The Annals of Statistics
1993, Vol. 21, No. 1, 286–298

BALANCED IMPORTANCE RESAMPLING FOR THE BOOTSTRAP

BY JAMES G. BOOTH,[1] PETER HALL AND ANDREW T. A. WOOD

Australian National University, Australian National University and CSIRO, and Australian National University

We show that the method of importance resampling, introduced by Vernon Johns and Anthony Davison, may be enhanced by balancing the resamples. It is demonstrated that "balanced importance resampling" improves on both "balanced uniform resampling" and "random importance resampling", from the viewpoint of statistical efficiency. Moreover, the range of applications for which efficient resampling methods may be applied is extended to include statistics which are smooth functions of solutions of estimating equations.

1. Introduction. In this paper we show that the method of importance resampling for approximating a bootstrap distribution function, introduced by Johns (1988) and Davison (1988), can be enhanced by balancing the resamples. That is to say, the resamples should be drawn in such a way that the total number of times each data value appears, rather than the probability of drawing that value, is prescribed. We give a concise prescription of the balancing proportions for achieving maximum asymptotic efficiency. In addition, we extend the range of application of importance resampling methods to statistics which are smooth functions of solutions of estimating equations. Thus, the examples involving M-estimators considered by Johns (1988), and "smooth means" problems, are included as special cases. This is an important extension because, on contemporary computers, the time required to generate resamples (either balanced or random) is very small. Thus, efficient resampling methods are only important in problems where the main computational constraint is finding the estimator.

The emphasis in this paper is upon practical considerations and upon simple derivations of the relevant formulas. Indeed, we shall show how the formulae for the asymptotic efficiencies of both importance and balanced importance resampling methods follow easily from an assumption of asymptotic normality of the bootstrap distribution in question. Examination of these formulas shows that balanced importance resampling is slightly more efficient than random importance resampling for approximating bootstrap tail probabilities and slightly better than balanced uniform resampling in the center of the distribution. The technique therefore provides for unified efficient resampling in

Received April 1990; revised January 1992.

[1]On leave from Department of Statistics, University of Florida, Gainesville, Florida 32611.

AMS 1991 *subject classifications*. Primary 62G09; secondary 65C05.

Key words and phrases. Asymptotic efficiency, estimating equation, exponential tilting, Monte Carlo approximation, smooth function.

bootstrap problems. Technical details may be found in an unpublished technical report [Hall (1990b)], and formal results are outlined in Section 5.

We now introduce the sampling framework and some notation which will be used throughout the paper. Let X denote a k-dimensional random variable with distribution F and let β be a p-dimensional parameter or characteristic of F satisfying

$$E\{\psi_r(X;\beta)\} = 0, \qquad r = 1,\dots,p,$$

for some smooth vector of functions $\psi = (\psi_1,\dots,\psi_p)^T$. Here $E\{\cdot\}$ denotes expectation with respect to F. If $\mathscr{X} = \{x_1,\dots,x_n\}$ is a random sample from F, then an estimate $\hat\beta$ of β may be obtained as a solution of the estimating equations

$$\sum_{i=1}^{n} \psi_r(x_i;\beta) = 0, \qquad r = 1,\dots,p. \tag{1.1}$$

We shall assume that a solution of (1.1) always exists and is a consistent estimate of β as $n \to \infty$. Suppose that our interest lies in a smooth scalar function $\eta = g(\beta)$ of the parameter β which we estimate by $\hat\eta = g(\hat\beta)$. We note that a simple but important class of problems which fall within this framework consists of situations in which g is a smooth function of the mean vector, that is, $\beta = E\{X\}$. In such problems, the functions, ψ_r, $r = 1,\dots,p$, have the simple form: $\psi_r(X,\beta) = X^{(r)} - \beta_r$, $r = 1,\dots,p$, where $X^{(r)}$ denotes the rth component of the random vector X.

