Statistica Sinica 4(1994), 281-295

ANALYTICAL APPROXIMATIONS TO BOOTSTRAP DISTRIBUTION FUNCTIONS USING SADDLEPOINT METHODS

Thomas J. DiCiccio, Michael A. Martin and G. Alastair Young*

*Stanford University and University of Cambridge**

Abstract: We develop analytical approximations to bootstrap distribution functions of statistics that are smooth functions of vector means. In particular, our technique is applicable in the case of bootstrap inference for a population mean, for a Studentized mean, and for other more complex situations, such as inference for population variances or correlation coefficients. The approximations are based on the application of a tail probability approximation of DiCiccio and Martin (1991) to a saddlepoint approximation for the joint density of several means. Our method extends the work of Davison and Hinkley (1988), who proposed the use of saddlepoint methods to replace bootstrap resampling primarily in the case of linear statistics, and the work of Daniels and Young (1991), who considered the problem of inference for a Studentized mean. Our technique produces accurate approximations over the entire range of the distribution function and is easy to implement. It has two critical advantages over standard resampling techniques: it can yield significant computational savings; and it is more accurate than standard resampling approaches based on 5,000 or 10,000 resamples. We illustrate these points by applying our technique to estimate the bootstrap distribution of a bivariate correlation coefficient for both real and simulated data; and the method performs well in each case. Finally, we illustrate the power and flexibility of our technique in the very complex problem of estimating the bootstrap distribution of a Studentized, transformed correlation coefficient.

Key words and phrases: Asymptotic approximations, correlation coefficient, cumulant generating function, Fisher's z transformation, iterated bootstrap, moment generating function, normal approximation, pivotal statistic, resampling, simulation, Studentized statistic, tail probability approximations, vector means.

1. Introduction

Although bootstrap procedures often provide accurate inference in complex problems, they are typically very computationally intensive. Considerable attention has been focused recently on the development of analytical methods for approximating bootstrap distributions that do not require extensive and time-consuming simulations. Davison and Hinkley (1988) proposed the idea that saddlepoint approximations could be used to replace bootstrap resampling in the case of an unstudentized sample mean. They developed analytical approximations to bootstrap distribution functions that worked well when the statistic of interest was linear, or if there was identification with a linear estimating equation. However, their methods met with limited success when dealing with non-linear statistics. Daniels and Young (1991) extended the work of Davison and Hinkley to the problem of approximating the density and tail probabilities of a Studentized mean. Their method involves, first, development of a bivariate saddlepoint approximation, and then use of a non-linear transformation to derive the joint density of the Studentized mean and another statistic. The marginal density and tail probabilities of the Studentized mean are then approximated by successive numerical integrations or Laplace approximations. In principle, their method could be applied to more complicated statistics than the Studentized mean, but in practice it is very cumbersome to use, even in the relatively simple case of a correlation coefficient.

In this paper, we propose a general method of analytical approximation to bootstrap distribution functions for statistics that are expressible as smooth functions of vector means. The method is easy to apply and has two critical advantages over standard bootstrap resampling algorithms. First, bootstrap calculations based on large numbers of resamples are computationally burdensome, but our method can lead to substantial computational savings. Second, our method provides highly accurate approximations to the entire distribution function, even in the extreme tails. In particular, we demonstrate that our analytical approximation is closer (in either an L^2 or L^∞ sense) to an "exact" bootstrap distribution function evaluation than standard bootstrap distribution function estimates based on 5,000 or 10,000 resamples. Moreover, our method is flexible in that it can deal with problems ranging from the very simple (e.g. Studentized mean), to a little harder (e.g. correlation coefficient), to the very difficult (e.g. Studentized correlation coefficient).

The technique introduced here is useful whenever accurate approximations to distribution functions are required. In particular, the approximation leads to simple methods for constructing bootstrap and iterated bootstrap confidence intervals. The latter confidence intervals are of particular interest, as iterated bootstrap computations are often prohibitively expensive. DiCiccio, Martin and Young (1992a, b) investigated the use of analytical approximations, similar to those introduced here, to replace the inner level of resampling in an iterated bootstrap computation. It should be stressed here that our aim in this paper is quite distinct from that pursued in the problem of constructing accurate two-sided confidence intervals. The accuracy of distribution function approximations in the latter setting is judged solely by the extent to which they ultimately produce two-sided confidence intervals with good coverage properties. Indeed, it turns out in

DiCiccio, Martin and Young's (1992a, b) development of that methodology that *cruder* approximations than those discussed here suffice. Their approximations are related to those given here in that they correspond essentially to the leading term of our more accurate distribution function approximations, and they use only approximate solutions to the system of saddlepoint equations that arise in our development. In this paper, we present the full, accurate approximation to the bootstrap distribution function.

