Statistica Sinica 4(1994), 407-421

THE BOOTSTRAP METHOD WITH SADDLEPOINT APPROXIMATIONS AND IMPORTANCE RESAMPLING

Zehua Chen and Kim-Anh Do

National University of Singapore and Queensland University of Technology

Abstract: An approach combining saddlepoint approximation and importance resampling is developed for approximating the bootstrap distribution of a statistic which is a smooth function of sample means. The idea is to approximate the distribution of the linear part of the statistic, say Y_0, by the saddlepoint technique and then to correct the approximation by the conditional expectation of the quadratic part of the statistic, say Y_1, given Y_0, where the conditional expectation is to be approximated by importance resampling. Techniques for simulating the conditional expectation are developed. The approach is compared with the smoothed importance resampling method through examples. It turns out that, with negligible extra work, significant efficiency gains can be achieved over importance resampling by use of our approach.

Key words and phrases: Bootstrap, conditional expectation, distribution function estimation, importance resampling, saddlepoint approximation.

1. Introduction

Various methods have been developed in the literature for approximating bootstrap distribution functions or quantiles of bootstrap distributions. A naïve method is provided by Monte Carlo simulation. Since the Monte Carlo method needs a huge number of resamplings, efforts have been made either to improve the efficiency of the naïve Monte Carlo method or to develop theoretical approximations. Johns (1988) proposed the method of importance resampling for improving the efficiency of Monte Carlo simulation. Davison and Hinkley (1988) used saddlepoint approximations for the bootstrap distribution of the unstudentized mean of a sample. Daniels and Young (1991) developed a saddlepoint technique for the Studentized mean of a sample. The methods of Daniels and Young has been recently extended to more general statistics by DiCiccio, Martin and Young (1994). DiCiccio, Field and Fraser (1990) proposed a method which can be viewed as a version of the Lugannani-Rice (1980) approximation using an approximate saddlepoint. In this article we combine theoretical approximation and efficient simulation together and provide a hybrid method. These different methods represent different ways of approximating the bootstrap distribution function. DiCiccio et al.'s (1990) method does not need simulation at all but requires some sophisticated theory. The Monte Carlo method (including importance resampling) does not need any sophisticated theory but needs a huge amount of simulation. However, the hybrid method, by combining theoretical approximation and simulation together, requires less sophisticated theory and less amount of simulation. The idea of the hybrid method might also be applied to the problems for which a decomposition of the quantity to be estimated is possible and accurate theoretical approximations can be made to certain components while the others can be simulated.

Throughout this paper, we assume that the statistic of interest is a smooth function of sample means, for example: Studentized mean, Studentized variance, correlation coefficient, etc. A statistic of this kind usually admits an asymptotic expansion in powers of $n^{-\frac{1}{2}}$, n being the sample size, with the first and second terms linear and quadratic in the sample means respectively. Through a device given in Barndorff-Nielson and Cox (1989), an asymptotic expansion of the distribution function of the statistics can be obtained in terms of the cumulative distribution function (CDF) and the probability density function (PDF) of the linear term and the conditional expectation of the quadratic term given the linear term. Our idea is to approximate the CDF and the PDF of the linear term by the saddlepoint technique and to approximate the conditional expectation by importance resampling. The estimate of the distribution function of the statistic is then obtained by plugging in these approximations in the asymptotic expansion. The quantiles can be obtained through a Cornish-Fisher type device. This approach is compared with smoothed importance resampling through numerical examples. Our simulation results show that, with negligible extra work, significant efficiency gains can be achieved over smoothed importance resampling by the use of our approach.

The paper is organized as follows. Section 2 describes the general methodology. Section 3 illustrates the application of the approach to the Studentized mean. Section 4 contains a simulation study comparing the approach with smoothed importance resampling. Some technical details are given in an appendix.

2. Methodology

Let $S_n = n^{1/2}A(\bar{\mathbf{X}})$ be the statistic of interest, where $\bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$, $\mathbf{X}_i$ are i.i.d. d-dimensional random vectors such that $E\mathbf{X}_i = \mu$, and $A(\mathbf{X})$ is a d-variate function with $A(\mu) = 0$. Suppose that A has continuous third derivatives in a neighbourhood of μ and that the random vector $\mathbf{X}$ has a finite variance-covariance matrix. Denote $a_j = \partial A(\mathbf{X})/\partial X^{(j)}|_{\mathbf{X}=\mu}$ and $a_{jk} = \partial^2 A(\mathbf{X})/\partial X^{(j)}\partial X^{(k)}|_{\mathbf{X}=\mu}$ where superscripts denote the components of a vector.