Under mild regularity conditions the variance of $\hat\eta$ in repeated sampling is given by

$$\mathrm{Var}\{\hat\eta\} = n^{-1} D_g^T H^{-1} \mathrm{Cov}\{\psi(X,\beta)\}(H^T)^{-1} D_g + o(n^{-1}),$$

where H is a $p \times p$ positive definite matrix with elements $H_{rs} = E\{\partial\psi_r(X,\beta)/\partial\beta_s\}$, $r,s = 1,\dots,p$, and D_g is the p vector with elements $\partial g/\partial\beta_r$, $r = 1,\dots,p$. Thus, a consistent estimate of the variance of $n^{1/2}\hat\eta$ is

$$\hat\sigma^2 = \hat D_g^T \hat H^{-1} \hat\Sigma_\psi (\hat H^T)^{-1} \hat D_g, \tag{1.2}$$

where $\hat H_{rs} = n^{-1}\sum_{i=1}^n \partial\psi_r(x_i;\beta)/\partial\beta_s|_{\beta=\hat\beta}$, $\hat D_g$ has elements $\partial g/\partial\beta_r|_{\beta=\hat\beta}$ and

$$\hat\Sigma_\psi = n^{-1}\sum_{i=1}^{n} \psi\big(x_i;\hat\beta\big)\psi\big(x_i;\hat\beta\big)^T.$$

Let Q denote the distribution function of the statistic $T = n^{1/2}(\hat\eta - \eta)/\hat\sigma$. We shall assume that T is asymptotically pivotal in the sense that it has a standard normal limiting distribution. If $\hat Q$ is an estimate of Q and $\hat t_\alpha$ denotes the quantile of $\hat Q$ satisfying $\hat Q(\hat t_\alpha) = \alpha$, then an approximate $100\alpha\%$ level, equal-tailed confidence interval for η is given by

$$\Big(\hat\eta - n^{-1/2}\hat t_{(1/2)(1+\alpha)}\hat\sigma,\ \hat\eta - n^{-1/2}\hat t_{(1/2)(1-\alpha)}\hat\sigma\Big). \tag{1.3}$$

When $\hat Q$ is the bootstrap estimate of Q, intervals of the form (1.3) are called

percentile-t bootstrap confidence intervals. Percentile-t bootstrap confidence intervals typically enjoy second-order accuracy; that is, the boundaries are correct to order $O_p(n^{-1})$ [see, e.g., Hall (1988)].

In the following section we discuss the bootstrap method of estimating Q, and its Monte Carlo approximant based on "random uniform" resampling. In Section 3 we discuss "random importance" and "balanced importance" resampling approximants for the bootstrap estimator, and derive formulas for their asymptotic efficiencies relative to random uniform resampling. In Section 4 we give the results of a simulation study involving a sample from a Weibull distribution. In this example, $\hat{\beta}$ is the maximum likelihood estimate for the Weibull scale and index parameters and η is the mean of the Weibull distribution. A confidence interval for η could therefore be obtained using standard asymptotic normality results for maximum likelihood estimators. However, this approach leads to confidence intervals which are only first-order correct; that is, the boundaries are correct to order $O_p(n^{-1/2})$. Thus, bootstrap methods are relevant even in fully parametric problems. Section 5 outlines regularity conditions and formal results.

2. Bootstrap estimation of Q. Let $\mathscr{X}^*$ denote a resample of size n drawn randomly with replacement from $\mathscr{X}$, and let $T^* = n^{1/2}(\hat{\eta}^* - \hat{\eta})/\hat{\sigma}^*$ denote the version of T in which $\hat{\eta}^*$ and $\hat{\sigma}^*$ are computed in the same manner as $\hat{\eta}$ and $\hat{\sigma}$ but using the resample $\mathscr{X}^*$ instead of $\mathscr{X}$. The theoretical bootstrap estimate of $Q(t) = P(T \le t)$ is given by

$$\hat{Q}(t) = P(T^* \le t | \mathscr{X}). \tag{2.1}$$

Except in rare cases the estimate (2.1) is not directly computable. In practice $\hat{Q}$ is typically approximated by simulating B "random uniform" resamples $\mathscr{X}_1^*, \ldots, \mathscr{X}_B^*$, and calculating

$$\hat{Q}_{ru}(t) = B^{-1} \sum_{b=1}^{B} I(T_b^* \le t), \tag{2.2}$$

where T_b^* is the value of T^* obtained using the bth resample $\mathscr{X}_b^*$, $b = 1, \ldots, B$.