In Section 2, we give a theoretical development of the technique. Section 3 contains the results of some numerical investigations of the technique in the case of a correlation coefficient from two data sets: Efron's (1982) Law School data relating LSAT and GPA scores of students entering 15 American law schools, and a data set simulated from a bivariate log normal distribution. A simulation study was also conducted to assess the accuracy of our technique against that of standard bootstrap resampling algorithms. The power of our technique is demonstrated further in Section 4, by its application in the complex circumstance of a Studentized, transformed correlation coefficient. Some conclusions and remarks are presented in Section 5.

2. The Technique

Suppose we are interested in obtaining a bootstrap estimate of the sampling distribution of an estimator of a parameter θ that is expressible as a smooth function of vector means. Assume that the data consists of n independent and identically distributed observations of a d-dimensional random vector $X = (X_1, \ldots, X_d)$ and denote the jth observation of X by $(X_{1j}, \ldots, X_{dj})$, $j = 1, \ldots, n$. Let $f_1, \ldots, f_k$ be real-valued, measurable functions on $\mathbb{R}^d$ and define

$$Z_j = \left[f_1(X_{1j}, \ldots, X_{dj}), \ldots, f_k(X_{1j}, \ldots, X_{dj})\right] = [Z_{1j}, \ldots, Z_{kj}], \qquad j = 1, \ldots, n,$$

and

$$\overline{Z} = [\overline{Z}_1, \ldots, \overline{Z}_k] = n^{-1} \sum_{j=1}^{n} Z_j.$$

Denote by μ the mean vector $[E(\overline{Z}_1), \ldots, E(\overline{Z}_k)] = [E\{f_1(X)\}, \ldots, E\{f_k(X)\}]$. Let $\theta = g(\mu)$, where g is a real-valued function having continuous gradient that is non-zero at $\overline{Z}$, and suppose that θ is estimated by $\hat{\theta} = g(\overline{Z})$. Note that $\hat{\theta}$ could be a sample mean, a Studentized mean, or another more complex statistic, such as a sample variance or correlation coefficient. For instance, θ is the bivariate correlation coefficient and $\hat{\theta}$ is the bivariate sample correlation coefficient if one takes $d = 2$, $k = 5$, $f_1(X_{1j}, X_{2j}) = X_{1j}$, $f_2(X_{1j}, X_{2j}) = X_{2j}$, $f_3(X_{1j}, X_{2j}) = X_{1j}^2$,

$f_4(X_{1j}, X_{2j}) = X_{2j}^2$, $f_5(X_{1j}, X_{2j}) = X_{1j}X_{2j}$, and $g(\overline{Z}) = (\overline{Z}_5 - \overline{Z}_1\overline{Z}_2)\{(\overline{Z}_3 - \overline{Z}_1^2)(\overline{Z}_4 - \overline{Z}_2^2)\}^{-\frac{1}{2}}$.

Bootstrap inference for θ can be made by resampling observations with replacement from the original n observations of $(X_1, \ldots, X_d)$. Formally, n observations correspond to a d-dimensional random vector $X^* = (X_1^*, \ldots, X_d^*)$, where the jth observation of X^* is denoted by $(X_{1j}^*, \ldots, X_{dj}^*)$, $j = 1, \ldots, n$, and the resamples are chosen according to the rule

$$P\Big\{(X_{1l}^*, \ldots, X_{dl}^*) = (X_{1j}, \ldots, X_{dj}) | (X_{1m}, \ldots, X_{dm}),\quad m = 1, \ldots, n\Big\} = n^{-1}, \qquad l, j = 1, \ldots, n.$$

Put

$$Z_l^* = \Big[f_1(X_{1l}^*, \ldots, X_{dl}^*), \ldots, f_k(X_{1l}^*, \ldots, X_{dl}^*)\Big], \qquad l = 1, \ldots, n,$$

and

$$\overline{Z}^* = \Big[\overline{Z}_1^*, \ldots, \overline{Z}_k^*\Big] = n^{-1}\sum_{l=1}^{n} Z_l^*.$$

A common bootstrap argument for constructing inference about θ is that the sampling distribution of $\hat{\theta} = g(\overline{Z})$ can be approximated by the distribution of $\hat{\theta}^* = g(\overline{Z}^*)$ conditional on $\overline{Z}$. Here, we propose that bootstrap resampling to approximate the sampling distribution of $\hat{\theta}$ can be avoided by applying a tail probability approximation discussed by DiCiccio and Martin (1991) to a saddlepoint approximation for the joint density of $\overline{Z}_1^*, \ldots, \overline{Z}_k^*$.