Then, the statistic S_n can be expanded as

$$S_n = \sum_{j=1}^{d} a_j n^{1/2}(\bar{X}-\mu)^{(j)} + n^{-\frac{1}{2}}\frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} a_{jk} n^{1/2}(\bar{X}-\mu)^{(j)} n^{1/2}(\bar{X}-\mu)^{(k)} + O_p(n^{-1}).$$

Let

$$Y_0 = \sum_{j=1}^{d} a_j n^{1/2}(\bar{X}-\mu)^{(j)}$$

and

$$Y_1 = \frac{1}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} a_{jk} n^{1/2}(\bar{X}-\mu)^{(j)} n^{1/2}(\bar{X}-\mu)^{(k)}$$

Let F_0 and f_0 denote the CDF and PDF of Y_0 respectively. Then from Barndorff-Nielson and Cox (1989, p.77), an asymptotic expansion of F_{S_n} can be obtained as follows:

$$F_{S_n}(s) = F_0(s) - n^{-1/2} f_0(s)\mu_{1.0}(s) + O(n^{-1}),$$

where $\mu_{1.0}(s)$ is the conditional expectation of Y_1 given $Y_0 = s$. Let $\widetilde{Y}_0 = n^{-1/2}Y_0$, i.e

$$\widetilde{Y}_0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} a_j(X_i^{(j)} - \mu^{(j)}) = \frac{1}{n}\sum_{i=1}^{n} V_i,$$

say, where $V_i = \sum_{j=1}^{d} a_j(X_i^{(j)} - \mu^{(j)})$ are i.i.d. Let $\widetilde{F}_0$ and $\widetilde{f}_0$ denote the CDF and PDF of $\widetilde{Y}_0$ respectively. Then, in terms of $\widetilde{F}_0$ and $\widetilde{f}_0$, the asymptotic expansion of F_{S_n} can be expressed as

$$F_{S_n}(s) = \widetilde{F}_0(n^{-1/2}s) - \widetilde{f}_0(n^{-1/2}s)\mu_{1.0}(s)/n + O(n^{-1}). \quad (1)$$

Note that the second term in the above expansion is actually of order $O(n^{-\frac{1}{2}})$ since $n^{-1/2}\widetilde{f}_0(n^{-1/2}s)$ is of order $O(1)$.

The CDF $\widetilde{F}_0$ and PDF $\widetilde{f}_0$ can be readily approximated by simple saddlepoint approximations while the conditional expectation $\mu_{1.0}(s)$ can be approximated through importance resampling. Thus, we propose to plug in the saddlepoint approximations of $\widetilde{F}_0$ and $\widetilde{f}_0$ and the importance resampling approximation of $\mu_{1.0}(s)$ into the asymptotic expansion (1) to obtain an estimate of $F_{S_n}(s)$.

The reader might wonder whether we can get more efficiency by using the hybrid method than by just simulating the distribution function of the statistics itself. In our application, the distribution function F_{S_n} will be replaced by its bootstrap estimate which is, at best, $O(n^{-1})$ distance away from F_{S_n}. So, when it comes to approximate the bootstrap estimate, as long as the error rate is of concern, it is enough to achieve an accuracy of order $O(n^{-1})$. If the distribution function itself were to be simulated, a simulation size needed to achieve an accuracy of order $O(n^{-1})$ is required. The simulation size to achieve this is proportional to n^2, (c.f. Chen and Do (1992), Section 2). But, by using the hybrid method, a smaller simulation size is required to achieve an accuracy of order $O(n^{-1/2})$ in the simulation of $\mu_{1.0}$. Such a simulation size is proportional to $n^{5/4}$, (c.f. (9) and (10) below). Thus, in general, the hybrid method will be more efficient than the direct simulation of the distribution function, and the efficiency gain will increase as n increases. This is vindicated in our simulation studies.

Saddlepoint Approximation. The saddlepoint technique was introduced by Daniels (1954) to approximate the PDF of the mean of i.i.d. random variables. A formula for approximating tail probabilities of a distribution using saddlepoint technique was developed by Lugannani and Rice (1980). We use Daniels' formula to approximate the PDF $\widetilde{f}_0$ and Lugannani and Rice's formula to approximate the CDF $\widetilde{F}_0$. The following formulae are adopted from Daniels (1987):

$$\widetilde{F}_0(s) = \Phi(\hat{\xi}) - \phi(\hat{\xi})\{1/\hat{Z} - 1/\hat{\xi} + O(n^{-\frac{3}{2}})\}, \quad (2)$$

$$\widetilde{f}_0(s) = \{n/(2\pi K''(\hat{T}))\}^{\frac{1}{2}} \exp\{n(K(\hat{T}) - \hat{T}s)\}\{1 + O(n^{-1})\}, \quad (3)$$

where

$$\hat{Z} = \hat{T}(nK''(\hat{T}))^{\frac{1}{2}}, \quad \hat{\xi} = \text{sgn}T \cdot \{2n(\hat{T}s - K(\hat{T}))\}^{\frac{1}{2}}, \quad (4)$$

$\hat{T}$ is the root of $K'(t) = s$, K is the cumulant generating function of V_1, and Φ and ϕ are the CDF and PDF respectively of the standard normal distribution.

