation to the required conditional probability is to just use the leading term of (8); that is,

$$\text{pr}(Y^1 \leq a^1 \mid Y^2 = a^2, \ldots, Y^k = a^k) \approx \Phi(r). \quad (9)$$

This alternative approximation is much easier to compute than the full approximation (8), but it is also significantly less accurate in our experience. Typically, the error in approximation (8) is of order $O(n^{-3/2})$, while the error in approximation (9) is of order $O(n^{-\frac{1}{2}})$.

We now outline the expression of the various components of tail probability approximation (8) in terms of the original functions $b, \bar{b}, \ell$, and $\bar{\ell}$. To this end, note that $\bar{y}^1$ maximizes $\bar{\ell}(y)$ subject to $y^2, \ldots, y^k$ being fixed at their conditioned values $a^2, \ldots, a^k$, and let $\bar{x}$ be the value of x maximizing $\ell(x)$ subject to $g^2(x) = a^2, \ldots, g^k(x) = a^k$. Then $\bar{y}^1 = g^1(\bar{x})$ and $\bar{x} = \tilde{x}(\bar{y}^1, a^2, \ldots, a^k)$. Hence,

$$r = \text{sgn}\{a^1 - g^1(\bar{x})\}\big(2[\ell(\bar{x}) - \ell\{\tilde{x}(a^1, \ldots, a^k)\}]\big)^{\frac{1}{2}}.$$

Next, observe from (7) that

$$\frac{b^*(a^1)}{b^*(\bar{y}^1)} = \left(\frac{\det[\bar{\Lambda}\{\tilde{x}(a^1, \ldots, a^k)\}]\det\{\Theta(\bar{x})\}}{\det\{\bar{\Lambda}(\bar{x})\}\det[\Theta\{\tilde{x}(a^1, \ldots, a^k)\}]}\right)^{\frac{1}{2}} \frac{b\{\tilde{x}(a^1, \ldots, a^k)\}}{b(\bar{x})}, \quad (10)$$

which is readily computed using values of ℓ_{ij} and the Lagrange multipliers $\lambda_\alpha(a^1,\ldots,a^k)$ $(\alpha = 1,\ldots,k)$ obtained in finding $\tilde{x}(a^1,\ldots,a^k)$.

Now, to compute $\ell^{*(1)}(y^1)$, it is convenient to work with the definition of $\ell^*(y^1)$ involving $\bar{\ell}$. Then,

$$\begin{aligned}\ell^{*(1)}(y^1) &= \frac{d}{dy^1}[\bar{\ell}\{\tilde{y}(y^1,a^2,\ldots,a^k)\}] \\ &= \frac{d}{dy^1}[\bar{\ell}\{y^1,a^2,\ldots,a^k,\tilde{y}^{k+1}(y^1,a^2,\ldots,a^k),\ldots,\tilde{y}^p(y^1,a^2,\ldots,a^k)\}] \\ &= \bar{\ell}_i\{\tilde{y}(y^1,a^2,\ldots,a^k)\}\tilde{y}^i_1(y^1,a^2,\ldots,a^k),\end{aligned}$$

where $\tilde{y}^i_1(y^1,\ldots,y^k) = \partial\tilde{y}^i/\partial y^1$ and the index i runs from 1 to p. However, $\tilde{y}^\alpha_1(y^1,a^2,\ldots,a^k)$ equals 1 for $\alpha = 1$ and zero for $\alpha = 2,\ldots,k$. Moreover, $\bar{\ell}_{i'}\{\tilde{y}(y^1,a^2,\ldots,a^k)\} = 0$ for $i' = k+1,\ldots,p$ since $\tilde{y}(y^1,a^2,\ldots,a^k)$ maximizes $\bar{\ell}$ subject to the first k components of y being held fixed at the values $y^1,a^2,\ldots,a^k$, respectively. Hence,

$$\ell^{*(1)}(y^1) = \bar{\ell}_1\{\tilde{y}(y^1,a^2,\ldots,a^k)\}.$$

Now, the Lagrangian for maximizing $\bar{\ell}(y)$ subject to the first k components of y being held fixed is $\bar{H}(y) = \bar{\ell}(y) + \lambda_\alpha y^\alpha$, where the index α runs from 1 to k. Note that the Lagrange multipliers $\lambda_\alpha(y^1,a^2,\ldots,a^k)$ $(\alpha = 1,\ldots,k)$ are the same as those for maximizing $\ell(x)$ subject to $g^1(x) = y^1, g^2(x) = a^2,\ldots,g^k(x) = a^k$. A Lagrange multiplier argument involving $\bar{H}(y)$ yields $\lambda_\alpha(y^1,a^2,\ldots,a^k) = -\bar{\ell}_\alpha\{\tilde{y}(y^1,a^2,\ldots,a^k)\}$ $(\alpha = 1,\ldots,k)$. In particular, we have

