Biometrika (1993), **80**, 4, *pp.* 781–90
Printed in Great Britain

Analytical approximations to conditional distribution functions

BY THOMAS J. DICICCIO, MICHAEL A. MARTIN
Department of Statistics, Sequoia Hall, Stanford University, Stanford, California 94305-4065, U.S.A.

AND G. ALASTAIR YOUNG
Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England

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 the conditional distribution function of Y^1 given $Y^2, \ldots, Y^k$ $(k \leqslant p)$, where each Y^i is a smooth function 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, sometimes 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 for $Y^1, \ldots, Y^k$, by an approach of Phillips (1983) and Tierney, Kass & Kadane (1989). An accurate approximation to the required conditional probability is then obtained by applying a marginal tail probability approximation of DiCiccio & Martin (1991) to the conditional density of Y^1 given $Y^2, \ldots, Y^k$. 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 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.

Several authors have discussed the use of saddlepoint methods to approximate conditional distributions. Skovgaard (1987) investigated the case of a bivariate mean to develop 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. S. Wang, in the unpublished technical report 'Saddlepoint approximations in conditional inference', extended Skovgaard's results to include the case of approximating the con-

ditional distribution of a mean given $p-1$ nonlinear functions of the means. The techniques of Skovgaard and Wang share several elements that limit their applicability. First, because they are based solely on saddlepoint approximations, the methods require knowledge of the cumulant generating function of the random vector of interest. Secondly, 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 severe in practice. Finally, their methods require the number of conditioning variables to be exactly $p-1$. However, many times, an ancillary of lower dimension than $p-1$ exists, and conditional inference given that ancillary is desired.

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 $k\leqslant p$. The vector X is not restricted to a vector of means, although that is the case that is usually of most interest. Consequently, we are not restricted to using a saddlepoint approximation for the density of X. Also, the variable whose conditional distribution is sought may be a smooth, nonlinear function of X, giving our method considerable generality. Moreover, our method allows the dimension of the conditioning variable to be smaller 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 & Martin's (1991) tail probability approximation to a marginal density approximation proposed by Phillips (1983) and Tierney, Kass & Kadane (1989). A theoretical contribution of the paper is to show that the marginal density approximations of Phillips (1983) and Tierney et al. (1989) are equivalent.

We describe our technique in § 2. Two examples are given in § 3: the first, a simple illustrative example; and the second, 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 $f_X(x)=cb(x)\exp\{l(x)\}$, $x=(x^1,\ldots,x^p)$. Let $\hat{x}=(\hat{x}^1,\ldots,\hat{x}^p)$ be the point maximizing $l(x)$ 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 $l(x)$ and its derivatives are $O(n)$. Interest centres on approximating conditional tail probabilities

$$\mathrm{pr}\,(Y^1\leqslant a^1\,|\,Y^2=a^2,\ldots,Y^k=a^k)\quad(k\leqslant p),$$

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 we assume have continuous gradients that are nonzero 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 et al. (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. 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 nonzero gradients in an $n^{-\frac{1}{4}}$-neighbourhood of $\hat{x}$. Denote the Hessian of this transformation by $J\{x(y)\}$. Then the probability density function of Y is

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

where $\bar{b}(y) = b\{x(y)\}/|J\{x(y)\}|$ and $\bar{l}(y) = l\{x(y)\}$. Let $\hat{y}$ be the value of y maximizing $\bar{l}(y)$, and let $\tilde{y} = \tilde{y}(y^1, \ldots, y^k)$ be the value of y maximizing $\bar{l}(y)$ subject to the first k components of y being held fixed at the values $y^1, \ldots, y^k$. Let

$$\bar{l}_i(y) = \partial\bar{l}(y)/\partial y^i, \quad \bar{l}_{ij}(y) = \partial^2\bar{l}(y)/\partial y^i \partial y^j \quad (i, j = 1, \ldots, p).$$

Then, Phillips' (1983) approximation to the marginal density of $Y^1, \ldots, Y^k$ is