Estimation of $\mu_{1.0}(s)$. The conditional expectation $\mu_{1.0}(s)$ can be approximated through resampling. Since $\mu_{1.0}(s)$ can not be directly simulated, we need first to define a theoretical estimate of $\mu_{1.0}(s)$. Let $R(t)$ be a probability density function satisfying $\int tR(t)dt = 0$, $\int t^2R(t)dt = 1$ and where $\int t^iR^2(t)dt$ is finite for $i = 0, 1, 2$. Define

$$\mu_{01}(s) = E\{Y_1R_h(s - Y_0)\},$$
$$\mu_0(s) = E\{R_h(s - Y_0)\},$$

and

$$\widetilde{\mu}_{1.0}(s) = \mu_{01}(s)/\mu_0(s),$$

where

$$R_h(t) = \frac{1}{h}R\left(\frac{t}{h}\right).$$

Let $f_{01}(y_0, y_1)$ be the joint PDF of Y_0 and Y_1. It follows from the properties of $R(t)$ that

$$\mu_{01}(s) = \int y_1 f_{01}(s, y_1)dy_1 + \frac{1}{2}h^2 \int f''_{01}(s, y_1)dy_1 + o(h^2)$$

and

$$\mu_0(s) = f_0(s) + \frac{1}{2} f_0''(s) h^2 + o(h^2),$$

where the double prime indicate the second derivatives with respect to s. Thus

$$\widetilde{\mu}_{1.0}(s) = \mu_{1.0}(s) + \frac{h^2}{2 f_0(s)} \Big\{ \int y_1 f_{01}''(s, y_1) dy_1 - f_0''(s) \mu_{1.0}(s) \Big\} + o(h^2). \quad (6)$$

Now let $\mathcal{X}_b^* = \{\mathbf{X}_{b1}^*, \ldots, \mathbf{X}_{bn}^*\}$, $b = 1, \ldots, B$, be B samples from the distribution of $\mathbf{X}$ and let (Y_{b0}^*, Y_{b1}^*) be the corresponding values of (Y_0, Y_1) calculated from the bth sample. Let

$$\hat{\mu}_{01}(s) = \frac{1}{B} \sum_{b=1}^{B} Y_{b1}^* R_h(s - Y_{b0}^*),$$

$$\hat{\mu}_0(s) = \frac{1}{B} \sum_{b=1}^{B} R_h(s - Y_{b0}^*).$$

Then an approximation to $\widetilde{\mu}_{1.0}(s)$ can be provided by

$$\hat{\mu}_{1.0}(s) = \hat{\mu}_{01}(s) / \hat{\mu}_0(s), \quad (7)$$

which is the naïve Monte Carlo approximation. A similar estimation of conditional distribution of form (7) was proposed by Booth, Hall and Wood (1992).

In the context of bootstrap, the efficiency of the naïve Monte Carlo simulation can be improved by various resampling schemes with each aiming at a certain particular purpose. The importance resampling scheme proposed by Johns (1988) stands out in simulating bootstrap distribution functions. The basic idea of importance resampling is to place different rather than equal resampling probabilities at each data point such that it is more likely for a bootstrap statistic to assume a value in the vicinity of a given point of interest, then that point may be approximated with greater accuracy. The importance resampling method has been successfully extended to simulating smoothed bootstrap distribution functions as well, (see Chen and Do (1992)). In the context of the ordinary unsmoothed bootstrap, a refinement of importance resampling, namely, balanced importance resampling, has also been developed, (see Booth, Hall and Wood (1993)). Balanced importance resampling can perform better than ordinary importance resampling, but can not be readily extended to handle the smoothed bootstrap. To improve the efficiency of the naïve Monte Carlo approximation (7), we now suggest an exponential tilted version of importance resampling analogous to that proposed by Chen and Do (1992). Let $f(\mathbf{X})$ denote the PDF of $\mathbf{X}$. Define

$$g(\mathbf{X}) = \exp\Big\{\alpha n^{-1/2} \sum_{j=1}^{d} a_j (X^{(j)} - \mu^{(j)}) + \beta\Big\} f(\mathbf{X}),$$

where α is a constant to be chosen and β is the constant such that $g(\mathbf{X})$ is a density function. Let $\mathcal{X}_b^\dagger = \{\mathbf{X}_{b1}^\dagger \dots \mathbf{X}_{bn}^\dagger\}$, $b = 1, \dots, B$, be B samples from the distribution with density $g(\mathbf{X})$ and let $(Y_{b0}^\dagger, Y_{b1}^\dagger)$ be the corresponding values of (Y_0, Y_1) calculated from the bth sample. Then define

$$\mu_{01}^\dagger(s) = \frac{1}{B}\sum_{b=1}^{B} Y_{b1}^\dagger R_h(s - Y_{b0}^\dagger)\exp\{-\alpha Y_{b0}^\dagger - n\beta\},$$
$$\mu_0^\dagger(s) = \frac{1}{B}\sum_{b=1}^{B} R_h(s - Y_{b0}^\dagger)\exp\{-\alpha Y_{b0}^\dagger - n\beta\},$$

and

$$\mu_{1.0}^\dagger(s) = \mu_{01}^\dagger(s)/\mu_0^\dagger(s). \tag{8}$$

This is the importance resampling approximation to $\tilde{\mu}_{1.0}(s)$.