$$\ell^{*(1)}(y^1) = -\lambda_1(y^1,a^2,\ldots,a^k). \tag{11}$$

The second derivative $\ell^{*(2)}(y^1)$ is harder to compute, and there does not seem to be a closed form expression for it in general. However, it is readily approximated numerically using the formula

$$\ell^{*(2)}(y^1) \approx \frac{\ell\{\tilde{x}(y^1+\delta,a^2,\ldots,a^k)\} - 2\ell\{\tilde{x}(y^1,a^2,\ldots,a^k)\} + \ell\{\tilde{x}(y^1-\delta,a^2,\ldots,a^k)\}}{\delta^2}, \tag{12}$$

for a small value of δ. It is convenient in this instance to work with the definition of $\ell^*(y^1)$ involving $\ell(x)$. Further details concerning computation of (12) are given in Section 3. Tail probability approximation (8) may then be computed.

An important special case of our approximation (8) occurs when $k = p$; that is, when the number of conditioning variables is $p-1$. This is the only case considered by Skovgaard (1987) and Wang (1991). In that case, the marginalization step to approximate the marginal density of $Y^1, \ldots, Y^k$ is not required and the function $\bar{\ell}$ and its derivatives are easily specified. Therefore the conditional density $f_{Y^1|Y^2,\ldots,Y^p}(y^1 \mid Y^2 = a^2, \ldots, Y^p = a^p)$ is proportional to $\bar{b}(y^1, a^2, \ldots, a^p)\exp\{\bar{\ell}(y^1, a^2, \ldots, a^p)\}$. Consequently, approximation (8) assumes the particularly simple form

$$\begin{aligned} &\mathrm{pr}(Y^1 \leq a^1 \mid Y^2 = a^2, \ldots, Y^p = a^p) \\ &\qquad \approx \Phi(r) + \phi(r)\left[\frac{1}{r} + \frac{\{-\bar{\ell}_{11}(\bar{y}^1, a^2, \ldots, a^p)\}^{\frac{1}{2}}}{\bar{\ell}_1(a^1, \ldots, a^p)} \frac{\bar{b}(a^1, \ldots, a^p)}{\bar{b}(\bar{y}^1, a^2, \ldots, a^p)}\right], \end{aligned} \tag{13}$$

where $r = \mathrm{sgn}(a^1 - \bar{y}^1)[2\{\bar{\ell}(\bar{y}^1, a^2, \ldots, a^p) - \bar{\ell}(a^1, \ldots, a^p)\}]^{\frac{1}{2}}$ and $\bar{y}^1$ maximizes $\bar{\ell}(y)$ subject to $y^2, \ldots, y^p$ being held fixed at their conditioned values, $a^2, \ldots, a^p$, respectively.

3. Computational Details

In this section, we outline some computational details for computing (8). Our aim is to provide a formal algorithm for the method. Assume that the conditioning values $a^2, \ldots, a^k$ are specified, and we wish to approximate $\mathrm{pr}(Y^1 \leq a^1 \mid Y^2 = a^2, \ldots, Y^k = a^k)$ for various values of a^1.

Step 1: Find $\tilde{x}$, the value of x which maximizes $\ell(x)$ subject to the constraints $g^1(x) = a^1, \ldots, g^k(x) = a^k$. Typically, this step involves the solution of a system of $(p+k)$ non-linear equations in as many unknowns to obtain $\tilde{x}$ and the Lagrange multipliers $\lambda_1(a^1, \ldots, a^k), \ldots, \lambda_k(a^1, \ldots, a^k)$ needed for later steps. Of course, the system of equations referred to here is linear in the λ_i's, and so a little algebra usually results in the reduction of the problem to that of solving a system of p nonlinear equations in p unknowns.

Step 2: Find $\bar{x}$, the value of x maximizing $\ell(x)$ subject to $g^2(x) = a^2, \ldots, g^k(x) = a^k$. This step typically requires solution of $(p+k-1)$ non-linear equations in as many unknowns.

