Biometrika (1991), 78, 1, *pp.* 169–79
Printed in Great Britain

Saddlepoint approximation for the studentized mean, with an application to the bootstrap

BY H. E. DANIELS AND G. A. YOUNG
Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K.

SUMMARY

A saddlepoint technique is used to approximate to the density and tail probability of the studentized mean of a random sample. The motivation was to replace bootstrapping of the studentized mean in the way Davison & Hinkley (1988) used the saddlepoint approximation for the unstudentized mean. The method involves first obtaining a bivariate saddlepoint approximation, then, after a nonlinear transformation, integrating out an unwanted variable either numerically or by a Laplace approximation. The tail probability is similarly evaluated either by a further numerical integration or by a Laplace approximation of the Temme type. Two difficulties arise.

(i) The nonlinearity of the transformation may result in Laplace approximations failing in the tail when the sample is not large. But numerical integration always works.

(ii) In the bootstrap application the saddlepoint approximation may itself break down when the data set contains an outlier.

Some key words: Bootstrap; Nonlinear transformation; Saddlepoint approximation; Studentized mean.

1. INTRODUCTION

The purpose of this paper is to devise a saddlepoint technique for approximating to the density and tail probability of the studentized mean of a random sample. Difficulties arise from the fact that a nonlinear transformation is involved and it is convenient to consider first the general problem of a nonlinear transformation of which this is a special case. Our motivation was to use such a technique to replace resampling in the way Davison & Hinkley (1988) used the saddlepoint approximation in the case of an unstudentized mean. For remarks concerning the value of studentization in the bootstrap context, see, for example, Hall (1988). We discuss that application in the final section of the paper.

The method of finding the density involves first obtaining a bivariate saddlepoint approximation, then integrating out an unwanted variable which can be performed either numerically or by using a Laplace approximation, as was done in a different context by Tierney, Kass & Kadane (1989). The tail probability can similarly be evaluated by further numerical integration or by a Laplace approximation of the Temme type (Barndorff-Nielsen & Cox, 1989, p. 82). Similar approximations have been considered by DiCiccio, Field & Fraser (1990) and others.

We find that for some distributions the nonlinearity of the transformation may result in Laplace approximations being unacceptably inaccurate in the tail when n is not large. A similar point was made by Leonard, Hsu & Tsui (1989). In practice, direct numerical integration is probably the safest procedure to use, and little accuracy is lost after renormalization.

In the bootstrap application we encounter the further difficulty that the saddlepoint approximation itself may break down when the data set contains extreme outliers.

2. THE STUDENTIZED MEAN

We wish to approximate to the distribution of $\lambda = \bar{x}/s$, where $\bar{x} = n^{-1}\Sigma x_j$, $s^2 = n^{-1}\Sigma (x_j - \bar{x})^2$ and $x_1, \ldots, x_n$ is a random sample from some distribution $F(x)$. For economy of notation random variables and observations will not always be distinguished; the interpretation is never ambiguous. The divisor n rather than $n-1$ has been used for s^2 to simplify the formulae, so that Student's t is related to λ by $t = \lambda\sqrt{(n-1)}$. Since we do not necessarily assume $E(x) = 0$ this could be noncentral t.

Assume that the density $f(x)$ dies away at least as fast as a normal density in the tails. This will certainly be true in the bootstrap application but is otherwise a rather severe restriction.

The joint moment generating function of X and X^2 is

$$M(T, U) = E\{\exp(TX + UX^2)\}.$$

Then

$$M^n(T, U) = E[\exp\{nT\bar{x} + nU(s^2 + \bar{x}^2)\}] = E[\exp\{nTs\lambda + nUs^2(1+\lambda^2)\}] \quad (2{\cdot}1)$$

from which a bivariate saddlepoint approximation can be found for the joint density of $s\lambda$ and $s^2(1+\lambda^2)$. A transformation then gives the joint density of λ and s. To approximate to the marginal density of λ one has either to integrate out s numerically or to find a Laplace approximation to replace the integration, which raises certain difficulties. Similar considerations arise for the marginal distribution of s.

3. THE GENERAL PROBLEM

Suppose there are two random variables X, Y whose joint moment generating function is

$$M(T, U) = \exp\{K(T, U)\} = E\{\exp(TX + UY)\} \quad (3{\cdot}1)$$

and $E(X) = \xi$, $E(Y) = \eta$.

The means $\bar{x}, \bar{y}$ from a sample of n observations are related to new quantities a, b by a transformation $\bar{x} = \bar{x}(a, b)$, $\bar{y} = \bar{y}(a, b)$ which is $(1-1)$ but is in general nonlinear. In our case $Y = X^2$, $a = \lambda$, $b = s$. The density of the means is

$$f_n(\bar{x}, \bar{y}) = \frac{n^2}{(2\pi i)^2} \int\int e^{n\{K(T,U) - T\bar{x} - U\bar{y}\}}\, dT\, dU, \quad (3{\cdot}2)$$

integration being along admissible paths in the T and U planes. The usual saddlepoint approximation is

$$\hat{f}_n(\bar{x}, \bar{y}) = \frac{n}{2\pi} \frac{e^{n\{K(\hat{T},\hat{U}) - \hat{T}\bar{x} - \hat{U}\bar{y}\}}}{(\hat{K}_{TT}\hat{K}_{UU} - \hat{K}^2_{TU})^{\frac{1}{2}}}, \quad (3{\cdot}3)$$

where

$$\hat{K}_T \equiv K_T(\hat{T}, \hat{U}) = \bar{x}, \quad \hat{K}_U \equiv K_U(\hat{T}, \hat{U}) = \bar{y}, \quad (3{\cdot}4)$$