To actually carry out the importance resampling approximation, we need to choose the constant α in the resampling probability density function $g(\mathbf{X})$ and the bandwidth h in the definition of $\tilde{\mu}_{1.0}(s)$. To this end, let (ℓ, Q) be the weak limit of (Y_0, Y_1). Denote the conditional expectation and the conditional variance of Q given $\ell = s$ by $e(s)$ and $v(s)$ respectively. In the appendix, we prove that

$$E^\dagger\left(\mu_{1.0}^\dagger(s) - \mu_{1.0}(s)\right)^2 \sim \frac{\sqrt{2\pi}k_0 v(s)\exp\{-\alpha s + \frac{\alpha^2}{2} + \frac{s^2}{2}\}}{Bh} + h^4\Big[\frac{1}{2}e''(s) - se'(s)\Big]^2, \tag{9}$$

where $k_0 = \int R^2(s)ds$ and $E^\dagger$ indicates that the expectation is taken with respect to the density g. It follows from (9) that the constants α and h should be chosen, respectively, as

$$\alpha = s$$

and

$$h = \Big(\frac{\sqrt{2\pi}k_0 v(s)}{B[e''(s) - 2se'(s)]^2}\Big)^{\frac{1}{5}}. \tag{10}$$

Substituting (2), (3) and (8) into (1), we obtain an approximation to $F_{S_n}(s)$ as

$$F_{S_n}(s) = \hat{F}_0(s) - n^{-1/2}\hat{f}_0(s)\mu_{1.0}^\dagger(s) + O(n^{-1/2}h^2) + O(n^{-1}), \tag{11}$$

where

$$\hat{F}_0 = \Phi(\hat{\xi}) - \phi(\hat{\xi})[1/\hat{Z} - 1/\hat{\xi}],$$
$$\hat{f}_0(s) = \exp\{n(K(\hat{T}) - \hat{T}s/\sqrt{n}\}/[2\pi K''(\hat{T})]^{\frac{1}{2}},$$

and $\hat{Z}$ and $\hat{\xi}$ are as defined in (4) where $\hat{T}$ is the root of $K'(t) = n^{-1/2}s$.

If, alternatively, we want to estimate the quantiles of F_{S_n}, an approximation formula for the quantiles can be deduced from (11) by a Cornish-Fisher argument.

Let $U_\alpha, U_{\alpha 0}$ be the α-quantile of F_{S_n} and $\hat{F}_0$ respectively. Since, in practice, B will be large enough such that $O(h^2) \sim O(n^{-1/2})$, we have

$$\begin{aligned}\alpha &= \hat{F}_0(U_\alpha) - n^{-1/2}\hat{f}_0(U_\alpha)\mu^{\dagger}_{1.0}(U_\alpha) + O(n^{-1}) \\ &= \hat{F}_0(U_{\alpha 0}).\end{aligned} \tag{12}$$

Expanding the right-hand side of (12) at $U_{\alpha 0}$ up to order $O(n^{-1})$ and then inverting it, we obtain an asymptotic expansion for U_α as

$$U_\alpha = U_{0\alpha} + n^{-1/2}\mu^{\dagger}_{1.0}(U_{0\alpha}) + O(n^{-1}).$$

We conclude this section with a remark. The methodology developed in this section is mainly to be applied to approximating the bootstrap distribution of S_n. In this case, the distribution function to be approximated is to be replaced by a bootstrap distribution function.

3. Application to the Bootstrap

In this section, we apply the methodology developed in the previous section to the approximation of bootstrap distribution functions. We only illustrate the application in the framework of the smoothed bootstrap. However, the application is not restricted to the smoothed bootstrap. The methodology can be applied similarly in the unsmoothed bootstrap while still retaining the desired accuracy.

Before continuing, we give a brief explanation of how the error rate $O(n^{-1})$ can be retained in the un-smoothed bootstrap while a density function does not exist. There is always a continuous distribution function, say $\hat{F}$, with a density function such that the distribution function of S_n under $\hat{F}$, $\hat{F}_{S_n}$, differs from the bootstrap distribution function F_{n,S_n} only by an error of order $O(n^{-1})$. For example, we can take $\hat{F}$ as

$$\hat{F}(x) = \begin{cases} 0, & x < x_{(1)} - \epsilon, \\ \frac{1}{n\epsilon}(x - x_{(1)} + \epsilon), & x_{(1)} - \epsilon \le x < x_{(1)}, \\ F_n(x_{(i)}) + \frac{1}{n\epsilon}(x - x_{(i+1)} + \epsilon), & x_{(i+1)} - \epsilon \le x < x_{(i+1)}, \\ & i = 1, \ldots, n-1, \\ F_n(x), & x_{(i)} \le x < x_{(i+1)} - \epsilon, \\ & i = 1, \ldots, n-1, \end{cases}$$

where the $x_{(1)}, \ldots, x_{(n)}$ are order statistics. With an appropriate ϵ, it follows from an Edgeworth expansion argument that $\hat{F}_{S_n} = F_{n,S_n} + O(n^{-1})$. In fact, the difference can be made to any order of n by appropriately choosing ϵ. Thus $\hat{F}_{S_n}$ can well be taken as an approximation to F_{S_n} with the same error rate $O(n^{-1})$ and the hybrid method can be applied to $\hat{F}$. Let $\hat{K}(t)$ and $K_n(t)$ denote the cumulant generating functions of $\widetilde{Y}_0$ under $\hat{F}$ and F_n respectively. It follows from a Taylor expansion argument that the difference between the two cumulant generating functions can be made to any order of n by appropriately choosing ϵ. Thus in the saddlepoint approximation, if $\hat{K}$ is replaced by K_n, the same error rate of approximation can be retained, (c.f. Field and Ronchetti (1990, Section 4.3)). Therefore, in the case of un-smoothed bootstrap, conceptually, we approximate the distribution function $\hat{F}_{S_n}$, but computationally, we do not need to know $\hat{F}$ and can treat all the empirical quantities as if they are their counterparts under $\hat{F}$ while still retaining the error rate $O(n^{-1})$.