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

where $\Omega(y)$ is the $p \times p$ matrix whose (i, j)th element is $-\bar{l}_{ij}(y)$, and $\Omega'(y)$ is the $(p-k) \times (p-k)$ submatrix of $\Omega(y)$ corresponding to $\{(i, j): i, j = k+1, \ldots, p\}$. Phillips' method assumes a p-dimensional transformation Y of the original variable X, even though only $k < p$ of the Y's are of interest. This assumption suggests that use of approximation (1) requires the explicit specification of $p-k$ 'nuisance' functions $g^{k+1}, \ldots, g^p$ of X, the choice of which could affect the accuracy of (1). Conventional approaches have dealt with this problem by assuming each of the 'nuisance' functions is an appropriate coordinate function, simplifying derivative calculations. A new result arising from our work is that $g^{k+1}, \ldots, g^p$ play no role in the computation of (1), and hence need not be specified at all.

Tierney et al. (1989) also provided a formula for the approximate marginal density of $Y^1, \ldots, Y^k$ but their derivation did not involve the 'nuisance' functions $g^{k+1}, \ldots, g^p$ assumed by Phillips. To describe their formula, we first need additional notation. Let $\tilde{x}$ be the value of x maximizing $l(x)$ subject to the constraints $g^1(x^1, \ldots, x^p) = y^1, \ldots, g^k(x^1, \ldots, x^p) = y^k$, and let $H(x)$ be the Lagrangian for this constrained maximization, $H(x) = l(x) + \lambda_\alpha\{g^\alpha(x) - y^\alpha\}$, where $\lambda_\alpha = \lambda_\alpha(y^1, \ldots, y^k)$ $(\alpha = 1, \ldots, k)$ and the convention applies whereby summation is assumed over indices appearing as both subscript and superscript. Let

$$l_i(x) = \partial l(x)/\partial x^i, \quad l_{ij}(x) = \partial^2 l(x)/\partial x^i \partial x^j, \quad H_{ij}(x) = \partial^2 H(x)/\partial x^i \partial x^j,$$

$$g^\alpha_i(x) = \partial g^\alpha(x)/\partial x^i, \quad g^\alpha_{ij}(x) = \partial^2 g^\alpha(x)/\partial x^i \partial x^j \quad (i, j = 1, \ldots, p; \alpha = 1, \ldots, k)$$

denote the partial derivatives of l, H and g^α, respectively. Define the $p \times p$ matrices $\Lambda(x) = \{-l^{ij}(x)\}$, the inverse of the matrix whose (i, j)th element is $-l_{ij}(x)$, and $\bar{\Lambda}(x) = \{-H^{ij}(x)\}$, the inverse of the matrix where (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 & Kadane's approximation to the marginal density of $Y^1, \ldots, Y^k$ is

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

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

Proposition 1 is proved in an unpublished technical report by the authors. In particular,

it is shown there that the following two relations hold:

$$\left[\frac{\det\{\Omega(\hat{y})\}}{\det\{\Omega'(\tilde{y})\}}\right]^{\frac{1}{2}}\frac{\bar{b}(\tilde{y})}{\bar{b}(\hat{y})}=\left[\frac{\det\{\bar{\Lambda}(\tilde{x})\}}{\det\{\Lambda(\hat{x})\}\det\{\Theta(\tilde{x})\}}\right]^{\frac{1}{2}}\frac{b(\tilde{x})}{b(\hat{x})}, \tag{3}$$

$$\bar{l}(\tilde{y})-\bar{l}(\hat{y})=l(\tilde{x})-l(\hat{x}). \tag{4}$$

In order to approximate conditional distribution functions, we first note 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

$$f_{Y^1|Y^2,\ldots,Y^k}(y^1\mid Y^2=a^2,\ldots,Y^k=a^k)\propto b^*(y^1)\exp\{l^*(y^1)\}, \tag{5}$$

for suitably defined functions l^* and b^*. We apply DiCiccio & Martin's (1991) tail probability formula to obtain approximations to conditional tail probabilities $\mathrm{pr}\,(Y^1\leqslant a^1\mid Y^2=a^2,\ldots,Y^k=a^k)$. 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 $\bar{l}(y)$ 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 $l(x)$ 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 (5) is given by