K_T, K_U etc. denoting derivatives. See, for example, Barndorff-Nielsen & Cox (1979). The relative error is $O(n^{-1})$. Then the approximate joint density of a, b is

$$\hat{g}_n(a,b)=\frac{\partial(\bar{x},\bar{y})}{\partial(a,b)}\hat{f}_n(\bar{x},\bar{y})=\frac{n}{2\pi}\frac{\partial(\bar{x},\bar{y})}{\partial(a,b)}|\hat{\Delta}|^{-\frac{1}{2}}e^{-n\hat{\Lambda}(a,b)}, \tag{3·5}$$

where

$$\hat{\Lambda}\equiv\hat{\Lambda}(a,b)=\hat{T}\bar{x}+\hat{U}\bar{y}-K(\hat{T},\hat{U})\geqslant 0, \tag{3·6}$$

$$\hat{\Delta}\equiv\hat{\Delta}(a,b)=\begin{bmatrix}\hat{K}_{TT} & \hat{K}_{TU}\\ \hat{K}_{TU} & \hat{K}_{UU}\end{bmatrix}. \tag{3·7}$$

The order of a, b is conventionally chosen to make the Jacobian positive.

The approximate marginal density $\hat{h}_n(a)$ of a can be found by integrating out b numerically over its range. It would be desirable to replace this integration by a Laplace approximation in the usual way, but care is needed because although $\hat{\Lambda}$ is always convex as a function of $(\bar{x},\bar{y})$ it may not be in terms of (a,b) for a nonlinear transformation.

Similarly, while an approximation to the tail probability can always be found by a further numerical integration one would like to replace it by a Temme approximation, but for the same reason it is necessary to proceed with caution.

The maximum of $\hat{g}_n(a,b)$ occurs, to $O(n^{-1})$, at (α,β), where $\hat{\Lambda}_\alpha=\hat{\Lambda}_\beta=0$. From (3·6)

$$\hat{\Lambda}_a=\hat{T}\frac{\partial\bar{x}}{\partial a}+\hat{U}\frac{\partial\bar{y}}{\partial a},\quad \hat{\Lambda}_b=\hat{T}\frac{\partial\bar{x}}{\partial b}+\hat{U}\frac{\partial\bar{y}}{\partial b}. \tag{3·8}$$

These vanish when $\hat{T}=0$, $\hat{U}=0$, in which case $\bar{x}(\alpha,\beta)=\xi$, $\bar{y}(\alpha,\beta)=\eta$. Also

$$\hat{\Lambda}_{aa}=\hat{T}\frac{\partial^2\bar{x}}{\partial a^2}+\hat{U}\frac{\partial^2\bar{y}}{\partial a^2}+\frac{\partial\hat{T}}{\partial a}\frac{\partial\bar{x}}{\partial a}+\frac{\partial\hat{U}}{\partial a}\frac{\partial\bar{y}}{\partial a}, \tag{3·9}$$

with similar formulae for $\hat{\Lambda}_{ab}$, $\hat{\Lambda}_{bb}$. Differentiating (3·4) leads to

$$\frac{\partial\hat{T}}{\partial a}=\frac{\hat{K}_{UU}\partial\bar{x}/\partial a-\hat{K}_{TU}\partial\bar{y}/\partial a}{|\hat{\Delta}|},\quad \frac{\partial\hat{U}}{\partial a}=\frac{-\hat{K}_{TU}\partial\bar{x}/\partial a+\hat{K}_{TT}\partial\bar{y}/\partial a}{|\hat{\Delta}|}. \tag{3·10}$$

Then

$$\frac{\partial\hat{T}}{\partial a}\frac{\partial\bar{x}}{\partial a}+\frac{\partial\hat{U}}{\partial a}\frac{\partial\bar{y}}{\partial a}=\frac{\hat{K}_{UU}(\partial\bar{x}/\partial a)^2-2\hat{K}_{TU}\partial\bar{x}/\partial a\partial\bar{y}/\partial a+\hat{K}_{TT}(\partial\bar{y}/\partial a)^2}{|\hat{\Delta}|} \tag{3·11}$$

$$=D_a'\hat{\Delta}^{-1}D_a,$$

where

$$D_a'=\left[\frac{\partial\bar{x}}{\partial a},\frac{\partial\bar{y}}{\partial a}\right],\quad D_b'=\left[\frac{\partial\bar{x}}{\partial b},\frac{\partial\bar{y}}{\partial b}\right]. \tag{3·12}$$

The other corresponding quantities are $D_a'\hat{\Delta}^{-1}D_b$ and $D_b'\hat{\Delta}^{-1}D_b$. At the maximum (α,β) the first two terms in (3·9) vanish, and

$$\begin{bmatrix}\hat{\Lambda}_{\alpha\alpha} & \hat{\Lambda}_{\alpha\beta}\\ \hat{\Lambda}_{\alpha\beta} & \hat{\Lambda}_{\beta\beta}\end{bmatrix}=\begin{bmatrix}D_\alpha'\hat{\Delta}^{-1}D_\alpha & D_\alpha'\hat{\Delta}^{-1}D_\beta\\ D_\beta'\hat{\Delta}^{-1}D_\alpha & D_\beta'\hat{\Delta}^{-1}D_\beta\end{bmatrix}=\begin{bmatrix}D_\alpha'\\ D_\beta'\end{bmatrix}\hat{\Delta}^{-1}[D_\alpha,D_\beta]. \tag{3·13}$$