We first review DiCiccio and Martin's (1991) tail probability approximation. Consider a continuous random vector $Y = (Y_1, \ldots, Y_k)$ having probability density function of the form

$$f_Y(y) \propto b(y)\exp\{\ell(y)\}, \quad y = (y_1, \ldots, y_k). \tag{1}$$

Suppose that the function ℓ is maximized at $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_k)$ and that for $j = 1, \ldots, k$, $Y_j - \hat{y}_j$ is $O_p(n^{-\frac{1}{2}})$ as $n \to \infty$. For each fixed y, assume that $\ell(y)$ and its partial derivatives are $O(n)$ and that $b(y)$ is $O(1)$. Now consider a random variable $W = g(Y)$, where the real-valued function g has continuous gradient that is non-zero at $\hat{y}$. To calculate an approximation to the tail probability $P(W \leq w)$, let $\tilde{y} = \tilde{y}(w)$ be the value of y that maximizes $\ell(y)$ subject to the constraint $g(y) = w$. Let

$$r(w) = \operatorname{sgn}\{w - g(\hat{y})\}\Big(2[\ell(\hat{y}) - \ell\{\tilde{y}(w)\}]\Big)^{\frac{1}{2}},$$

and assume that r is an increasing function of w. Approximations to the distribution function of W can be based on normal approximations to the distribution of $R = r(W)$. More formally, provided $w - g(\hat{y})$ is $O(n^{-\frac{1}{2}})$,

$$P(W \leq w) = \Phi(r) + O(n^{-\frac{1}{2}}), \tag{2}$$

where $r = r(w)$ and Φ is the standard normal distribution function.

DiCiccio and Martin (1991) provide a more accurate approximation to the distribution of R. Put $\ell_i(y) = \partial\ell(y)/\partial y_i$, $\ell_{ij}(y) = \partial^2\ell(y)/\partial y_i\partial y_j$, $g_i(y) = \partial g(y)/\partial y_i$, and $g_{ij}(y) = \partial^2 g(y)/\partial y_i \partial y_j$, $i, j = 1, \dots, k$. Define

$$J_{ij}(w) = -\ell_{ij}\{\tilde{y}(w)\} + \frac{\ell_p\{\tilde{y}(w)\}}{g_p\{\tilde{y}(w)\}} g_{ij}\{\tilde{y}(w)\}, \qquad i, j = 1, \dots, k,$$

where p is any index such that $g_p\{\tilde{y}(w)\}$ is non-zero. Such an index p always exists by virtue of the assumptions about g. Denote the matrix $\{J_{ij}(w)\}$ by $J(w)$ and its inverse $J(w)^{-1}$ by $\{J^{ij}(w)\}$. Finally, let

$$Q(w) = \sum_{i=1}^{k} \sum_{j=1}^{k} J^{ij}(w) g_i\{\tilde{y}(w)\} g_j\{\tilde{y}(w)\}, \quad D(w) = \left\{Q(w)|J(w)|/|J(\hat{w})|\right\}^{-\frac{1}{2}}.$$

Then, the improved tail probability approximation is

$$P(W \leq w) = \Phi(r) + \varphi(r)\left[\frac{1}{r} + D(w)\frac{g_j\{\tilde{y}(w)\}}{\ell_j\{\tilde{y}(w)\}}\frac{b\{\tilde{y}(w)\}}{b(\hat{y})}\right] + O(n^{-\frac{3}{2}}), \tag{3}$$

where $r = r(w)$, φ is the standard normal density function, and j is any index for which $g_j\{\tilde{y}(w)\}$ does not vanish.

Note that the random vector Y used in the above construction is assumed to be continuous, whereas we have in mind a random vector $\overline{Z}^*$ that is discrete; indeed, the distribution of $\overline{Z}^*$ given $Z_1, \dots, Z_n$ has $\binom{2n-1}{n}$ atoms. However, the size of the largest atom of the distribution of $\overline{Z}^*$ given $Z_1, \dots, Z_n$ shrinks exponentially quickly to 0 as $n \to \infty$. Consequently, even for quite small n, the distribution of $\overline{Z}^*$ given $Z_1, \dots, Z_n$ can be regarded as continuous for practical purposes. It is therefore convenient to refer to the "joint density of $\overline{Z}_1^*, \dots, \overline{Z}_k^*$ given $Z_1, \dots, Z_n$" as the density corresponding to a continuous approximation to the joint distribution of $\overline{Z}_1^*, \dots, \overline{Z}_k^*$ given $Z_1, \dots, Z_n$.