$$\begin{aligned}b^*(y^1)&=\left(\frac{\det\{\Omega(\hat{y})\}}{\det[\Omega'\{\tilde{y}(y^1,a^2,\ldots,a^k)\}]}\right)^{\frac{1}{2}}\frac{\bar{b}\{\tilde{y}(y^1,a^2,\ldots,a^k)\}}{\bar{b}(\hat{y})}\\&=\left(\frac{\det[\bar{\Lambda}\{\tilde{x}(y^1,a^2,\ldots,a^k)\}]}{\det\{\Lambda(\hat{x})\}\det[\Theta\{\tilde{x}(y^1,a^2,\ldots,a^k)\}]}\right)^{\frac{1}{2}}\frac{b\{\tilde{x}(y^1,a^2,\ldots,a^k)\}}{b(\hat{x})},\end{aligned} \tag{6}$$

and $l^*(y^1)$ is given by

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

see (3) and (4), respectively. DiCiccio & Martin's (1991) tail probability approximation for densities of the form (5) is

$$\mathrm{pr}\,(Y^1\leqslant a^1\mid Y^2=a^2,\ldots,Y^k=a^k)\eqsim\Phi(r)+\phi(r)\left[\frac{1}{r}+\frac{\{-l^{*(2)}(\bar{y}^1)\}^{\frac{1}{2}}}{l^{*(1)}(a^1)}\frac{b^*(a^1)}{b^*(\bar{y}^1)}\right], \tag{7}$$

where $\bar{y}^1$ maximizes $l^*(y^1)$,

$$r=\mathrm{sgn}\,(a^1-\bar{y}^1)[2\{l^*(\bar{y}^1)-l^*(a^1)\}]^{\frac{1}{2}},$$

$l^{*(1)}(y^1)=dl^*(y^1)/dy^1$, $l^{*(2)}(y^1)=d^2l^*(y^1)/d(y^1)^2$ denote the first two derivatives of $l^*(y^1)$, and Φ and ϕ denote standard normal distribution and density functions, respectively. Expression of the various components of approximation (7) in terms of the original functions b and l is given in an Appendix. A simpler approximation to the required conditional probability is to use just the leading term of (7); that is,

$$\mathrm{pr}\,(Y^1\leqslant a^1\mid Y^2=a^2,\ldots,Y^k=a^k)\eqsim\Phi(r). \tag{8}$$

This alternative approximation is much easier to compute than the full approximation (7), but it is also significantly less accurate in our experience. Typically, the error in approximation (7) is of order $O(n^{-3/2})$, while the error in (8) is of order $O(n^{-\frac{1}{2}})$.

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

An important special case of approximation (7) occurs when $k = p$, that is when the number of conditioning variables is $p - 1$. This is the only case for which the techniques of Skovgaard (1987) and of S. Wang, in the technical report mentioned in § 1, apply. Here, the marginalization step to approximate the marginal density of $Y^1, \ldots, Y^k$ is unnecessary and the function $\bar{l}$ and its derivatives are easily specified. Therefore, the conditional density

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

is proportional to $\bar{b}(y^1, a^2, \ldots, a^p) \exp\{\bar{l}(y^1, a^2, \ldots, a^p)\}$, and approximation (7) assumes the simple form

$$\begin{aligned} &\mathrm{pr}\,(Y^1 \leqslant a^1 \mid Y^2 = a^2, \ldots, Y^p = a^p) \\ &\quad \doteq \Phi(r) + \phi(r)\left[\frac{1}{r} + \frac{\{-\bar{l}_{11}(\bar{y}^1, a^2, \ldots, a^p)\}^{\frac{1}{2}}}{\bar{l}_1(a^1, \ldots, a^p)} \frac{\bar{b}(a^1, \ldots, a^p)}{\bar{b}(\bar{y}^1, a^2, \ldots, a^p)}\right], \end{aligned} \tag{9}$$