Since $|[D_\alpha,D_\beta]|=\partial(\bar{x},\bar{y})/\partial(\alpha,\beta)$ it follows that

$$\hat{\Lambda}_{\alpha\alpha}\hat{\Lambda}_{\beta\beta}-\hat{\Lambda}_{\alpha\beta}^2=\left\{\frac{\partial(\bar{x},\bar{y})}{\partial(\alpha,\beta)}\right\}^2\Big/|\hat{\Delta}|>0. \tag{3·14}$$

So $\hat{\Lambda}(a,b)$ remains convex in the vicinity of (α,β). However, for moderate values of n nonconvexity outside this region may affect the approximations.

4. LAPLACE APPROXIMATIONS

To approximate to the integration over b we need $b_0 = b_0(a)$ such that $\hat{\Lambda}_b(a, b_0) = 0$ and $\hat{\Lambda}_{bb}(a, b_0) > 0$. Assuming the latter to be satisfied, the Laplace approximation to the marginal density of a is

$$\hat{h}_n(a) = \sqrt{\left(\frac{n}{2\pi}\right)} \frac{\partial(\bar{x}, \bar{y})}{\partial(a, b_0)} \frac{e^{-n\hat{\Lambda}(a, b_0)}}{|\hat{\Delta}(a, b_0)|^{\frac{1}{2}} \hat{\Lambda}_{bb}^{\frac{1}{2}}(a, b_0)} \tag{4·1}$$

$$= \sqrt{\left(\frac{n}{2\pi}\right)} \frac{\partial(\bar{x}, \bar{y})}{\partial(a, b_0)} G^{-\frac{1}{2}}(a, b_0)\, e^{-n\hat{\Lambda}(a, b_0)}, \tag{4·2}$$

where

$$\hat{G}(a, b_0) = |\Delta(a, b_0)| \left(\hat{T}_0 \frac{\partial^2 \bar{x}}{\partial b_0^2} + \hat{U}_0 \frac{\partial^2 \bar{y}}{\partial b_0^2} \right) + \hat{K}_{UU} \left(\frac{\partial \bar{x}}{\partial b_0} \right)^2 - 2\hat{K}_{TU} \frac{\partial \bar{x}}{\partial b_0} \frac{\partial \bar{y}}{\partial b_0} + \hat{K}_{TT} \left(\frac{\partial \bar{y}}{\partial b_0} \right)^2 \tag{4·3}$$

and $\hat{T}_0$, $\hat{U}_0$, b_0 are determined from the three equations

$$\hat{T}_0 \frac{\partial \bar{x}}{\partial b_0} + \hat{U}_0 \frac{\partial \bar{y}}{\partial b_0} = 0, \quad K_T(\hat{T}_0, \hat{U}_0) = \bar{x}(a, b_0), \quad K_U(\hat{T}_0, \hat{U}_0) = \bar{y}(a, b_0). \tag{4·4}$$

These equations are easily solved using packaged iterative routines. Throughout the paper, numerical computations were performed using the NAG root finding algorithm CØ5NBF. The key to finding suitable starting values for the iteration is to observe that (4·4) is satisfied with $\hat{T}_0 = \hat{U}_0 = 0$, $b_0 = \beta$, at $a = \alpha$.

Under appropriate conditions the tail probability of a can be found from (4·2) by an approximation due to Temme (1982) which is analogous to that arrived at by Lugannani & Rice (1980) using complex variable methods. Maximizing $\hat{\Lambda}(a, b_0(a))$ with respect to a should give $a = \alpha$, $b_0(a) = \beta$ as in § 3. Differentiating $\hat{\Lambda}_b(a, b_0) = 0$ with respect to a we find

$$\frac{db_0}{da} = -\frac{\hat{\Lambda}_{ab}(a, b_0)}{\hat{\Lambda}_{bb}(a, b_0)}. \tag{4·5}$$

Also

$$\frac{d\hat{\Lambda}(a, b_0)}{da} = \hat{\Lambda}_a(a, b_0) + \hat{\Lambda}_b(a, b_0) \frac{db_0}{da} = \hat{\Lambda}_a(a, b_0), \tag{4·6}$$

$$\frac{d^2\hat{\Lambda}(a, b_0)}{da^2} = \hat{\Lambda}_{aa}(a, b_0) + \hat{\Lambda}_{ab}(a, b_0) \frac{db_0}{da} = \hat{\Lambda}_{aa}(a, b_0) - \frac{\{\hat{\Lambda}_{ab}(a, b_0)\}^2}{\hat{\Lambda}_{bb}(a, b_0)} = \hat{\Lambda}_{aa.b}(a, b_0). \tag{4·7}$$