Step 3: Form $r = \text{sgn}\{a^1 - g^1(\bar{x})\}[2\{\ell(\bar{x}) - \ell(\tilde{x})\}]^{\frac{1}{2}}$, and hence approximation (9) to the required conditional distribution function.

Step 4: Calculation of the factor $b^*(a^1)/b^*(\bar{y}^1)$ follows readily from (10). This step requires calculation of the derivatives $\ell_{ij}(x)$ and $g^{\alpha}_{ij}(x)$ $(\alpha = 1, \ldots, k)$. These derivatives are readily calculated either analytically or numerically.

Step 5: The first derivative $\ell^{*(1)}(a^1)$ is given by $-\lambda_1(a^1, \ldots, a^k)$ from Step 1.

Step 6: The second derivative $\ell^{*(2)}(\bar{y}^1)$ is calculated numerically using (12). In order to calculate (12), the quantities $\tilde{x}(\bar{y}^1 + \delta, a^2, \ldots, a^k)$ and $\tilde{x}(\bar{y}^1 - \delta, a^2, \ldots, a^k)$ are needed. In the former case, $\tilde{x}(\bar{y}^1 + \delta, a^2, \ldots, a^k)$ is that value of x maximizing $\ell(x)$ subject to the constraints $g^1(x) = \bar{y}^1 + \delta, g^2(x) = a^2, \ldots, g^k(x) = a^k$, and so may be found by solving the system of equations from Step 1 except with a^1 replaced by $\bar{y}^1 + \delta$. Similarly, $\tilde{x}(\bar{y}^1 - \delta, a^2, \ldots, a^k)$ may be found by solving the same system of equations with a^1 replaced by $\bar{y}^1 - \delta$. Approximation (8) may then be computed using (10), (11) and (12).

In implementation of the algorithm described above we have used the packages Minpack and NAG to solve the required systems of equations.

4. Examples

4.1. Conditional inference for a normal mean when the coefficient of variation is known.

Let $W_1, \ldots, W_n$ be a sample from a normal distribution $N(\mu, c^2\mu^2)$, $\mu > 0$, with known coefficient of variation $c > 0$. The mean μ is the parameter of interest. For simplicity, let $c = 1$. Let $T = \overline{W}$ and $U = \sum_{i=1}^{n}(W_i - \overline{W})^2$. The statistic $Z = g^2(T, U) = n^{\frac{1}{2}}T/(U + nT^2)^{\frac{1}{2}}$ is ancillary; see Hinkley (1977) and Lehmann (1991). Lehmann (1991, p.549) gives the conditional density of $V = g^1(T, U) = (U + nT^2)^{\frac{1}{2}}$ given $Z = z$:

$$f_{V|Z}(v \mid Z = z) = k\mu^{-n}v^{n-1}\exp\{-\tfrac{1}{2}(\mu^{-1}v - n^{\frac{1}{2}}z)^2\}. \qquad (14)$$

In this case, the number of conditioning variables, k, is equal to $p - 1$, so the simpler approximation (13) may be used to approximate conditional tail probabilities $\text{pr}(V \leq v \mid Z = z)$. Of course, in this instance, tail probabilities can be computed easily through numerical integration of the exact density (14), so this example serves merely to illustrate

the accuracy of our method. Suppose the true value of μ is 1. Then, the joint density of T and U is proportional to $u^{(n-3)/2}\exp\{-\frac{1}{2}n(t-1)^2 - \frac{1}{2}u\}$. We choose $b(t,u) = 1$ and $\ell(t,u) = -\frac{1}{2}n(t-1)^2 - \frac{1}{2}u + \frac{1}{2}(n-3)\log u$ here, rather than the more obvious choice $b(t,u) = u^{(n-3)/2}$, $\ell(t,u) = -\frac{1}{2}n(t-1)^2 - \frac{1}{2}u$, because in the latter case $\partial\ell(t,u)/\partial u$ never vanishes, meaning that $(\hat{t},\hat{u})$ falls on the boundary of possible (t,u) values. Now, $\bar{b}(v,z) = 2n^{-\frac{1}{2}}v^2$ and $\bar{\ell}(v,z) = zvn^{\frac{1}{2}} - \frac{1}{2}n - \frac{1}{2}v^2 + (n-3)\log v + \frac{1}{2}(n-3)\log(1-z^2)$. The constrained maximizing point $\bar{v} = \bar{v}(z)$ of $\bar{\ell}(v,z)$ subject to z being fixed at its conditioned value may be computed algebraically here, so approximation (13) becomes