We note that, conditional on $\mathscr{X}$, the approximation $\hat{Q}_{ru}$ is unbiased for $\hat{Q}$ in (2.1). Moreover, if the assumed asymptotic normality of T implies that of T^* then $\hat{Q}_{ru}$ has variance given by

$$B \operatorname{Var}\{\hat{Q}_{ru}(t) | \mathscr{X}\} = \Phi(t)\{1 - \Phi(t)\} + o_p(1) \tag{2.3}$$

as $n \to \infty$. Here Φ denotes the standard normal distribution function. Thus, for example, for t close to the 5th or 95th percentile of $\hat{Q}$, approximately $B = 7500$ resamples are required to estimate $\hat{Q}(t)$ to two decimal places of accuracy, using (2.2). Conversely, if $\hat{t}_{\alpha, ru}$ denotes the quantile of $\hat{Q}_{ru}$ satisfying $\hat{Q}_{ru}(\hat{t}_{\alpha, ru}) = \alpha$, a Taylor series argument shows that

$$\hat{t}_{\alpha, ru} - \hat{t}_\alpha \approx \{\alpha - \hat{Q}_{ru}(\hat{t}_\alpha)\}\phi(\hat{t}_\alpha)^{-1},$$

where ϕ denotes the standard normal density. Thus, since $\phi(\hat{t}_\alpha) \approx 0.1$ for $\alpha = 0.5$ or 0.95, use of $B = 7500$ resamples results in an estimate of $\hat{t}_\alpha$ which is accurate to one decimal place.

3. Efficient bootstrap simulation. In this section we discuss ways of reducing the number of resamples B used to approximate the exact bootstrap distribution function $\hat{Q}$ in (2.1) while retaining the same level of accuracy as the standard approximant $\hat{Q}_{ru}$ in (2.2). Two methods which have received a considerable amount of attention in the literature are "random importance resampling" [Johns (1988)] and "balanced uniform resampling" [Davison, Hinkley and Schechtman (1986)]. The first of these methods involves selecting B independent "with replacement" resamples of size n from $\mathscr{X}$ with unequal probabilities assigned to the sample values $x_1, \ldots, x_n$. This approach is described in more detail in Section 3.1. In the balanced method, on the other hand, the bth resample of size n consists of the $(nb - n + 1)$th to nbth elements of a random permutation of the ordered set consisting of B replications of $\mathscr{X}$. In the latter method, each sample value is chosen exactly B times. This means that, in the balanced case, the resampling statistics $T_1^*, \ldots, T_B^*$ are conditionally correlated given $\mathscr{X}$. Both approaches lead to unbiased estimates of $\hat{Q}$ in (2.1) which have smaller asymptotic variances than the estimate $\hat{Q}_{ru}$ based on the same number of resamples. The importance resampling approach is generally much more efficient than (2.2) in the tails, whereas balancing leads to substantial improvements in the center of the distribution. In Section 3.2 we discuss a technique called "balanced importance resampling" which combines aspects of both approaches and which contains the balanced uniform method just described as a special case. Balanced importance resampling leads to an estimate for $\hat{Q}$ with bias of order $O(B^{-1})$ but with an asymptotic variance uniformly smaller than that obtained by either random importance or balanced uniform resampling.