The tail probability is

$$\hat{Q}_n(a_1) = \int_{a_1}^{\infty} \hat{h}_n(a)\, da = \int_{a_1}^{\infty} \sqrt{\left(\frac{n}{2\pi}\right)} \frac{\partial(\bar{x}, \bar{y})}{\partial(a, b_0)} \frac{e^{-n\hat{\Lambda}(a, b_0)}\, da}{|\hat{\Delta}(a, b_0)|^{\frac{1}{2}} \hat{\Lambda}_{bb}^{\frac{1}{2}}(a, b_0)}, \tag{4·8}$$

which we further approximate by Temme's method (Barndorff-Nielsen & Cox, 1989, p. 82). Since $\hat{\Lambda}(\alpha, \beta) = 0$, let

$$w = \sqrt{\{2\hat{\Lambda}(a, b_0)\}}\, \mathrm{sgn}\,(a - \alpha). \tag{4·9}$$

Then

$$\frac{dw}{da} = \frac{\hat{\Lambda}_a(a, b_0)}{w} \tag{4·10}$$

and, expanding $\hat{\Lambda}$ about α in (4·9), we find

$$\left(\frac{dw}{da}\right)_{a=\alpha} = \hat{\Lambda}^{\frac{1}{2}}_{aa.b}(\alpha, \beta). \tag{4·11}$$

Then $\hat{Q}_n(a_1)$ transforms to

$$\hat{Q}_n(a_1) = \int_{w_1}^{\infty} \sqrt{\left(\frac{n}{2\pi}\right)} \psi(a, b_0)\, \mathrm{e}^{-\frac{1}{2}nw^2}\, dw, \tag{4·12}$$

where

$$\psi(a, b_0) = \frac{\partial(\bar{x}, \bar{y})}{\partial(a, b_0)} \frac{da/dw}{|\hat{\Delta}(a, b_0)|^{\frac{1}{2}} \hat{\Lambda}^{\frac{1}{2}}_{bb}(a, b_0)} \tag{4·13}$$

and $w_1 = w(a_1)$.

Temme's device is to replace $\psi(a, b_0)$ by $\psi(\alpha, \beta) - \{\psi(\alpha, \beta) - \psi(a, b_0)\}$ and then integrate by parts. From (3·14) and (4·11) it is found that $\psi(\alpha, \beta) = 1$, and using (4·10) we can rearrange (4·12) as

$$\hat{Q}_n(a_1) = 1 - \Phi(w_1\sqrt{n}) - \sqrt{\left(\frac{n}{2\pi}\right)} \int_{w_1}^{\infty} e^{-\frac{1}{2}nw^2}\, w\, dw \left\{\frac{1}{w} - \frac{\partial(\bar{x}, \bar{y})/\partial(a, b_0)}{\hat{\Lambda}_a(a, b_0)|\hat{\Delta}(a, b_0)|^{\frac{1}{2}} \hat{\Lambda}^{\frac{1}{2}}_{bb}(a, b_0)}\right\}. \tag{4·14}$$

On integrating by parts and replacing a_1, w_1 by a, w in the final result we obtain

$$\hat{Q}_n(a) = 1 - \Phi(w\sqrt{n}) - \frac{\phi(w\sqrt{n})}{\sqrt{n}} \left\{\frac{1}{w} - \frac{\partial(\bar{x}, \bar{y})/\partial(a, b_0)}{\hat{\Lambda}_a(a, b_0)|\hat{\Delta}(a, b_0)|^{\frac{1}{2}} \hat{\Lambda}^{\frac{1}{2}}_{bb}(a, b_0)}\right\}, \tag{4·15}$$

the remainder being incorporated in the error $O(n^{-\frac{3}{2}})$. Here $\Phi(.)$, $\phi(.)$ are the usual standard normal distribution function and density, and w is defined by (4·9).

5. Marginal distributions of $\lambda = \bar{x}/s$ and s

For this application, we first take $a = \lambda$, $b = s$

$$\bar{x} = s\lambda, \quad \bar{y} = s^2(1 + \lambda^2), \quad \frac{\partial(\bar{x}, \bar{y})}{\partial(\lambda, s)} = 2s^2$$

to find the marginal distribution of λ. The marginal distribution of s can be found from the same formulae by taking $a = s$, $b = \lambda$.

A good illustration of the care needed when using Laplace approximation is provided by the example of a mixture of normal distributions with

$$F(x) = \tfrac{1}{2}\{\Phi(x + c) + \Phi(x - c)\}. \tag{5·1}$$

Here

$$K(T, U) = -\tfrac{1}{2}c^2 + \frac{T^2 + c^2}{2(1 - 2U)} + \log\cosh\left(\frac{cT}{1 - 2U}\right) - \tfrac{1}{2}\log(1 - 2U). \tag{5·2}$$

When $c = 0$ the distribution is $N(0, 1)$ and the joint density of λ and s is approximated by

$$\hat{g}_n(\lambda, s) = As^{n-1}\, e^{-\frac{1}{2}ns^2(1+\lambda^2)}. \tag{5·3}$$

The approximate marginal densities of λ and s are found to be

$$\hat{h}_n(\lambda) = B(1+\lambda^2)^{-\frac{1}{2}n}, \quad \hat{k}_n(s) = Cs^{n-2}\, e^{-\frac{1}{2}ns^2}, \tag{5·4}$$

where A, B, C are Stirling approximations to the normalizing constants. In this case all three are exact after renormalization and both $\hat{h}_n(\lambda)$ and $\hat{k}_n(s)$ can be obtained either by direct integration of $\hat{g}_n(\lambda, s)$ or by Laplace approximations. Moreover Temme approximations to the tail probabilities of λ and s will be found to agree well with the values obtained by direct integration.