We now return to the smoothed bootstrap. The smoothed bootstrap amounts to replacing the role of the empirical distribution function F_n in the original bootstrap by that of a smoothed version $\hat{F}$. The possibility of using the smoothed bootstrap to improve the performance of the ordinary bootstrap was first discussed in Efron's pioneer paper, see Efron (1979). Silverman and Young (1987) and Young (1988) described a particular version of $\hat{F}$ with density function

$$\hat{f}_\lambda(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda}Q\left(\frac{z - Z_i}{\lambda}\right),$$

where Q is a probability density function with zero mean and unit variance. These authors have developed criteria for determining whether it is advantageous to use the smoothed bootstrap instead of the original bootstrap. The choice of the smoothing parameter λ has been considered by De Angelis and Young (1992a,b). In particular, a general algorithm was developed for the smoothed bootstrap which relies on constructing a bootstrap estimate of the mean squared error of the smoothed bootstrap estimator and choosing the smoothing parameter to minimize this quantity.

We shall illustrate the application to the Studentized mean. In our illustration, take Q to be ϕ, the density function of the standard normal distribution. Throughout, let Z^* denote the random variable associated with this density and let $\hat{\mu}_j$ denote the jth moment of $\hat{f}_\lambda(z)$, i.e. $\hat{\mu}_j = \int z^j \hat{f}_\lambda(z)dz$.

Studentized Mean. The Studentized mean of the data $\mathcal{Z}$ is given by

$$S_n = n^{1/2}\hat{\sigma}^{-1}(\bar{Z} - \mu_z),$$

where

$$\bar{Z} = n^{-1}\sum_{i=1}^{n} Z_i, \quad \mu_z = EZ_i, \quad \text{and} \quad \hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2.$$

The smoothed bootstrap distribution of S_n is the distribution of S_n^*, given by

$$S_n^* = n^{1/2}\hat{\sigma}^{*-1}(\bar{Z}^* - \mu_{z^*}),$$

where

$$\bar{Z}^* = n^{-1}\sum_{i=1}^{n} Z_i^*,\ \mu_{z^*} = E^* Z_i^* = \bar{Z}, \quad \text{and} \quad \hat{\sigma}^{*^2} = n^{-1}\sum_{i=1}^{n}(Z_i^* - \bar{Z}^*)^2.$$

Asymptotic Expansion. After some trivial calculation, it can be shown that

$$S_n^* = Y_0^* + n^{-\frac{1}{2}} Y_1^* + O_p(n^{-1}),$$

where

$$Y_0^* = n^{1/2}\hat{\sigma}^{-1}(\bar{Z}^* - \bar{Z}),$$

$$Y_1^* = n\hat{\sigma}^{-3}\Big\{\bar{Z}(\bar{Z}^* - \bar{Z})^2 - \frac{1}{2}(\bar{Z}^* - \bar{Z})n^{-1}\sum_{i=1}^{n}(Z_i^{*^2} - Z_i^2 - \lambda^2)\Big\}.$$

Cumulant Generating Function. In the context of the Studentized mean, the V in the definition of the cumulant generating function $K(t)$ is given by $V^* = (Z^* - \bar{Z})/\hat{\sigma}$. Hence

$$K(t) = -\frac{t\bar{Z}}{\hat{\sigma}} + \frac{t^2\lambda^2}{2\hat{\sigma}^2} + \log\left(n^{-1}\sum_{i=1}^{n} e^{tZ_i/\hat{\sigma}}\right). \tag{13}$$

Exponential Tilted Density. The exponential tilted density used in resampling for the Studentized mean is as follows:

$$\hat{g}(z) = \exp\{\alpha n^{-1/2}\hat{\sigma}^{-1}(z - \bar{Z}) + \beta\}\hat{f}_\lambda(z),$$

where $\beta = -K(n^{-1/2}\alpha)$ and K is given by (13).

For the technique of resampling under the exponential tilted distribution, see Chen and Do (1992).