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

3. Examples

3·1. *Circular normal distribution*

Consider a random sample $(X_1, Y_1)', \ldots, (X_n, Y_n)'$ from a bivariate normal distribution with mean vector $(\theta \cos \lambda, \theta \sin \lambda)'$ and identity variance matrix. Define the statistics $R = g^1(\bar{X}, \bar{Y}) = (\bar{X}^2 + \bar{Y}^2)^{\frac{1}{2}}$ and $W = g^2(\bar{X}, \bar{Y}) = \arctan(\bar{Y}/\bar{X})$, which estimate θ and λ, respectively. Suppose the conditional distribution of R given $W = \lambda_0$ is of interest. In this instance, neither Wang's nor Skovgaard's methods can be applied since g^1 is nonlinear. The joint density of $(\bar{X}, \bar{Y})$ is $f_{\bar{X},\bar{Y}}(x, y) = b(x, y) \exp\{l(x, y)\}$, where

$$b(x, y) = n/(2\pi), \quad l(x, y) = -\tfrac{1}{2}n\{(x - \theta \cos \lambda)^2 + (y - \theta \sin \lambda)^2\},$$

from which the joint density of R and W is easily found to be $f_{R,W}(r, w) = \bar{b}(r, w) \exp\{\bar{l}(r, w)\}$, where

$$\bar{b}(r, w) = nr/(2\pi), \quad \bar{l}(r, w) = -\tfrac{1}{2}n\{(r \cos w - \theta \cos \lambda)^2 + (r \sin w - \theta \sin \lambda)^2\}.$$

A little algebra yields that approximation (9) to the conditional distribution of R given $W = \lambda_0$ is

$$\mathrm{pr}\,(R \leqslant a \mid W = \lambda_0) \doteq \Phi[n^{\frac{1}{2}}\{a - \theta \cos(\lambda - \lambda_0)\}]. \tag{10}$$

Approximation (10) is particularly simple here since the second term on the right-hand side of (9) vanishes.

In this example the exact conditional tail probability may be calculated analytically. Direct integration shows

$$\mathrm{pr}\,(R \leqslant a \mid W = \lambda_0) = I(a)/I(\infty),$$

where

$$I(x) = \exp(-\tfrac{1}{2}n\theta^2 s^2)\left([\exp(-\tfrac{1}{2}n\theta^2 c^2) - \exp\{-\tfrac{1}{2}n(x-\theta c)^2\}]/(2\pi) + \frac{n^{\frac{1}{2}}\theta c}{(2\pi)^{\frac{1}{2}}}[\Phi\{n^{\frac{1}{2}}(x-\theta c)\} - \Phi(-n^{\frac{1}{2}}\theta c)]\right),$$

with $c = \cos(\lambda - \lambda_0)$ and $s = \sin(\lambda - \lambda_0)$.

We carried out a small numerical study to assess the accuracy of our approximation. For this example, we let the true values of θ and λ be 25 and $\pi/6$, respectively. We constructed approximations to $\text{pr}(R \leqslant a \mid W = \lambda_0)$ for various values of n and $\lambda_0 = \frac{1}{2}$, comparing (10) with an exact tail probability, computed from the formula above. The results are presented in Table 1. Approximation (10) performs very well throughout.

Table 1. *Approximations to conditional probabilities* $\text{pr}(R \leqslant a \mid W = \lambda_0)$ *for Example* 3·1

	$n = 5$			$n = 10$			$n = 20$	
a	Exact	(10)	a	Exact	(10)	a	Exact	(10)
23·9	0·0072	0·0073	24·2	0·0061	0·0061	24·4	0·0040	0·0040
24·0	0·0131	0·0132	24·3	0·0142	0·0142	24·5	0·0137	0·0137
24·3	0·0600	0·0606	24·5	0·0590	0·0595	24·6	0·0392	0·0394
24·4	0·0912	0·0924	24·6	0·1059	0·1070	24·7	0·0944	0·0950
24·5	0·1330	0·1351	24·7	0·1749	0·1770	24·8	0·1923	0·1940
24·6	0·1864	0·1897	24·9	0·3797	0·3843	24·9	0·3357	0·3387
24·7	0·2515	0·2562	25·0	0·5037	0·5088	25·0	0·5089	0·5124
25·0	0·4991	0·5062	25·1	0·6279	0·6324	25·1	0·6810	0·6838
25·4	0·8155	0·8186	25·3	0·8322	0·8342	25·2	0·8211	0·8227
25·6	0·9115	0·9126	25·4	0·9000	0·9009	25·3	0·9145	0·9151
25·7	0·9425	0·9430	25·5	0·9452	0·9455	25·5	0·9883	0·9883
25·9	0·9786	0·9787	25·6	0·9724	0·9725	25·6	0·9967	0·9969
26·1	0·9933	0·9933	25·8	0·9946	0·9946	25·7	0·9992	0·9992
26·2	0·9965	0·9965	25·9	0·9979	0·9979	25·8	0·9998	0·9998