However, as c increases from zero the situation changes. Figure 1 shows contours of $\log \hat{g}_n(\lambda, s)$ for $c = 0, 1, 2$ and $n = 10$. The distribution is centred around the point $\lambda = 0$, $s = \sqrt{(1+c^2)}$ at which $\hat{\Lambda}_\lambda = 0$, $\hat{\Lambda}_s = 0$. When $c = 0$ the maxima required for the Laplace approximations are unique for all λ and s. But when $c = 2$ this is no longer true when s

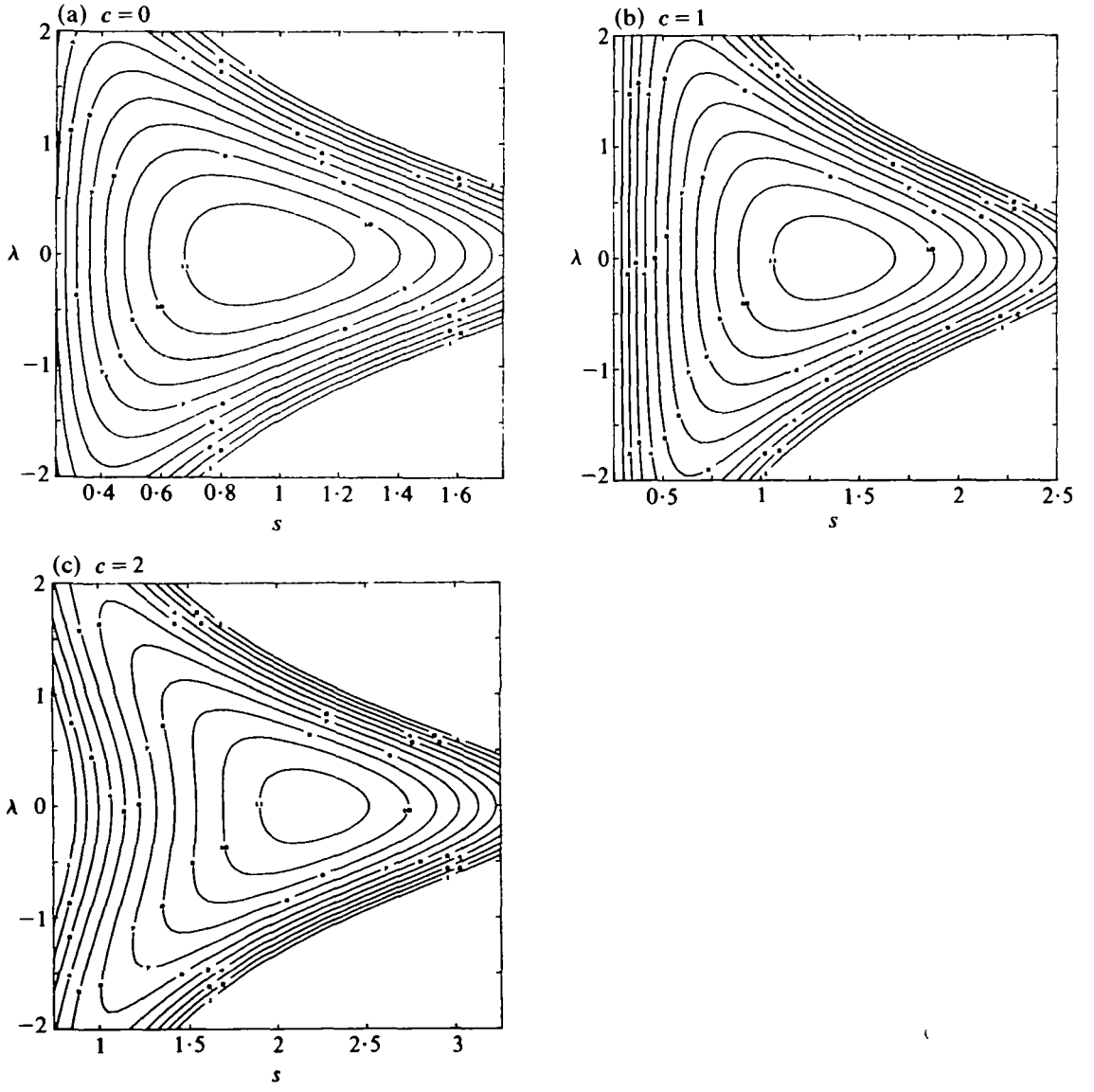

Fig. 1. Contours of $\log \hat{g}_n(\lambda, s)$ for normal mixture distribution, $c = 0, 1, 2$, sample size $n = 10$. In each plot there are 11 contours with equally spaced heights between 0·0 and −10·0.

is small enough, and integration of λ for such values of s cannot be replaced by a simple Laplace approximation. On the other hand, integration of s for given λ to find the marginal density of λ can always be replaced by a Laplace approximation.

These facts are borne out by the calculation shown in Tables 1 and 2, in which the density and tail probability of λ and s obtained by various methods for $n = 10$, $c = 1, 2$ are compared with the 'exact' value computed from 5×10^6 simulations. Here, as in all the tables of the paper, simulated density figures were obtained by numerical differentiation of the simulated tail probability figures, using a 5-point Lagrange formula on a grid with intervals of 0·05: this procedure is equivalent to a 5-point kernel smoothing of the corresponding histogram of simulated values of the statistic. The simulation sizes are such that the fourth significant figure in any simulated value is unreliable. In the case of λ it will be seen that the simulated values of the density $\hat{h}_n(\lambda)$ are well matched by those obtained by both numerical integration of s and the use of a Laplace approximation. The tail probability $\hat{Q}_n(\lambda)$ obtained from a Temme approximation agrees remarkably well with the simulated values.