In order to apply tail probability approximation (3) in the bootstrap context, an approximation of the form (1) is required to the joint density of $\overline{Z}_1^*, \dots, \overline{Z}_k^*$ given $Z_1, \dots, Z_n$. The approach of Daniels and Young (1991) suggests using a saddlepoint approximation to this joint density. The cumulant generating function of $Z_{11}^*, \ldots, Z_{k1}^*$ given $Z_1, \ldots, Z_n$ is

$$\begin{aligned} K(T_1, \ldots, T_k) &= \log\left[E\{\exp(T_1 Z_{11}^* + \cdots + T_k Z_{k1}^*)|Z_1, \ldots, Z_n\}\right] \\ &= \log\left\{n^{-1}\sum_{j=1}^{n}\exp(T_1 Z_{1j} + \cdots + T_k Z_{kj})\right\}. \qquad (4) \end{aligned}$$

The usual saddlepoint approximation to the density of $\overline{Z}_1^*, \ldots, \overline{Z}_k^*$ given $Z_1, \ldots, Z_n$ is

$$\hat{h}_n(\zeta_1, \ldots, \zeta_k) \propto |\hat{\Delta}(\zeta_1, \ldots, \zeta_k)|^{-\frac{1}{2}} \exp\left[n\left\{K(\hat{T}_1, \ldots, \hat{T}_k) - \sum_{l=1}^{k}\hat{T}_l \zeta_l\right\}\right], \qquad (5)$$

where

$$K_{T_l}(\hat{T}_1, \ldots, \hat{T}_k) = \zeta_l, \quad l = 1, \ldots, k, \qquad (6)$$

are the saddlepoint equations, $K_{T_l} = \partial K(T_1, \ldots, T_k)/\partial T_l$, and $\hat{\Delta} = \{K_{T_l T_m}(\hat{T}_1, \ldots, \hat{T}_k)\}$ is the $k \times k$ matrix of second-order partial derivatives $K_{T_l T_m}(T_1, \ldots, T_k) = \partial^2 K(T_1, \ldots, T_k)/\partial T_l \partial T_m$, $l, m = 1, \ldots, k$, evaluated at $\hat{T}_1, \ldots, \hat{T}_k$. The cumulant generating function (4) has a very simple form, so its derivatives K_{T_l}, $l = 1, \ldots, k$, and $K_{T_l T_m}$, $l, m = 1, \ldots, k$, are easy to calculate algebraically and also very easy to compute. General reviews of saddlepoint methods are given by Barndorff-Nielsen and Cox (1979, 1989) and Reid (1988).

Approximations to $P(\hat{\theta}^* \le w|Z_1, \ldots, Z_n)$ can be obtained by applying tail probability approximation (3) to the approximate density (5) by putting $Y = \overline{Z}^*$, $y = \zeta = (\zeta_1, \ldots, \zeta_k)$, $W = \hat{\theta}^* = g(Y)$, $b(y) = |\hat{\Delta}(\zeta)|^{-\frac{1}{2}}$, and $\ell(y) = n\{K(\hat{T}_1, \ldots, \hat{T}_k) - \sum_{l=1}^{k}\hat{T}_l\zeta_l\}$. It is easily verified that the saddlepoint approximation provided by (5) satisfies the conditions assumed on the density (1).

3. A Simple Example: The Correlation Coefficient

We illustrate the methodology presented in Section 2 for the case of inference for a bivariate correlation coefficient. The data consists of n pairs $(X_{11}, X_{21}), \ldots, (X_{1n}, X_{2n})$. In an obvious notation,

$$\overline{Z} = \left[\overline{X}_1, \overline{X}_2, \overline{X_1^2}, \overline{X_2^2}, \overline{X_1 X_2}\right] = \left[\overline{Z}_1, \overline{Z}_2, \overline{Z}_3, \overline{Z}_4, \overline{Z}_5\right].$$

The bivariate sample correlation coefficient is given by

$$\hat{\theta} = g(\overline{Z}) = \frac{\overline{Z}_5 - \overline{Z}_1\overline{Z}_2}{\left\{(\overline{Z}_3 - \overline{Z}_1^2)(\overline{Z}_4 - \overline{Z}_2^2)\right\}^{\frac{1}{2}}}.$$

In this case, the cumulant generating function of $Z^*_{11}, \ldots, Z^*_{k1}$ given $Z_1, \ldots, Z_n$ is

$$K(T_1, T_2, T_3, T_4, T_5) \\ = \log\left\{n^{-1}\sum_{j=1}^{n}\exp\left(T_1X_{1j} + T_2X_{2j} + T_3X_{1j}^2 + T_4X_{2j}^2 + T_5X_{1j}X_{2j}\right)\right\}.$$

The derivatives and second derivatives of K are simple to calculate.