Optimal Bandwidth in the Estimation of $\mu_{1.0}(s)$. To calculate the optimal bandwidth h in the estimation of $\mu_{1.0}(s)$, we need first to calculate the limit version of $E(Y_1^* \mid Y_0^* = s)$ and $\mathrm{Var}(Y_1^* \mid Y_0^* = s)$. Let

$$X_{1n} = n^{1/2}\hat{\sigma}^{-1}(\bar{Z}^* - \bar{Z}), \qquad X_{2n} = n^{1/2}\hat{\tau}^{-1}\left(n^{-1}\sum_{i=1}^{n} Z_i^{*^2} - \hat{\mu}_2\right),$$

where $\hat{\tau}^2 = \int(z^2 - \hat{\mu}_2)^2\hat{f}_\lambda(z)dz$. By the multivariate central limit theorem

$$\begin{pmatrix} X_{1n} \\ X_{1n} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\mathbf{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}$$

and

$$\hat{\rho} = (\hat{\mu}_3 - \hat{\mu}_1\hat{\mu}_2)/\hat{\sigma}\hat{\tau}.$$

We can write Y_0^* and Y_1^* as

$$Y_1^* = \hat{\mu}_1\hat{\sigma}^{-1}X_{1n}^2 - \frac{1}{2}\hat{\tau}\hat{\sigma}^{-2}X_{1n}X_{2n}, \quad Y_0^* = X_{1n}.$$

Hence the limit of $E(Y_1^* \mid Y_0^* = s)$ and $\mathrm{Var}(Y_1^* \mid Y_0^* = s)$ are given, respectively, by

$$e(s) = \left(\frac{\hat{\mu}_1}{\hat{\sigma}} - \frac{1}{2}\frac{\hat{\tau}\hat{\rho}}{\hat{\sigma}^2}\right)s^2, \tag{14}$$

$$v(s) = \frac{\hat{\tau}^2(1-\hat{\rho}^2)s^2}{4\hat{\sigma}^4}. \tag{15}$$

Substituting (14) and (15) into (10), we obtain

$$h = \left[\frac{\sqrt{2\pi}k_0\hat{\tau}^2(1-\hat{\rho}^2)s^2}{4B(2\hat{\mu}_1\hat{\sigma} - \hat{\tau}\hat{\rho})^2(1-2s^2)^2}\right]^{\frac{1}{5}}. \tag{15a}$$

4. Simulation Study

We summarize a small simulation study which compares the performance of our method relative to smoothed asymptotic importance resampling in the context of estimating the bootstrap distribution function of a Studentized mean. We shall use the subscripts (SPI) to denote the values of estimates resulting from our method employing saddlepoint approximations to estimate the linear part and importance resampling to simulate the quadratic part conditional on the linear part; the subscript (SAI) denotes values of estimates obtained using smoothed asymptotic importance resampling method, that is, a slight perturbation is added to the importance resamples as described by Chen and Do (1992). We report results for the case where the sample size $n = 10, 20$, or 30; where $B = 200$ and for various tail probabilities $p = 0.025$, 0.05, 0.1, 0.15, 0.2, 0.25, and 0.4 with corresponding t-quantiles $s = t_{n-1}^{-1}(\alpha)$. Assume that the underlying parent population is Normal which gives us conservative estimates of the efficiency gains. Other assumptions for underlying distributions, especially skewed distributions, are also appropriate and usually give better efficiency gains than what we report here. We chose a generic value $\lambda = 1$; the choice of the smoothing parameter should not affect the relative efficiency gains since importance resampling in the second stage should balance out whatever bias is introduced in the first stage. Extensive work on choice of λ has been investigated by De Angelis and Young (1992a,b) and Silverman and Young (1987). We do not focus attention on this issue here.

Let $F_{S_n}(s) \equiv p$ denote the tail probability of the Studentized mean which is fixed in advance. Let $\hat{F}_{\mathrm{SPI}}(s)$ and $\hat{F}_{\mathrm{SAI}}(s)$ denote the approximations to $F_{S_n}(s)$, using B resamples from our method and B importance resamples from the smoothed asymptotic method respectively. We computed the average $\hat{D}_{\mathrm{SPI}}$ of $[\hat{F}_{\mathrm{SPI}}(s) - F_{S_n}(s)]^2$ over $M = 10,000$ independent samples and $\hat{D}_{\mathrm{SAI}}$ is calculated similarly. Finally, we took the ratio $r = \hat{D}_{\mathrm{SAI}}/\hat{D}_{\mathrm{SPI}}$ as a measure of the efficiency of our method in combining saddlepoint approximation with importance resampling relative to smoothed importance resampling alone.

Table 1. Column 4 presents efficiencies of our hybrid SPI method relative to smoothed asymptotic importance (SAI) method in the problem of estimating the distribution function for the Studentized mean. Here $B = 200$, $n = 10$, and the underlying parent population is Normal.

p	$s = t_{n-1}^{-1}(p)$	Johns' A_p	$r = \hat{D}_{\mathrm{SAI}}/\hat{D}_{\mathrm{SPI}}$
0.025	−2.262	−2.178	1.25
0.05	−1.833	−1.894	1.59
0.1	−1.383	−1.575	1.77
0.15	−1.100	−1.371	1.91
0.2	−0.883	−1.216	2.27
0.25	−0.703	−1.078	2.94
0.4	−0.261	−0.773	5.03

Table 2. Same as for Table 1 except that $n = 20$.