In the case of s, however, while the values of the density $\hat{k}_n(s)$ obtained by numerical integration of λ agree well with the simulated values, those obtained from a Laplace approximation become progressively worse with decreasing s, as would be expected from the contours in Fig. 1. As s approaches a critical value the quadratic approximation near the maximum with respect to λ for given s fails and a quartic approximation becomes appropriate. In the case $c = 2$, with decreasing s this ultimately separates into two

Table 1. *Normal mixture distribution,* $c = 1$. *Simulated figures based on* 5 000 000 *samples*

(*a*) *Tail probabilities and densities of* λ. *Laplace and integrated densities renormalized*

λ	$\hat{Q}_n(\lambda)$ Simulated	$\hat{Q}_n(\lambda)$ Temme	$\hat{Q}_n(\lambda)$ Integrated	$\hat{h}_n(\lambda)$ Simulated	$\hat{h}_n(\lambda)$ Laplace	$\hat{h}_n(\lambda)$ Integrated
0·1	0·3842	0·3838	0·3847	1·1249	1·1173	1·1162
0·3	0·1937	0·1931	0·1949	0·7452	0·7515	0·7513
0·5	0·0834	0·0829	0·0844	0·3699	0·3736	0·3739
0·7	0·0334	0·0330	0·0339	0·1567	0·1561	0·1564
0·9	0·0132	0·0130	0·0135	0·0603	0·0610	0·0612
1·1	0·0054	0·0053	0·0055	0·0231	0·0238	0·0239
1·3	0·0023	0·0023	0·0024	0·0091	0·0096	0·0097
1·5	0·0011	0·0010	0·0011	0·0044	0·0041	0·0041
1·7	0·0005	0·0005	0·0005	0·0019	0·0018	0·0018
1·9	0·0002	0·0002	0·0002	0·0009	0·0008	0·0009

(*b*) *Tail probabilities and densities of s. Integrated densities renormalized, Laplace densities not renormalized*

s	$\hat{Q}_n(s)$ Simulated	$\hat{Q}_n(s)$ Integrated	$\hat{k}_n(s)$ Simulated	$\hat{k}_n(s)$ Integrated	$\hat{k}_n(s)$ Laplace
0·5	0·9993	0·9993	0·0112	0·0116	0·0152
0·7	0·9894	0·9912	0·1176	0·1184	0·1409
0·9	0·9338	0·9345	0·4990	0·5030	0·5775
1·1	0·7736	0·7729	1·1084	1·1167	1·2604
1·3	0·5103	0·5091	1·4240	1·4275	1·5958
1·5	0·2480	0·2475	1·1031	1·1038	1·2265
1·7	0·0848	0·0847	0·5367	0·5312	0·5879
1·9	0·0199	0·0199	0·1634	0·1620	0·1788
2·1	0·0031	0·0032	0·0323	0·0317	0·0349
2·3	0·0003	0·0003	0·0040	0·0040	0·0044

Table 2. *Normal mixture distribution, $c = 2$. Simulated figures based on 5000 000 samples*

(*a*) *Tail probabilities and densities of* λ. *Laplace and integrated densities renormalized*

λ	$\hat{Q}_n(\lambda)$ Simulated	$\hat{Q}_n(\lambda)$ Temme	$\hat{Q}_n(\lambda)$ Integrated	$\hat{h}_n(\lambda)$ Simulated	$\hat{h}_n(\lambda)$ Laplace	$\hat{h}_n(\lambda)$ Integrated
0·1	0·3816	0·3815	0·3825	1·1447	1·1347	1·1333
0·3	0·1897	0·1896	0·1918	0·7469	0·7441	0·7436
0·5	0·0820	0·0819	0·0839	0·3594	0·3596	0·3597
0·7	0·0342	0·0341	0·0355	0·1477	0·1496	0·1498
0·9	0·0149	0·0148	0·0157	0·0589	0·0606	0·0608
1·1	0·0076	0·0070	0·0075	0·0253	0·0257	0·0259
1·3	0·0036	0·0036	0·0039	0·0116	0·0118	0·0119
1·5	0·0021	0·0021	0·0022	0·0052	0·0059	0·0060
1·7	0·0013	0·0013	0·0013	0·0027	0·0032	0·0033
1·9	0·0009	0·0008	0·0008	0·0016	0·0019	0·0019
2·1	0·0006	0·0006	0·0005	0·0012	0·0012	0·0012
2·3	0·0005	0·0004	0·0003	0·0008	0·0008	0·0008