We consider two examples. The first involves the correlation coefficient for Efron's (1982, p.10) Law School data where the two variates are average LSAT and GPA scores for the 1973 entering classes of 15 American law schools. The second example uses 50 data points simulated from a bivariate log normal distribution with true correlation coefficient 0.378. For these examples, the estimated correlation coefficients were 0.776 and 0.438, respectively. In each case, tail probability approximations (2) and (3) for $\hat{\theta}^*$ given $Z_1, \ldots, Z_n$ were computed, as well as simulated true values. Most calculations involved in computing approximation (3) in each case were straightforward, merely requiring knowledge of the derivatives of K and g. As already mentioned, the derivatives of K are straightforward to compute, even in settings much more complex than this; see Section 4. In simple problems, such as this, the derivatives of g are also easy to compute algebraically, but this will generally not be feasible. In complex problems, there are two possible solutions: use of numerical derivatives; or use of computer algebra to automate the computations. We have used both of these ideas with success. In this example, derivatives of g were computed algebraically. The solution of the saddlepoint equations (6) as part of the constrained maximization of $\ell(\zeta)$ subject to $g(\zeta) = w$ poses the only real numerical challenge in constructing our approximations. The constrained maximization problem can be reduced algebraically to that of solving a system of 11 nonlinear equations in 11 unknowns. The NAG subroutine C05NCF was readily used to solve the system of equations, with no significant problems. The solution of the saddlepoint equations in the unconstrained maximization of $\ell(\zeta)$ is known. In that case, $\hat{T}_1 = \cdots = \hat{T}_5 = 0$. Values close to these can be used as starting values in the constrained maximization step. Finally, NAG subroutines F01AAF and F03AAF were used to find the inverse and determinant of $J(\zeta)$ respectively.

The results of our study are reported in Table 1. The simulated true values in the case of the Law School data were each computed using 5,000,000 simulations and the simulated values obtained in the case of simulated lognormal data were each calculated using 100,000 simulations. In both cases, tail probability approximation (3) performs very well, remaining close to the simulated true values even in the extreme tails of the distribution. Approximation (2), the naive normal approximation to the distribution of $R = r(\hat{\theta}^*)$, does not perform well in the tails, although it seems adequate in the center of the distribution. It was apparent from our numerical investigation that larger sample sizes, say 30 to 50, are required for the approximation to be accurate when the data come from heavy-tailed distributions such as the log normal. Notably, our approximation did not perform well when the data comprised a sample of size 15 from a bivariate log normal distribution, but improved substantially when the sample size was increased to 30 and subsequently to 50. Unreported simulations suggest that for lighter-tailed distributions, such as the bivariate normal, smaller sample sizes, say 10 to 15, are sufficient to ensure that our approximation is highly accurate.

A significant practical advantage of our technique is its accuracy compared to standard bootstrap algorithms. Table 1 suggests that our approximation is consistently close to bootstrap distribution function estimates based on very large numbers of resamples, which, for all practical purposes, can be regarded as exact. How does our method compare with more common bootstrap approaches, which employ 5,000 or 10,000 resamples? To address this question, we carried out a large simulation study of distribution function estimates using our method (denoted $\hat{F}_{an}$), and bootstrap approximations based on 5,000 and 10,000 resamples (denoted $\hat{F}_{5000}$ and $\hat{F}_{10000}$, respectively). In each case, accuracy is assessed relative to an "exact" bootstrap distribution based on 1,000,000 resamples. For each of four underlying bivariate populations, G, and each of three sample sizes $n = 20, 30$ and 50, the following simulation was carried out:

(1) Generate a data set of size n from G;

(2) By drawing 1,000,000 resamples, evaluate the "exact" bootstrap distribution function, $\hat{F}$, of the correlation coefficient, for each point in a grid of 201 equally-spaced points $y_1, \ldots, y_{201}$ in $[-1, 1]$;

(3) Evaluate $\hat{F}_{an}$, and, by resampling, bootstrap estimates $\hat{F}_{5000}$ based on 5000 resamples, and $\hat{F}_{10000}$ based on 10,000 resamples;

(4) For each of $\tilde{F} = \hat{F}_{an}, \hat{F}_{5000}$, and $\hat{F}_{10000}$, compute the L^2 and L^∞ error measures

$$\begin{aligned} \mathrm{err}_1(\tilde{F}) &= \sum_{y_i} \left\{ \tilde{F}(y_i) - \hat{F}(y_i) \right\}^2; \\ \mathrm{err}_2(\tilde{F}) &= \sup_{y_i} \left| \tilde{F}(y_i) - \hat{F}(y_i) \right|, \end{aligned}$$

(5) Repeat steps (1) to (4) 100 times, summing the error measures err_1 and err_2.