p	$s = t_{n-1}^{-1}(p)$	Johns' A_p	$r = \hat{D}_{\mathrm{SAI}}/\hat{D}_{\mathrm{SPI}}$
0.025	−2.093	−2.178	1.45
0.05	−1.729	−1.894	1.96
0.1	−1.328	−1.575	2.35
0.15	−1.066	−1.371	2.49
0.2	−0.361	−1.216	3.03
0.25	−0.688	−1.078	3.97
0.4	−0.257	−0.773	8.29

Table 3. Same as for Table 1 except that $n = 30$.

p	$s = t_{n-1}^{-1}(p)$	Johns' A_p	$r = \hat{D}_{\text{SAI}}/\hat{D}_{\text{SPI}}$
0.025	−2.045	−2.178	3.14
0.05	−1.699	−1.894	3.42
0.1	−1.311	−1.575	3.85
0.15	−1.055	−1.371	4.21
0.2	−0.854	−1.216	4.86
0.25	−0.683	−1.078	5.96
0.4	−0.256	−0.773	15.41

Tables 1, 2, and 3 report our simulation results in the problem of estimating the distribution of a Studentized mean using $B = 200$, $n = 10, 20$, or 30 and employing the optimal bandwidth value of h as given in (15a). For each table, column 1 gives the tail probability; column 2 gives the corresponding t-quantiles $s = t_{n-1}^{-1}(p)$ from which the conditional importance resampling probabilities generating the quadratic parts Y_{1b} $(b = 1, \ldots, B)$ are obtained as

$$p_i = \exp\{sn^{-\frac{1}{2}}\hat{\sigma}^{-1}Z_i\}\Big/\Big[\sum_{j=1}^{n}\exp\{sn^{-\frac{1}{2}}\hat{\sigma}^{-1}Z_j\}\Big];$$

column 3 gives the values of Johns' A_p from which we calculated Johns' asymptotic resampling probabilities to be of the form

$$p_i = \exp\{A_p n^{-\frac{1}{2}}\hat{\sigma}^{-1}Z_i\}\Big/\Big[\sum_{j=1}^{n}\exp\{A_p n^{-\frac{1}{2}}\hat{\sigma}^{-1}Z_j\}\Big].$$

Smoothed importance resamples are easily generated by introducing a perturbation to Johns' importance resamples. Exact details have been described in Chen and Do (1992). Column 4 presents the efficiency gains of our method relative to smoothed asymptotic importance resampling method. The most dominant feature that emerges from our study is that our method (SPI) always performs better than smoothed asymptotic importance resampling method (SAI); especially towards the centre of the distribution. It can be seen from our simulation study that our SPI method has the additional advantage that the efficiency gains increases with n so that as n increases beyond 10 our SPI method may also outperform other methods that do not possess this property, for example, balanced importance resampling (see Booth, Hall, Wood (1993)). Note that the SAI method performs very well relative to uniform resampling towards the tails, so we can deduce that the efficiency gains of our SPI method relative to uniform resampling is excellent throughout, for example when $n = 20$, our SPI method outperforms uniform resampling by at least a factor of 7 irrespective of the quantiles we are trying to estimate. This feature alone is a significant improvement over existing efficient resampling methods where efficiency gains are only good at the tails of the distribution (such as asymptotic importance resampling or balanced importance resampling) or only superior towards the centre of the distribution (such as balanced resampling or antithetic resampling). We also investigated the relative performance of our hybrid method in comparison to using the saddlepoint method alone. We consistently observed that the former is superior to the latter, for example, define $r' = \hat{D}_{\mathrm{SADDLE}}/\hat{D}_{\mathrm{SPI}}$, then for $n = 10$ we obtained $r' = 1.55, 1.82$ when $p = 0.05, 0.10$ respectively.

Appendix

In this appendix, we prove that the asymptotic mean squared error of $\mu^{\dagger}_{1.0}(s)$ is given by

$$E^{\dagger}\left(\mu^{\dagger}_{1.0}(s) - \mu_{1.0}(s)\right)^2 \sim \frac{\sqrt{2\pi}k_0 v(s)\exp(-\alpha s + \frac{\alpha^2}{2} + \frac{s^2}{2})}{Bh} + h^4\left(\frac{1}{2}e''(s) - se'(s)\right)^2. \tag{A1}$$

Proof. First, note that

$$E^{\dagger}\mu^{\dagger}_0(s) = \mu_0(s),$$
$$E^{\dagger}\mu^{\dagger}_{01}(s) = \mu_{01}(s).$$

Expanding $\mu^{\dagger}_{1.0}(s) = \mu^{\dagger}_{01}(s)/\mu^{\dagger}_0(s)$ at $(\mu_0(s), \mu_{01}(s))$ and noting that

$$\mu^{\dagger}_0(s) - \mu_0(s) = O_p(B^{-\frac{1}{2}}) = \mu^{\dagger}_{01}(s) - \mu_{01}(s),$$

we have

$$\mu^{\dagger}_{1.0}(s) = \widetilde{\mu}_{1.0}(s) + \frac{1}{\mu_0(s)}\left[\mu^{\dagger}_{01}(s) - \widetilde{\mu}_{1.0}(s)\mu^{\dagger}_0(s)\right] + O_p(B^{-1}).$$