$$\mathrm{pr}(V \le v \mid Z = z) \approx \Phi(r) + \phi(r)\left\{\frac{1}{r} + \frac{(\bar{v}^2+n-3)^{\frac{1}{2}}v^3}{(zvn^{\frac{1}{2}} - v^2 + n - 3)\bar{v}^3}\right\},$$

where $r = \mathrm{sgn}(v-\bar{v})[2\{\bar{\ell}(\bar{v},z) - \bar{\ell}(v,z)\}]^{\frac{1}{2}}$ and $\bar{v} = \frac{1}{2}zn^{\frac{1}{2}}[1 + \{1 + 4z^{-2}n^{-1}(n-3)\}^{\frac{1}{2}}]$.

Table 1 reports approximations $\mathrm{pr}(V \le v \mid Z = z)$ when $n = 10$ and $z = 0.75$ for various values of v. In this instance, our approximation is very close to exact values based on numerical integration of (14). The simple approximation (9) tends to perform poorly.

Wang (1991) considers tail probabilities for $V^2 = n^{-1}\sum W_i^2$ given $Z = z$. His method, which is based on a saddlepoint approximation to the joint distribution of several means, is not valid in our case as it can only be used to find the conditional distribution of a linear function of means, here the mean of W^2 itself, given $(p-1)$ smooth functions of the means.

4.2. Saddlepoint approximations and an application to the conditional bootstrap.

Skovgaard (1987) and Wang (1991) consider the special case where $X = (X^1,\ldots,X^p)$ is a vector of means and the density $f_X(x)$ is approximated by a saddlepoint approximation. In that setting, consider n observations of a p-dimensional random vector $W = (W_1,\ldots,W_p)$. Denote the cumulant generating function of W by $K(T_1,\ldots,T_p)$. Then the usual saddlepoint approximation to the joint density of $X = (\overline{W}_1,\ldots,\overline{W}_p)$ is proportional to

$$\hat{f}_X(x^1,\ldots,x^p) \propto |\hat{\Delta}(x^1,\ldots,x^p)|^{-\frac{1}{2}}\exp[n\{K(\hat{T}_1,\ldots,\hat{T}_p) - \sum_{i=1}^{p}\hat{T}_i x^i\}],$$

where the saddlepoint $(\hat{T}_1,\ldots,\hat{T}_p)$ satisfies

$$K_{T_i}(\hat{T}_1,\ldots,\hat{T}_p) = x^i \qquad (i = 1,\ldots,p)$$

$K_{T_i} = \partial K(T_1,\ldots,T_p)/\partial T_i$, and $\hat{\Delta} = \{K_{T_iT_j}(\hat{T}_1,\ldots,\hat{T}_p)\}$ is the $k \times k$ matrix of second-order partial derivatives $K_{T_iT_j}(T_1,\ldots,T_p) = \partial^2 K(T_1,\ldots,T_p)/\partial T_i \partial T_j$ $(i,j = 1,\ldots,p)$ evaluated at $\hat{T}_1,\ldots,\hat{T}_p$. General reviews of saddlepoint methods are given by Barndorff-Nielsen and Cox (1979) and Reid (1988).

Approximation (12) can be used to approximate conditional tail probabilities $\mathrm{pr}(Y^1 \leq a^1 \mid Y^2 = a^2,\ldots,Y^k = a^k)$ where $Y^1 = g^1(X),\ldots,Y^k = g^k(X)$ are smooth functions of the means. It is convenient in this instance to take $b(x^1,\ldots,x^p) = |\hat{\Delta}(x^1,\ldots,x^p)|^{-\frac{1}{2}}$ and $\ell(x^1,\ldots,x^p) = n\{K(\hat{T}_1,\ldots,\hat{T}_p) - \sum_{i=1}^p \hat{T}_i x^i\}$. Wang's (1991) method is only valid when the function g^1 is linear. Tierney, Kass and Kadane (1989) give an approximate marginal density formula for $Y^1,\ldots,Y^k$, from which application of (12) is straightforward.