The results of the study are given in Table 2. The figures clearly indicate that the analytical technique enjoys higher accuracy than the resampling techniques based on 5,000 or 10,000 resamples in the majority of situations considered. Moreover, as the sample size increases, the accuracy of the analytical technique improves dramatically, while that of the resampling methods only improves marginally. This phenomenon suggests that the analytical technique is preferable to resampling when sample sizes are moderate, both in terms of accuracy and computational expense, since the computational expense associated with resampling techniques becomes quickly worse as sample size increases.

The major advantage of using our method to obtain accurate approximations to bootstrap distribution functions is that it is typically much faster than direct simulation. Moreover, the computational savings increase dramatically as sample size increases. Use of our method was faster per distribution function evaluation than direct simulation based on 100,000 simulations by a factor of about 50 for the sample of size 15 and by a factor of about 100 for the sample of size 50. In addition to its accuracy and efficiency, a further advantage of our methodology is that it avoids the need for the number of resamples to be specified in bootstrap applications.

4. A More Complicated Example: The Studentized Correlation Coefficient

The bootstrap argument in Section 2 that the sampling distribution of $\hat{\theta}$ may be approximated by the bootstrap distribution of $\hat{\theta}^*$ is only reasonable if the statistic $\hat{\theta}$ is either an exact pivot or pivotal to a high order of approximation. For example, in the case of inference for a population mean, more accurate results may be obtained if inference is based on the Studentized mean $(\overline{X} - \mu)/\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of the variance of $\overline{X}$, rather than on $\overline{X}$ or $\overline{X} - \mu$. Consequently, it is important that the function g be chosen carefully, so that $g(\overline{Z})$ is approximately pivotal. Of course, choosing $\hat{\theta}$ to represent a Studentized statistic both complicates the form of g and increases the dimension of $\overline{Z}$, the vector of means necessary for the computation of $\hat{\theta}$. The latter consideration has the most important ramifications for our method because it implies also an increase in the dimension of the system of equations arising in the constrained maximization step.

Suppose inference for a correlation coefficient is to be based on a Studentized correlation coefficient. A natural approach in this instance is to work with the variance-stabilized, z-transformed correlation coefficient,

$$Z = \tanh^{-1}(\hat{\rho}),$$

rather than with $\hat{\rho}$ itself, and base inference on the Studentized, transformed coefficient. The variance of $n^{\frac{1}{2}}Z$ is

$$\rho^2(1-\rho^2)^{-2}\left\{\mu_{22}\mu_{11}^{-2} + \frac{1}{4}\left(\mu_{40}\mu_{20}^{-2} + \mu_{04}\mu_{02}^{-2} + 2\mu_{22}\mu_{20}^{-1}\mu_{02}^{-1}\right)\right.$$

$$- \left(\mu_{31}\mu_{20}^{-1}\mu_{11}^{-1} + \mu_{13}\mu_{02}^{-1}\mu_{11}^{-1}\right)\Big\}, \tag{7}$$

provided the bivariate moments $\mu_{ij} = E[\{X_{11} - E(X_{11})\}^i\{X_{21} - E(X_{21})\}^j]$, for $i, j = 0, \ldots, 4$, exist; see Kendall and Stuart (1979, p.312). The nonparametric delta method estimate of this variance is the obvious plug-in estimate of (7). Let $\hat{\theta}$ represent the transformed correlation coefficient, Studentized using the nonparametric delta method estimate of (7). Then, the function g is a function of the vector of means

$$\begin{aligned}\overline{Z} &= \left[\overline{X}_1, \overline{X}_2, \overline{X_1^2}, \overline{X_2^2}, \overline{X_1X_2}, \overline{X_1^2X_2}, \overline{X_1X_2^2}, \overline{X_1^2X_2^2}, \overline{X_1^3}, \overline{X_2^3}, \overline{X_1^3X_2}, \overline{X_1X_2^3}, \overline{X_1^4}, \overline{X_2^4}\right] \\ &= \left[\overline{Z}_1, \ldots, \overline{Z}_{14}\right].\end{aligned}$$

The cumulant generating function of $Z_{11}^*, \ldots, Z_{14,1}^*$ given $Z_1, \ldots, Z_n$ is

$$K(T_1, \ldots, T_{14}) = \log\left\{\frac{1}{n}\sum_{j=1}^{n}\exp\left(\sum_{i=1}^{14}T_iZ_{ij}\right)\right\}.$$

Derivatives of K are easily computed algebraically. Derivatives of g necessary for the computation of (3) are tedious to compute algebraically, so numerical approximations are preferable in this instance. The primary hurdle to the use of our technique in this example is the fact that the constrained maximization of $\ell(\zeta)$ subject to the constraint $g(\zeta) = w$ reduces to the problem of solving a system of 29 equations in 29 unknowns. Throughout our numerical investigations, the NAG subroutine C05NCF was used to solve the system of equations without any significant problems. Of course, the algorithm suffers some loss of efficiency as the dimension of the system of equations grows, but we still find that it is competitive with standard resampling schemes while still retaining an advantage in terms of accuracy.