(*b*) *Tail probabilities and densities of s. Integrated densities renormalized, Laplace densities not renormalized*

s	$\hat{Q}_n(s)$ Simulated	$\hat{Q}_n(s)$ Integrated	$\hat{k}_n(s)$ Simulated	$\hat{k}_n(s)$ Integrated	$\hat{k}_n(s)$ Laplace
0·7	0·9997	0·9997	0·0028	0·0028	0·0015
0·9	0·9987	0·9987	0·0071	0·0079	0·0039
1·1	0·9962	0·9958	0·0210	0·0234	0·0109
1·3	0·9879	0·9869	0·0725	0·0756	0·0348
1·5	0·9606	0·9587	0·2245	0·2289	0·1311
1·7	0·8860	0·8829	0·5544	0·5584	1·0094
1·9	0·7315	0·7275	0·9909	0·9941	1·4790
2·1	0·5028	0·4995	1·2310	1·2278	1·5105
2·3	0·2691	0·2675	1·0324	1·0275	1·1852
2·5	0·1068	0·1062	0·5806	0·5761	0·6450
2·7	0·0304	0·0302	0·2185	0·2153	0·2372
2·9	0·0060	0·0060	0·0542	0·0535	0·0584
3·1	0·0008	0·0008	0·0088	0·0089	0·0096
3·3	0·0001	0·0001	0·0009	0·0010	0·0011

symmetric maxima on either side of $\lambda = 0$, one of which the Laplace computer program detects and uses to produce half the correct value. When $c = 1$ the effect is less pronounced but still serious.

The danger can only be recognized by examining the behaviour of the density of λ near the maximum for each value of s. The safest procedure would seem to be to integrate λ numerically. Computationally such numerical integration is most conveniently performed using direct function evaluation on a regular grid, for then the numerical solution of (3·4) proceeds sequentially.

6. Application to the bootstrap

We now apply the present procedure in an attempt to replace bootstrap sampling of the studentized mean from a data set, in a way similar to Davison & Hinkley's (1988) use of a saddlepoint approximation for unstudentized means. How far the confidence intervals so obtained relate to the underlying distribution is a separate important issue

which is not addressed here; the simpler situation of an unstudentized mean has been discussed by us elsewhere (Young & Daniels, 1990).

We have a data set $x_1, \ldots, x_n$ from an unknown distribution $F(x)$ with mean μ and variance σ^2. Let $\bar{x} = n^{-1} \Sigma x_j$, $s_0^2 = n^{-1} \Sigma (x_j - \bar{x})^2$, and let $Y_1, \ldots, Y_n$ be a random sample from the empirical distribution function $\hat{F}_n(y)$ which has probability $1/n$ at each of $x_1, \ldots, x_n$, so that $E(Y_i) = \bar{x}$, $\text{var}(Y_i) = s^2$. The motivation for this application is the assumption that the sampling behaviour of $\bar{X} - \mu$ can be inferred from that of $\bar{Z} = \bar{Y} - \bar{x}$, where $z_j = y_j - \bar{x}$. Now $s^2 = n^{-1} \Sigma (Y_j - \bar{Y})^2$ and the required moment generating function is

$$M(T, U) = E\{\exp(TZ + UZ^2)\} = \frac{1}{n} \sum_{j=1}^{n} e^{T(x_j - \bar{x}) + U(x_j - \bar{x})^2},$$

from which $K(T, U) = \log M(T, U)$ and its derivatives can be computed.

Davison & Hinkley's data ($n = 10$), recentred to have zero mean, are $-8{\cdot}27$, $-7{\cdot}47$, $-4{\cdot}87$, $-2{\cdot}87$, $-1{\cdot}27$, $-0{\cdot}67$, $-0{\cdot}57$, $3{\cdot}93$, $6{\cdot}13$, $15{\cdot}93$. For this data set the simulated densities and tail probabilities of $\lambda = \bar{z}/s$ and s are compared with the various approximations in Table 3. As regards λ the agreement is quite good, though not as close as Davison & Hinkley's results for the unstudentized mean. However, the simulated density for s differs markedly from the saddlepoint densities computed either by numerical integration or by a Laplace approximation, which are themselves in good agreement. Notice that it appears to have a periodic variation superimposed on the values obtained by the other methods. This is related to the presence of the outlying observation $15{\cdot}93$ in the data set in a way we explain below, and is also probably responsible for the observed slight instability in the simulated density of λ.

What we are coming up against here is the limitation of the saddlepoint approximation itself when applied to an empirical distribution and hence to the bootstrap. Davison & Hinkley found it worked well for the unstudentized mean even with such an outlier in the data set, but calculation of s involves squaring the observations and the effect of the outlier is magnified.

Data sets with more extreme outliers than Davison & Hinkley's are found to cause similar trouble even with the unstudentized mean. To illustrate the way an outlier generates periodic variation in the density of the unstudentized mean, consider the following more extreme data set constructed by squaring Davison & Hinkley's $x_j - \bar{x}$, recentring at their mean and dividing by 100: $-0{\cdot}4621$, $-0{\cdot}4608$, $-0{\cdot}4492$, $-0{\cdot}3830$, $-0{\cdot}3109$, $-0{\cdot}2282$, $-0{\cdot}0896$, $0{\cdot}0927$, $0{\cdot}2183$, $2{\cdot}0723$. Table 4 compares the simulated and saddlepoint densities $\hat{f}_n(\bar{z})$ of $\bar{z}$ using this data set. There is a strong periodic variation of the simulated density about the saddlepoint values, the period being about $0{\cdot}25$.