We are particularly interested in applying (12) to estimate tail probabilities for the conditional bootstrap. Monte Carlo simulation to estimate conditional bootstrap distribution functions is extremely tedious, requiring careful stratification of bootstrap resamples according as to whether bootstrap versions of the conditioned variables approximately satisfy the original conditioning or not; see Hinkley and Schechtman (1987) and Davison and Hinkley (1988). In particular, a difficulty arises in deciding how close resamples need to come to the original conditioning criteria to be retained in the simulation. However, recent methods based on saddlepoint approximations have been proposed to approximate bootstrap distribution functions which avoid the need for resampling; see Davison and Hinkley (1988), Daniels and Young (1991), and DiCiccio, Martin and Young (1990). Analytical approximations such as these are particularly important in the conditional bootstrap context because they avoid the stratification problem discussed above. Davison and Hinkley (1988) consider conditional bootstrap inference for the ratio $\theta = E(V)/E(U)$, where (U_i, V_i) $(i = 1,\ldots,n)$ are pairs with common distribution function F. They suggest a suitable model for studying the conditional distribution of $T = \overline{V}/\overline{U}$ given $U_1,\ldots,U_n$ is $v_i = \theta u_i + u_i^{\alpha}\epsilon_i$, where ϵ_i are independent errors with zero mean and variance σ^2. For simplicity, let $\alpha = 1$. Then $\mathrm{Var}(\overline{V} \mid u_1,\ldots,u_n) = \sigma^2 c$ where $c = (\sum u_i^2)/(\sum u_i)^2$. The aim is to approximate the conditional bootstrap distribution of $T^* = \overline{V}^*/\overline{U}^*$ given the

bootstrap ancillary $A^* = (\sum U_i^{*2})/(\sum U_i^*)^2$, where (U_i^*, V_i^*) $(i = 1, \ldots, n)$ is a resample drawn at random with replacement from (U_i, V_i) $(i = 1, \ldots, n)$. Davison and Hinkley find approximations to

$$\mathrm{pr}(\overline{V}^*/\overline{U}^* \leq t \mid \overline{A}_1^* = a_1, \overline{A}_2^* = a_2) = \mathrm{pr}(\overline{V}^* - t\overline{U}^* \leq 0 \mid \overline{A}_1^* = a_1, \overline{A}_2^* = a_2),$$

where $\overline{A}_1^* = n^{-1}\sum U_i^*$ and $\overline{A}_2^* = n^{-1}\sum U_i^{*2}$, by applying Skovgaard's (1987) method. They condition on both $\overline{A}_1^*$ and $\overline{A}_2^*$ reasoning that, since for their data A^* is highly correlated with $\overline{A}_1^*$ and $\overline{A}_2^*$, "redundancy of a conditioning variable is harmless". Note that Skovgaard's method requires two conditioning variables in this case, because here $p = 3$ since $(X^1, X^2, X^3) = (\overline{U}^*, \overline{V}^*, \overline{U^2}^*)$. However, approximation (12) allows us to approximate conditional tail probabilities $\mathrm{pr}(\overline{V}^*/\overline{U}^* \leq t \mid A^* = a)$ directly.

Table 2 reports conditional tail probability approximations for $\overline{V}^*/\overline{U}^*$ for the data set of size 25 reported by Davison and Hinkley (1988, Table 3). We have repeated Davison and Hinkley's experiment with two conditioning variables $\overline{A}_1^*$ and $\overline{A}_2^*$ using approximation (13), and we have also obtained approximations to $\mathrm{pr}(\overline{V}^*/\overline{U}^* \leq t \mid A^* = a)$ using (12) which could not be obtained using their method. In both cases, the joint cumulant generating function of (U_i^*, V_i^*, U_i^{*2}) $(i = 1, \ldots, n)$ given (U_i, V_i) $(i = 1, \ldots, n)$ is

$$K(T_1, T_2, T_3) = \log\left\{n^{-1}\sum_{i=1}^{n}\exp(T_1U_i + T_2V_i + T_3U_i^2)\right\}.$$

In the latter case, we consider the conditional distribution of $Y^1 = g^1(X^1, X^2, X^3) = X^2/X^1$ given $Y^2 = g^2(X^1, X^2, X^3) = n^{-1}X^3/(X^1)^2$. In each case, the derivatives of K and g^i are easily calculated algebraically or numerically and the constrained maximization steps were carried out using Minpack's hybrd subroutine. In order to obtain the "exact" probabilities for the first case, resamples from the simulation experiments were stratified by requiring each bootstrap ancillary to be no more than one quarter of its standard deviation from its observed data value. This requirement meant that only about 10% of the bootstrap resamples drawn were actually retained in estimating the exact probability. In the second case, resamples for which the bootstrap ancillary was no more than one tenth

of its standard deviation from its observed data value were retained, resulting in a 10% retention rate again. The results reported in Table 2 are very encouraging. In particular, in the latter case when there is only one conditioning variable, approximation (12) performs very well, especially in the upper tail of the distribution. The simpler approximation (9) fails badly in this case, particularly in the center of the distribution. In the case of two conditioning variables, approximation (9) and approximation (13) both perform well, especially in the lower tail.