For this example, $\theta = 25$, $\lambda = \pi/6$ and $\lambda_0 = \frac{1}{2}$.

3·2. *Saddlepoint approximations and an application to the conditional bootstrap*

Skovgaard (1987) and S. Wang, in his technical report mentioned in § 1, consider the special case where X is a vector of means and the density $f_X(x)$ is approximated by a saddlepoint approximation. In this instance, it is necessary to know the cumulant generating function of X so that the saddlepoint approximation to $f_X(x)$ can be formed. Our methodology can then easily be applied in this setting by appropriate choice of the functions b and l. To describe the saddlepoint approximation, 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 saddlepoint approximation to the joint density of $X = (\bar{W}_1, \ldots, \bar{W}_p)$ is proportional to

$$\hat{f}_X(x^1, \ldots, x^p) \propto |\hat{\Delta}(x^1, \ldots, x^p)|^{-\frac{1}{2}} \exp\left[n\left\{K(\hat{T}_1, \ldots, \hat{T}_p) - \sum_{i=1}^{p} \hat{T}_i x^i\right\}\right], \qquad (11)$$

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

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

$K_{T_i} = \partial K(T_1, \ldots, T_p)/\partial T_i$, 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

$$K_{T_iT_j}(T_1, \ldots, T_p) = \partial^2 K(T_1, \ldots, T_p)/\partial T_i \partial T_j \quad (i, j = 1, \ldots, p)$$

evaluated at $\hat{T}_1, \ldots, \hat{T}_p$. General reviews of saddlepoint methods are given by Barndorff-Nielsen & Cox (1979) and Reid (1988).

Approximation (7) can be used to approximate conditional tail probabilities

$$\text{pr}\,(Y^1 \leqslant a^1 \mid Y^2 = a^2, \ldots, Y^k = a^k),$$

where $Y^1 = g^1(X), \ldots, Y^k = g^k(X)$ are smooth functions of the means, by noting that (11) is in the form (1) with

$$b(x^1, \ldots, x^p) = |\hat{\Delta}(x^1, \ldots, x^p)|^{-\frac{1}{2}}, \quad l(x^1, \ldots, x^p) = n\left\{K(\hat{T}_1, \ldots, \hat{T}_p) - \sum_{i=1}^{p} \hat{T}_i x^i\right\}.$$

Wang's method is only valid when the function g^1 is linear. Tierney et al. (1989) give an approximate marginal density formula for $Y^1, \ldots, Y^k$, from which application of (7) is straightforward.

We are particularly interested in applying (7) 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 (Hinkley & Schechtman, 1987; Davison & Hinkley, 1988). In particular, a difficulty arises in deciding how close resamples need to come to the conditioning criteria to be retained in the simulation. Moreover, the more stringently the conditioning criteria are adhered to, the fewer the resamples that can be retained in estimating the probability. Consequently, the total number of resamples that needs to be drawn to ultimately obtain satisfactory estimates can become overwhelmingly large. Recent methods based on saddlepoint approximations have been proposed to approximate bootstrap distribution functions without the need for any resampling (Davison & Hinkley, 1988; Daniels & Young, 1991; DiCiccio, Martin & Young, 1993). These kinds of approaches are particularly beneficial in a conditional bootstrap framework because they avoid the stratification problem completely.