The explanation is as follows. The data consist of nine observations closely clustered about their mean $-0{\cdot}2303$ and a widely separated outlier $2{\cdot}0723$. Consider the even more extreme data set consisting of $-0{\cdot}2303$ repeated nine times, and $2{\cdot}0723$. This is a two-point distribution with probability $\frac{9}{10}$ at $-0{\cdot}2303$ and $\frac{1}{10}$ at $2{\cdot}0723$. Then we are simulating samples of 10 centred Bernoulli variables whose means $\bar{z}$ take values on the lattice

$$-0{\cdot}2303, 0, 0{\cdot}2303, 0{\cdot}4606, \ldots, 1{\cdot}8420, 2{\cdot}0723$$

with binomial probabilities. In the actual data set the small variation of the first nine observations around $-0{\cdot}2303$ evens out these binomial 'spikes' into fluctuations with a period related to the lattice interval. In situations like this the saddlepoint approximation produces a smoothed version of the exact density.

Table 3. *Bootstrap example. Simulated figures based on* 1000 000 *samples. Laplace and integrated densities are renormalized*

(*a*) *Tail probabilities and densities of* λ

λ	$\hat{Q}_n(\lambda)$ Simulated	$\hat{Q}_n(\lambda)$ Temme	$\hat{Q}_n(\lambda)$ Integrated	$\hat{h}_n(\lambda)$ Simulated	$\hat{h}_n(\lambda)$ Integrated	$\hat{h}_n(\lambda)$ Laplace
−1·8	0·9978	0·9978	0·9983	0·0083	0·0071	0·0072
−1·6	0·9953	0·9954	0·9961	0·0182	0·0155	0·0150
−1·4	0·9900	0·9907	0·9915	0·0367	0·0311	0·0294
−1·2	0·9807	0·9820	0·9830	0·0576	0·0560	0·0554
−1·0	0·9661	0·9657	0·9680	0·0972	0·0970	0·1014
−0.8	0·9390	0·9383	0·9417	0·1821	0·1715	0·1761
−0·6	0·8901	0·8951	0·8950	0·3087	0·3064	0·3011
−0·4	0·8136	0·8203	0·8118	0·4777	0·5414	0·5270
−0·2	0·6837	0·6865	0·6700	0·8482	0·8866	0·8769
0·0	0·4769	0·4715	0·4624	1·1561	1·1589	1·1650
0·2	0·2490	0·2482	0·2402	1·0339	0·9993	1·0121
0·4	0·0910	0·0896	0·0884	0·5385	0·5129	0·5187
0·6	0·0245	0·0235	0·0239	0·1740	0·1659	0·1658
0·8	0·0053	0·0050	0·0053	0·0426	0·0396	0·0386
1·0	0·0011	0·0010	0·0011	0·0083	0·0082	0·0077

(*b*) *Densities of s*

s	$\hat{k}_n(s)$ Simulated	$\hat{k}_n(s)$ Integrated	$\hat{k}_n(s)$ Laplace
1·0	0·0002	0·0003	0·0002
2·0	0·0057	0·0055	0·0045
3·0	0·0596	0·0580	0·0509
4·0	0·1592	0·1247	0·1320
5·0	0·1187	0·1571	0·1609
6·0	0·1338	0·2000	0·2004
7·0	0·2852	0·2128	0·2107
8·0	0·1447	0·1601	0·1587
9·0	0·0740	0·0684	0·0688
10·0	0·0135	0·0122	0·0124
11·0	0·0005	0·0005	0·0006

Table 4. *New data set. Density of mean* $\bar{z}$ *of bootstrap sample of size* $n = 10$. *Simulated figures based on* 1000 000 *bootstrap samples. Saddlepoint figures computed using method of Davison & Hinkley* (1988)

$\bar{z}$	$\hat{f}_n(\bar{z})$ Simulated	$\hat{f}_n(\bar{z})$ Saddlepoint	$\bar{z}$	$\hat{f}_n(\bar{z})$ Simulated	$\hat{f}_n(\bar{z})$ Saddlepoint
−0·40	0·1275	0·0340	0·10	0·9845	1·5665
−0·35	0·5739	0·1769	0·15	0·9207	1·4059
−0·30	1·3050	0·5255	0·20	1·1088	1·2231
−0·25	1·7826	0·9687	0·25	1·0303	1·0337
−0·20	1·6081	1·2860	0·30	0·6494	0·8502
−0·15	1·2217	1·5088	0·35	0·3419	0·6816
−0·10	1·3538	1·6646	0·40	0·3156	0·5333
−0·05	1·9345	1·7467	0·45	0·3556	0·4076
0·00	2·1082	1·7522	0·50	0·2756	0·3047
0·05	1·5775	1·6876	0·55	0·1348	0·2228

In the case of the studentized mean the trouble would be to some extent diminished if s were replaced by an estimator such as the mean deviation which does not involve squaring the observations. Simulation methods could still be used, but an analytic approach does not seem feasible.

References

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc.* B **41**, 279–312.

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* London: Chapman & Hall.

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

DICICCIO, T. J., FIELD, C. A. & FRASER, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77–95.

HALL, P. G. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16**, 927–85.

LEONARD, T., HSU, J. S. J. & TSUI, K.-W. (1989). Bayesian marginal inference. *J. Am. Statist. Assoc.* **84**, 1051–8.

LUGANNANI, R. & RICE, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12**, 475–90.

TEMME, N. M. (1982). The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.* **13**, 239–53.

TIERNEY, L., KASS, R. E. & KADANE, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**, 425–33.

YOUNG, G. A. & DANIELS, H. E. (1990). Bootstrap bias. *Biometrika* **77**, 179–85.

[*Received April* 1990. *Revised September* 1990]