Appendix

Proof of Proposition 1

First note that since $\bar{\ell}(y) = \ell\{x(y)\}$, $\tilde{x}(y) = x(\tilde{y})$ and $\hat{x} = x(\hat{y})$, we have

$$\exp\{\ell(\tilde{x}) - \ell(\hat{x})\} = \exp[\ell\{\tilde{x}(y)\} - \ell\{x(\hat{y})\}] = \exp\{\bar{\ell}(\tilde{y}) - \bar{\ell}(\hat{y})\}.$$

Next, observe that

$$\frac{\bar{b}(\tilde{y})}{\bar{b}(\hat{y})} = \frac{b\{x(\tilde{y})\}\det[J\{x(\hat{y})\}]}{b\{x(\hat{y})\}\det[J\{x(\tilde{y})\}]} = \frac{b(\tilde{x})}{b(\hat{x})}\frac{\det\{J(\hat{y})\}}{\det\{J(\tilde{y})\}}.$$

Consequently, to show that (2) and (3) coincide, it remains to show

$$\frac{\det\{\bar{\Lambda}(\tilde{x})\}}{\det\{\Lambda(\hat{x})\}\det\{\Theta(\tilde{x})\}} = \frac{\det\{\Omega(\hat{y})\}\det\{J(\hat{y})\}^2}{\det\{\Omega'(\tilde{y})\}\det\{J(\tilde{y})\}^2}. \tag{A.1}$$

Note that since $\ell(x) = \bar{\ell}\{y(x)\}$, it follows that

$$\ell_i(x) = \bar{\ell}_k\{y(x)\}g_i^k(x) \qquad (i = 1, \ldots, p),$$

where the index k runs from 1 to p, and

$$\ell_{ij}(x) = \bar{\ell}_{kl}\{y(x)\}g_i^k(x)g_j^l(x) + \bar{\ell}_k\{y(x)\}g_{ij}^k(x) \qquad (i, j = 1, \ldots, p), \tag{A.2}$$

where the indices k, l run from 1 to p. Now, since $\ell_i(\hat{x}) = 0$ and $\bar{\ell}_k(\hat{y}) = 0$ $(i, k = 1, \ldots, p)$, we have

$$-\ell_{ij}(\hat{x}) = -\bar{\ell}_{kl}(\hat{y})g_i^k(\hat{x})g_j^l(\hat{x}),$$

and hence, inverting and taking determinants on both sides,

$$\det\{\Lambda(\hat{x})\} = [\det\{\Omega(\hat{y})\}\det\{J(\hat{y})\}^2]^{-1}. \tag{A.3}$$

Recall that $\tilde{y}$ maximizes $\bar{\ell}(y)$ subject to the first k components of y being fixed. Hence $\bar{\ell}_{i'}(\tilde{y}) = 0$, for $i' = k+1, \ldots, p$. Consequently, from (A.2) it follows that

$$\begin{aligned}\ell_{ij}(\tilde{x}) &= \bar{\ell}_{kl}(\tilde{y}) g_i^k(\tilde{x}) g_j^l(\tilde{x}) + \bar{\ell}_\alpha(\tilde{y}) g_{ij}^\alpha(\tilde{x}) \\ &= \bar{\ell}_{kl}(\tilde{y}) g_i^k(\tilde{x}) g_j^l(\tilde{x}) - \lambda_\alpha g_{ij}^\alpha(\tilde{x}),\end{aligned}$$

where the index α runs from 1 to k, and λ_α $(\alpha = 1, \ldots, k)$ are the Lagrange multipliers for the constrained maximization step to obtain $\tilde{y}$. Note as well that λ_α $(\alpha = 1, \ldots, k)$ are also the Lagrange multipliers for finding $\tilde{x}$. Thus,

$$\bar{\ell}_{kl}(\tilde{y}) g_i^k(\tilde{x}) g_j^l(\tilde{x}) = \ell_{ij}(\tilde{x}) + \lambda_\alpha g_{ij}^\alpha(\tilde{x}) = H_{ij}(\tilde{x}).$$