Since $\widetilde{\mu}_{1.0}(s) = \mu_{1.0}(s) + ch^2 + o(h^2)$, where

$$c = \frac{1}{2f_0(s)}\left[\int y_1 f''_{01}(s, y_1)dy_1 - f''_0(s)\mu_{1.0}(s)\right], \tag{A2}$$

(see (6)), it follows that

$$E^{\dagger}\left(\mu^{\dagger}_{1.0}(s) - \mu_{1.0}(s)\right)^2 = \frac{1}{\mu_0^2(s)}E^{\dagger}\left(\mu^{\dagger}_{01}(s) - \widetilde{\mu}_{1.0}(s)\mu^{\dagger}_0(s)\right)^2 + c^2h^4 + O_p(B^{-\frac{3}{2}}) + O_p(B^{-1}h^2). \tag{A3}$$

Now

$$E^{\dagger}\left(\mu_{01}^{\dagger}(s)-\tilde{\mu}_{1.0}(s)\mu_{0}^{\dagger}(s)\right)^{2}=\frac{k_{0}}{Bh}e^{-\alpha s-n\beta}\int(y_{1}-\tilde{\mu}_{1.0}(s))^{2}f_{01}(s,y_{1})dy_{1}+O(B^{-1}). \tag{A4}$$

Substituting (A2) into (A4), then (A4) into (A3), we obtain

$$E^{\dagger}\left(\mu_{01}^{\dagger}(s)-\mu_{1.0}(s)\right)^{2}=\frac{k_{0}e^{-\alpha s-n\beta}}{Bhf_{0}(s)}\mathrm{Var}(Y_{1}\mid Y_{0}=s)+c^{2}h^{4}+O_{p}(B^{-1}). \tag{A5}$$

Note that

$$\int y_{1}f_{01}(s,y_{1})dy_{1}=f_{0}(s)\mu_{1.0}(s). \tag{A6}$$

Differentiating both sides of (A6) with respect to s yields

$$\int y_{1}f_{01}''(s,y_{1})dy_{1}=f_{0}''(s)\mu_{1.0}(s)+2f_{0}'(s)\mu_{1.0}'(s)+\mu_{1.0}''(s)f_{0}(s).$$

Hence

$$c=\frac{1}{2f_{0}(s)}[2f_{0}'(s)\mu_{1.0}'(s)+\mu_{1.0}''(s)f_{0}(s)]. \tag{A7}$$

Since $Y_0 \rightarrow N(0,1)$ in distribution, $f_0(s) \rightarrow \phi(s)$ and $e^{-n\beta} \rightarrow e^{\alpha^2/2}$ (noting that $e^{-n\beta} = Ee^{\alpha Y_0}$). Substituting (A7) into (A5) and replacing $e^{-n\beta}$, $f_0(s)$, $\mathrm{Var}(Y_1|Y_0=s)$ and $\mu_{1.0}(s)$ by $e^{\alpha^2/2}, \phi(s), v(s)$ and $e(s)$ respectively yields (A1). The proof is complete.

Note: A detailed version of this paper which includes mathematical derivation for the Studentized variance exists in the form of a CMA technical report (The Australian National University) and can be obtained from the second author.

Acknowledgement

The helpful comments of two referees have significantly improved this paper.

References

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

Booth, J., Hall, P. and Wood, A. (1992). Bootstrap estimation of conditional distributions. *Ann. Statist.* **20**, 1594-1610.

Booth, J. G., Hall, P. and Wood, A. T. A. (1993). Balanced importance resampling for the bootstrap. *Ann. Statist.* **21**, 286-298.

Chen, Z. and Do, K.-A. (1992). Importance resampling for the smoothed bootstrap. *J. Statist. Comput. Simulation* **42**, 107-124.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

Daniels, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55**, 37-48.

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.

De Angelis, D. and Young, G. A. (1992a). Smoothing the bootstrap. *Internat. Statist. Rev.* **60**, 45-56.

De Angelis, D. and Young, G. A. (1992b). Bootstrapping the correlation coefficient: A comparison of smoothing strategies. *J. Statist. Comput. Simulation* **40**, 167-176.

DiCiccio, T. J., Field, C. A. and Fraser, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77-95.

DiCiccio, T. J., Martin, M. A. and Young, G. A. (1994). Analytic approximations to bootstrap distribution functions using saddlepoint methods. *Statistica Sinica* **4**, 281-295.

Easton, G. S. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics. *J. Amer. Statist. Assoc.* **81**, 420-430.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Field, C. A. and Ronchetti, E. (1990). *Small Sample Asymptotics.* Institute of Mathematical Statistics, Hayward, CA.

Johns, M. V. (1988). Importance sampling for bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **83**, 709-714.

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12**, 475-490.

Silverman, B. W. and Young, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74**, 469-479.

Young, G. A. (1988). A note on bootstrapping the correlation coefficient. *Biometrika* **75**, 370-373.

Depertment of Mathematics, National University of Singapore, Singapore 0511, Singapore.
School of Mathematics, Queensland University of Technology, Brisbane, Q 4001, Australia.

(Received July 1992; accepted December 1993)