A numerical study was carried out to assess the accuracy of our technique for estimating the distribution function of a Studentized, transformed correlation coefficient. The data we used consisted of pairs of test scores on 31 individuals reported by Ryan, Joiner and Ryan (1992, p.219). The results of our study are reported in Table 3. True simulated values were computed using 5,000,000 bootstrap replications. Despite the complexity of the problem, both the naive normal approximation (2) and the full approximation (3) perform remarkably well. The full approximation (3) tracks the simulated values very closely, suggesting that our technique has considerable utility even in very complex problems.

5. Concluding Remarks

Remark 1. As Daniels and Young (1991) remark, some care should be taken when using saddlepoint methods in the bootstrap context when the data set contains outlying observations. In those cases, our approximations to bootstrap distribution functions can be inaccurate in the tails and occasionally exhibit local non-monotonic behaviour. Efron's Law School Data contains one mild outlier, but this did not pose a significant problem for the technique.

Remark 2. Our approximations have a natural application in the construction of approximate bootstrap percentile (or, in the light of Section 4, percentile-t) confidence intervals. The resulting intervals typically share the properties of bootstrap intervals formed using resampling. For instance, Efron (1982) notes that bootstrap percentile intervals are equivariant under monotone transformations; intervals based on use of approximation (3) share this property because (3) is equivariant under invertible transformations of W; see DiCiccio and Martin (1991).

Remark 3. The iterated bootstrap (Hall (1986), Beran (1987), Hall and Martin (1988)) is a technique that produces highly accurate inferences. Unfortunately, it is very computationally expensive, even in its simplest applications. Iterated bootstrap computations are readily facilitated by an application of our technique to replace the final level of resampling. DiCiccio, Martin and Young (1992a, b) have investigated the use of approximations related to that given here in the iterated bootstrap context, and report very significant savings over standard techniques based on nested levels of resampling. In the iterated bootstrap setting, DiCiccio, Martin and Young, show that approximations based on the leading term of approximation (3), and that use approximate solutions to the saddlepoint equations (6), are sufficiently accurate to produce two-sided iterated bootstrap confidence intervals with high coverage accuracy.

Table 1. Simulated and Approximate values of $P(\hat{\theta}^* \leq w|Z_1,\ldots,Z_n)$ for the correlation coefficient

	Law School data, $n = 15$		
w	Approximation (2)	Approximation (3)	Simulated (SE)
0.25	0.121	0.193	0.201 (.002)
0.30	0.249	0.367	0.377 (.003)
0.35	0.500	0.685	0.702 (.004)
0.40	0.984	1.250	1.269 (.005)
0.45	1.886	2.235	2.252 (.007)
0.50	3.509	3.907	3.906 (.009)
0.55	6.313	6.659	6.648 (.011)
0.60	10.920	11.014	10.967 (.014)
0.70	28.309	26.860	26.719 (.020)
0.80	57.728	53.568	53.091 (.022)
0.85	74.123	69.074	68.052 (.021)
0.90	88.258	83.585	82.633 (.017)
0.95	97.509	95.514	95.487 (.009)
0.99	99.959	99.904	99.901 (.001)

	Lognormal data, $n = 50$		
w	Approximation (2)	Approximation (3)	Simulated (SE)
–0.10	0.008	0.014	0.020 (.004)
–0.05	0.030	0.050	0.060 (.008)
0.0	0.091	0.149	0.142 (.012)
0.05	0.242	0.381	0.362 (.019)
0.10	0.577	0.859	0.863 (.029)
0.15	1.268	1.730	1.779 (.042)
0.20	2.649	3.253	3.483 (.058)
0.25	5.394	5.990	6.215 (.076)
0.30	10.740	11.001	11.267 (.100)
0.40	35.637	33.496	33.580 (.150)
0.50	72.839	68.570	68.104 (.147)
0.60	94.087	92.165	91.364 (.089)
0.65	97.799	97.030	96.468 (.058)
0.70	99.282	99.040	98.790 (.035)
0.75	99.794	99.724	99.605 (.020)
0.80	99.948	99.929	99.905 (.010)
0.85	99.990	99.984	99.983 (.004)

Note: All table entries and standard errors are percentages. Standard errors of simulated values are given in parentheses.