Matrix inversion on both sides yields

$$\bar{\ell}^{kl}(\tilde{y}) = H^{ij}(\tilde{x}) g_i^k(\tilde{x}) g_j^l(\tilde{x}), \tag{A.4}$$

and hence

$$[\det\{\Omega(\tilde{y})\}]^{-1} = \det\{\bar{\Lambda}(\tilde{x})\} \det\{J(\tilde{y})\}^2. \tag{A.5}$$

Equation (A.4) implies that $\Theta(\tilde{x})$ is the $k \times k$ submatrix in the upper left corner of the inverse of $\Omega(\tilde{y})$. Therefore, by the formula for the determinant of a partitioned inverse (Draper and Smith, 1981, p.127),

$$\det\{\Omega^{-1}(\tilde{y})\} = \det\{\Theta(\tilde{x})\} \det[\{\Omega'(\tilde{y})\}^{-1}]. \tag{A.6}$$

Hence, from (A.5) and (A.6) it follows that

$$\frac{\det\{\bar{\Lambda}(\tilde{x})\}}{\det\{\Theta(\tilde{x})\}} = \frac{1}{\det\{\Omega'(\tilde{y})\} \det\{J(\tilde{y})\}^2}.$$

Combining (A.3) and (A.6), (A.1) obtains, and the proof is complete.

References

Barndorff-Nielsen, O.E. & Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. B* **52**, 485–96.

Daniels, H.E. & Young, G.A. (1991). Saddlepoint approximation for the studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169–79.

Davison, A.C. & Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–31.

DiCiccio, T.J. & Martin, M.A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika* **78**, 891–902.

DiCiccio, T.J., Martin, M.A. & Young, G.A. (1990). Analytical approximations to bootstrap distribution functions using saddlepoint methods. Technical Report No. 356, Department of Statistics, Stanford University.

Draper, N.R. & Smith, H. (1981). *Applied Regression Analysis*, Second edition. Wiley: New York.

Hinkley, D.V. (1977). Conditional inference about a normal mean with known coefficient of variation. *Biometrika* **64**, 105–8.

Hinkley, D.V. & Schechtman, E. (1987). Conditional bootstrap methods in the mean shift model. *Biometrika* **75**, 85–94.

Lehmann, E.L. (1991). *Testing Statistical Hypotheses*, Second Edition. Wadsworth and Brooks/Cole: Pacific Grove, California.

Phillips, P.C.B. (1983). Marginal densities of instrumental variables estimators in the general single equation case. *Adv. Econometrics* **2**, 1–24.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213–38.

Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24**, 875–87.

Tierney, L., Kass, R.E. & Kadane, J.B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**, 425–33, and correction note (1991), *Biometrika* **78**, 233-4.

Wang, S. (1991) Saddlepoint approximations in conditional inference. Unpublished technical report.

Table 1

Conditional tail probabilities pr($V \leq v \mid Z = 0.75$) for Example 4.1, $n = 10$.

v	Exact	Approximation (9)	Approximation (13)
1.5	0.0000	0.0003	0.0000
2.0	0.0005	0.0036	0.0004
2.5	0.0057	0.0234	0.0048
3.0	0.0331	0.0910	0.0300
3.5	0.1190	0.2394	0.1125
4.0	0.2920	0.4595	0.2828
4.5	0.5264	0.6878	0.5174
5.0	0.7471	0.8575	0.7409
5.5	0.8949	0.9494	0.8918
6.0	0.9664	0.9862	0.9653
6.5	0.9918	0.9971	0.9915
7.0	0.9985	0.9995	0.9984
7.5	0.9998	0.9999	0.9998
8.0	1.0000	1.0000	1.0000

Table 2

Approximations to conditional probabilities $\text{pr}(\bar{V}^*/\bar{U}^* \leq t \mid \bar{A}_1^* = 147.9, \bar{A}_2^* = 43120)$ and $\text{pr}(\bar{V}^*/\bar{U}^* \leq t \mid A^* = 0.07885)$ for $n = 25$ pairs of Example 4.2. The conditioned values chosen are the values of $\overline{A}_1$, $\overline{A}_2$ and A from the data.