Davison & 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 $T = \bar{V}/\bar{U}$ given $U_1, \ldots, U_n$ is $v_i = \theta u_i + u_i^{\alpha}\varepsilon_i$, where ε_i are independent errors with zero mean and variance σ^2. For simplicity, let $\alpha = 1$. Then $\text{var}\,(\bar{V} \mid u_1, \ldots, u_n) = \sigma^2 c$, where $c = (\sum u_i^2)/(\sum u_i)^2$. The aim is to approximate the conditional bootstrap distribution of $T^* = \bar{V}^*/\bar{U}^*$ 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 from (U_i, V_i) $(i = 1, \ldots, n)$. Davison & Hinkley approximate

$$\text{pr}\,(\bar{V}^*/\bar{U}^* \leqslant t \mid \bar{A}_1^* = a_1, \bar{A}_2^* = a_2) = \text{pr}\,(\bar{V}^* - t\bar{U}^* \leqslant 0 \mid \bar{A}_1^* = a_1, \bar{A}_2^* = a_2),$$

where $\bar{A}_1^* = n^{-1}\sum U_i^*$ and $\bar{A}_2^* = n^{-1}\sum U_i^{*2}$, by applying Skovgaard's (1987) method. They condition on both $\bar{A}_1^*$ and $\bar{A}_2^*$ reasoning that, since for their data A^* is highly correlated with $\bar{A}_1^*$ and $\bar{A}_2^*$, 'redundancy of a conditioning variable is harmless'. Note that their method requires two conditioning variables, since $X = (\bar{U}^*, \bar{V}^*, \bar{U}^{2*})$. However, approxi-

mation (7) allows us to approximate conditional tail probabilities pr $(\bar{V}^*/\bar{U}^* \leqslant t \mid A^* = a)$ with only one conditioning variable.

Table 2 reports conditional tail probability approximations for $\bar{V}^*/\bar{U}^*$ for the data set of size 25 reported by Davison & Hinkley (1988, Table 3). We repeated Davison & Hinkley's experiment with two conditioning variables $\bar{A}_1^*$ and $\bar{A}_2^*$ using approximation (9), and we have also calculated approximations to pr $(\bar{V}^*/\bar{U}^* \leqslant t \mid A^* = a)$ using (7) which could not be obtained using their method. 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 $Y^2 = g^2(X^1, X^2, X^3) = n^{-1}X^3/(X^1)^2$. The functions K and g^i and their derivatives are easily calculated and the constrained maximization steps can be carried out using commonly available numerical subroutines such as Minpack's HYBRD and NAG's C05NCF. In order to obtain the 'exact' probabilities for the first example, 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 resulted in only about 10% of the bootstrap resamples drawn being used in estimating the exact probability. For the second example, resamples for which the bootstrap ancillary was no more than one-tenth of its standard deviation from its observed data value were retained, again resulting in a 10% retention rate. The results reported in Table 2 are very encouraging. In particular, in the latter case

Table 2. *Approximations to conditional probabilities* pr $(\bar{V}^*/\bar{U}^* \leqslant t \mid \bar{A}_1^* = 147{\cdot}9,\ \bar{A}_2^* = 43120)$ *and* pr $(\bar{V}^*/\bar{U}^* \leqslant t \mid A^* = 0{\cdot}07885)$ *for* $n = 25$ *pairs of Example* 3·2

	Two conditioning variables		One conditioning variable	
t	Exact	Approximation (9)	Exact	Approximation (7)
7·8	0·0001	0·0001	0·0003	0·0004
7·9	0·0002	0·0002	0·0006	0·0009
8·0	0·0008	0·0007	0·0013	0·0020
8·1	0·0020	0·0020	0·0029	0·0042
8·2	0·0050	0·0051	0·0061	0·0083
8·3	0·0115	0·0117	0·0125	0·0157
8·4	0·0247	0·0249	0·0243	0·0289
8·5	0·0488	0·0489	0·0450	0·0511
8·6	0·0888	0·0892	0·0794	0·0868
8·7	0·1497	0·1506	0·1328	0·1411
8·8	0·2352	0·2363	0·2107	0·2187
8·9	0·3432	0·3452	0·3155	0·3219
9·0	0·4664	0·4709	0·4433	0·4456
9·1	0·5948	0·6108	0·5835	0·5863
9·2	0·7151	0·7243	0·7185	0·7188
9·3	0·8157	0·8266	0·8297	0·8298
9·4	0·8920	0·9021	0·9089	0·9092
9·5	0·9429	0·9511	0·9573	0·9576
9·6	0·9727	0·9786	0·9828	0·9827
9·7	0·9883	0·9918	0·9936	0·9938
9·8	0·9955	0·9973	0·9980	0·9981
9·9	0·9984	0·9992	0·9995	0·9995
10·0	0·9995	0·9998	0·9999	0·9999
10·1	0·9999	1·0000	1·0000	1·0000