Table 2. Comparison of analytical and resampling approximations to the bootstrap distribution of a correlation coefficient

Bivariate	n	Squared error (err_1)			Sup error (err_2)		
population, G		$\hat{F}_{an}$	$\hat{F}_{5000}$	$\hat{F}_{10000}$	$\hat{F}_{an}$	$\hat{F}_{5000}$	$\hat{F}_{10000}$
1	20	$\underline{0.0545}$	0.2277	0.1419	$\underline{0.4260}$	1.1335	0.8427
	30	$\underline{0.0064}$	0.2174	0.1059	$\underline{0.1546}$	1.1508	0.7865
	50	$\underline{0.0012}$	0.1586	0.0749	$\underline{0.0878}$	1.1106	0.7594
2	20	0.1477	0.2582	$\underline{0.1300}$	$\underline{0.6553}$	1.1615	0.8259
	30	$\underline{0.0447}$	0.2093	0.0928	$\underline{0.2983}$	1.1289	0.7747
	50	$\underline{0.0027}$	0.1640	0.0714	$\underline{0.1106}$	1.1475	0.7407
3	20	0.1461	0.1955	$\underline{0.1186}$	$\underline{0.7343}$	1.1420	0.8368
	30	$\underline{0.0110}$	0.1753	0.0764	$\underline{0.2378}$	1.1708	0.7790
	50	$\underline{0.0023}$	0.1147	0.0577	$\underline{0.1309}$	1.0579	0.7631
4	20	0.1996	0.1846	$\underline{0.0984}$	$\underline{0.7750}$	1.0746	0.7971
	30	$\underline{0.0193}$	0.1624	0.0687	$\underline{0.2857}$	1.1198	0.7238
	50	$\underline{0.0024}$	0.1241	0.0561	$\underline{0.1183}$	1.0516	0.7686

Note:
Estimator with smallest error is underlined.
Bivariate population 1 denotes $X \sim N(0,1)$, $Y \sim N(0,1)$, $\rho = 0$.
Population 2 denotes $X \sim |N(0,1)|$, $Y \sim |N(0,1)|$, $\rho = 0$.
Population 3 denotes $X = U + V$, $Y = U + W$, U, V, W independent $N(0,1)$ ($\rho = \frac{1}{2}$).
Population 4 denotes $X = U + V$, $Y = U + W$, U, V, W independent $|N(0,1)|$ ($\rho = \frac{1}{2}$).

Table 3. Simulated and Approximate values of $P(\hat{\theta}^* \leq w|Z_1,\ldots,Z_n)$ for the studentized correlation coefficient

	Test Score Data, $n = 31$		
w	Approximation (2)	Approximation (3)	Simulated (SE)
2.25	0.110	0.140	0.134 (0.002)
2.75	0.521	0.617	0.634 (0.004)
3.25	1.992	2.224	2.168 (0.007)
3.75	6.089	6.484	6.425 (0.011)
4.00	9.778	10.207	10.153 (0.014)
4.50	21.341	21.558	21.414 (0.018)
5.00	37.772	37.270	36.789 (0.022)
5.50	55.704	54.212	53.628 (0.022)
6.00	71.183	68.913	68.166 (0.021)
6.50	82.326	79.676	79.048 (0.018)
7.25	91.874	90.082	89.368 (0.014)
7.75	95.203	93.780	93.370 (0.011)
8.00	96.318	95.102	94.811 (0.010)
8.25	97.177	96.168	95.964 (0.009)
8.75	98.348	97.715	97.605 (0.007)
9.00	98.741	98.263	98.092 (0.006)
9.50	99.276	99.031	98.853 (0.004)
9.75	99.455	99.290	99.098 (0.003)

Note: All table entries and standard errors are percentages. Standard errors of simulated values are given in parentheses.

References

Barndorff-Nielsen, O. E. and Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser.B* **41**, 279-312.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–468.

Daniels, H. E. and Young, G. A. (1991). Saddlepoint approximation for the Studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169–179.

Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417–431.

DiCiccio, T. J. and 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. and Young, G. A. (1992a). Analytical approximations for iterated bootstrap confidence intervals. *Statistics and Computing* **2**, 161–171.

DiCiccio, T. J., Martin, M. A. and Young, G. A. (1992b). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika* **79**, 285–295.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

Hall, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14**, 1431–1452.

Hall, P. and Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661–671.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th edition. Charles Griffin, London.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci.* **3**, 213–238.

Ryan, B. F., Joiner, B. L. and Ryan, T. A. (1992). *Minitab Handbook*, 2nd edition. PWS-Kent, Boston.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.
Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K.

(Received February 1992; accepted May 1993)