	Two conditioning variables			One conditioning variable		
t	Exact†	(9)	(13)	Exact†	(9)	(12)
7.8	0.0001	0.0001	0.0001	0.0003	0.0008	0.0004
7.9	0.0002	0.0002	0.0002	0.0006	0.0016	0.0009
8.0	0.0008	0.0007	0.0007	0.0013	0.0033	0.0020
8.1	0.0020	0.0020	0.0020	0.0029	0.0066	0.0042
8.2	0.0050	0.0051	0.0051	0.0061	0.0126	0.0083
8.3	0.0115	0.0118	0.0117	0.0125	0.0231	0.0157
8.4	0.0247	0.0252	0.0249	0.0243	0.0405	0.0289
8.5	0.0488	0.0497	0.0489	0.0450	0.0684	0.0511
8.6	0.0888	0.0905	0.0892	0.0794	0.1108	0.0868
8.7	0.1497	0.1528	0.1506	0.1328	0.1715	0.1411
8.8	0.2352	0.2394	0.2363	0.2107	0.2534	0.2187
8.9	0.3432	0.3492	0.3452	0.3155	0.3565	0.3219
9.0	0.4664	0.4752	0.4709	0.4433	0.4763	0.4456
9.1	0.5948	0.6059	0.6018	0.5835	0.6031	0.5863
9.2	0.7151	0.7276	0.7243	0.7185	0.7241	0.7188
9.3	0.8157	0.8287	0.8266	0.8297	0.8264	0.8298
9.4	0.8920	0.9032	0.9021	0.9089	0.9023	0.9092
9.5	0.9429	0.9514	0.9511	0.9573	0.9513	0.9576
9.6	0.9727	0.9786	0.9786	0.9828	0.9786	0.9827
9.7	0.9883	0.9918	0.9918	0.9936	0.9918	0.9938
9.8	0.9955	0.9973	0.9973	0.9980	0.9972	0.9981
9.9	0.9984	0.9992	0.9992	0.9995	0.9992	0.9995
10.0	0.9995	0.9998	0.9998	0.9999	0.9998	0.9999
10.1	0.9999	1.0000	1.0000	1.0000	0.9999	1.0000
10.2	1.0000				1.0000	

† "Exact" probabilities based on 500,000 retained bootstrap resamples.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 452	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Analytical approximations to conditional distribution functions		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas J. DiCiccio Michael A. Martin G. Alastair Young		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE April 20, 1992 13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Also issued as Technical Report No. 394 on National Science Foundation Grant DMS89-05874

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ancillary statistic; conditional bootstrap, Laplace's method, marginal density, saddlepoint approximation, tail probability approximation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SUMMARY. Conditional inference plays a central role in statistics, but determination of relevant conditional distributions is often difficult. We develop analytical procedures that are accurate and easy to apply for approximating conditional distribution functions. For a continuous random vector $X = (X^1, \ldots, X^p)$, we estimate conditional tail probabilities

$$\text{pr}(Y^1 \leq a^1 | Y^2 = a^2, \ldots, Y^k = a^k), \qquad k \leq p,$$

where $Y^i = g^i(X^1, \ldots X^p)$ $(i = 1, \ldots, k)$ and $g^1, \ldots, g^k$ are smooth functions of X. Previous approaches have dealt with the cases where the variable whose conditional distribution is sought is a linear function of means, and where there are $p - 1$ conditioning variables. However, in many practical circumstances the statistic of interest is a nonlinear function of means and it is advantageous to condition on a lower-dimensional ancillary statistic. Our procedure first involves approximating the marginal density function $f_{Y^1,\ldots,Y^k}(y^1, \ldots, y^k)$ for $Y^1, \ldots, Y^k$, by an approach of Phillips (1983) and Tierney, Kass and Kadane (1989). An accurate approximation to the required conditional probability is then obtained by applying a marginal tail probability approximation of DiCiccio and Martin (1991) to the conditional density of Y^1 given $Y^2, \ldots, Y^k$ which satisfies

$$f_{Y^1|Y^2,\ldots,Y^k}(y^1 | Y^2 = a^2, \ldots, Y^k = a^k) \propto f_{Y^1,\ldots,Y^k}(y^1, a^2, \ldots, a^k).$$

Our method is illustrated in several examples, including one which uses a saddlepoint approximation for the density of X, and the method is applied for conditional bootstrap inference.