The conditioned values chosen are the values of $\bar{A}_1$, $\bar{A}_2$ and A from the data. 'Exact' probabilities are based on 500 000 retained bootstrap resamples.

when there is only one conditioning variable, approximation (7) performs very well, especially in the upper tail of the distribution.

Acknowledgements

We are grateful to the Editor, Associate Editor and referee for helpful comments which enabled us to produce a more incisive paper.

Appendix

Computation of approximation (7)

Here we outline the expression of the components of approximation (7) in terms of the original functions b and l. Note that $\bar{y}^1$ maximizes $\bar{l}(y)$ 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 $l(x)$ 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,

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

Next, observe from (6) that

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

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 computing $\tilde{x}(a^1, \ldots, a^k)$.

To compute $l^{*(1)}(y^1)$, we use the definition of $l^*(y^1)$ involving $\bar{l}$. Then, it follows that

$$l^{*(1)}(y^1) = \bar{l}_i\{\tilde{y}(y^1, a^2, \ldots, a^k)\}\, \tilde{y}_1^i(y^1, a^2, \ldots, a^k),$$

where $\tilde{y}_1^i(y^1, \ldots, y^k) = \partial \tilde{y}^i / \partial y^1$ and i runs from 1 to p. However, $\tilde{y}_1^\alpha(y^1, a^2, \ldots, a^k)$ equals 1 for $\alpha = 1$ and zero for $\alpha = 2, \ldots, k$. Moreover, $\bar{l}_{i'}\{\tilde{y}(y^1, a^2, \ldots, a^k)\} = 0$ for $i' = k+1, \ldots, p$ since $\tilde{y}(y^1, a^2, \ldots, a^k)$ maximizes $\bar{l}$ subject to the first k components of y being held fixed at the values $y^1, a^2, \ldots, a^k$, respectively. Hence,

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

The Lagrangian for maximizing $\bar{l}(y)$ subject to the first k components of y being held fixed is $\bar{H}(y) = \bar{l}(y) + \lambda_\alpha y^\alpha$, where α runs from 1 to k. The Lagrange multipliers $\lambda_\alpha(y^1, a^2, \ldots, a^k)$ $(\alpha = 1, \ldots, k)$ are the same as those for maximizing $l(x)$ subject to $g^1(x) = y^1$, $g^2(x) = a^2, \ldots, g^k(x) = a^k$. Then,

$$\lambda_\alpha(y^1, a^2, \ldots, a^k) = -\bar{l}_\alpha\{\tilde{y}(y^1, a^2, \ldots, a^k)\} \quad (\alpha = 1, \ldots, k).$$

In particular,

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

The second derivative $l^{*(2)}(y^1)$ cannot be expressed in closed form in general. However, it is easy to compute accurately through the use of numerical derivatives, or by methods such as those proposed by Fraser, Reid & Wong (1991).

References

Barndorff-Nielsen, O. E. & Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. 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. (1993). Analytical approximations to bootstrap distribution functions using saddlepoint methods. *Statist. Sinica.* To appear.

Fraser, D. A. S., Reid, N. & Wong, A. (1991). Exponential linear models: A two-pass procedure for saddlepoint approximations. *J. R. Statist. Soc.* B **53**, 483–92.

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

Phillips, P. C. B. (1983). Marginal densities of instrumental variables estimators in the general single equation case. *Adv. Economet.* **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. Prob.* **24**, 875–87.

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

[*Received April* 1992. *Revised June* 1993]