3.1. *Random importance resampling.* Under mild regularity conditions $T^* = \sum_{i=1}^n \varepsilon(x_i^*) + O_p(n^{-1/2})$, where

$$\varepsilon_i = \varepsilon(x_i) = n^{-1/2}\hat{\sigma}^{-1}\hat{D}_g^T\hat{H}^{-1}\psi\big(x_i;\hat{\beta}\big), \qquad i = 1, \ldots, n. \tag{3.1}$$

Suppose that a resample $\mathscr{X}^*$, instead of being selected by random uniform resampling as in Section 2, is selected from $\mathscr{X}$ by random sampling with replacement from a "tilted" distribution which assigns probability, $p_i = \exp(\theta\varepsilon_i)/\sum_{j=1}^n \exp(\theta\varepsilon_j)$ to the ith sample value x_i, $i = 1, \ldots, n$. Note that $p_i = n^{-1}\exp\{\theta\varepsilon_i - \kappa(\theta)\}$, where $\kappa(\theta) = \log(n^{-1}\sum_{i=1}^n e^{\theta\varepsilon_i})$ is the empirical cumulant generating function for the ε_i's in (3.1). Note also that the random uniform resampling case is recovered in the case $\theta = 0$. Let $E_\theta\{\cdot|\mathscr{X}\}$ denote expectation under this resampling rule. Then we have

$$\hat{Q}(t) = E_\theta\big\{I(T^* \le t)e^{-n\{\theta\bar{\varepsilon}^* - \kappa(\theta)\}}|\mathscr{X}\big\}$$

for any value of θ, where $\bar{\varepsilon}^*$ is the value that the statistic $\bar{\varepsilon}$ takes when

computed using the resample $\mathscr{X}^*$ rather than $\mathscr{X}$; that is, $\bar{\varepsilon}^* = n^{-1}\sum_{i=1}^n \varepsilon(x_i^*)$. Thus, there is an entire family of unbiased estimates of $\hat{Q}$ (conditional on $\mathscr{X}$) indexed by the tilting parameter θ, with elements given by

$$\hat{Q}_{ri}(t;\theta) = B^{-1}\sum_{b=1}^{B} I(T_b^* \le t)e^{-n\{\theta\bar{\varepsilon}_b^* - \kappa(\theta)\}}. \tag{3.2}$$

In (3.2), T_b^* and $\bar{\varepsilon}_b^*$ are computed using the bth importance resample $\mathscr{X}_b^*$. In the case $\theta = 0$, (3.2) reduces to (2.2).

Now, let Y denote a standard normal variate. Then the asymptotic variance of (3.2) is given by

$$\begin{aligned} B\,\mathrm{Var}_\theta\{\hat{Q}_{ri}(t;\theta)|\mathscr{X}\} &= \mathrm{Var}_\theta\{I(T_b^* \le t)e^{-n\{\theta\bar{\varepsilon}_b^* - \kappa(\theta)\}}|\mathscr{X}\} \\ &\approx \mathrm{Var}\{I(Y + \theta \le t)e^{-\theta Y - (1/2)\theta^2}\} \\ &= \Phi(t + \theta)e^{\theta^2} - \Phi(t)^2. \end{aligned} \tag{3.3}$$

The approximation in (3.3) derives from the assumed asymptotic normality of the statistic T, which implies under regularity conditions (see Section 5) that T^* and $n\bar{\varepsilon}^*$ are approximately normally distributed with mean θ and variance 1. Thus, an efficient Monte Carlo estimate of the theoretical bootstrap value $\hat{Q}(t)$ is obtained by selecting resamples from the tilted distribution with parameter θ which minimizes (3.3). A modification of this approach for the case where t is positive, utilizing the asymmetry about $t = 0$ of the minimized asymptotic variance formula (3.3), is described in Section 3.3.

3.2. *Balanced importance resampling.* Let B_i, $i = 1, \ldots, n$, be nonnegative integers summing to nB and let $\mathscr{X}^\flat$ denote the ordered set of nB elements in which the first B_1 elements equal x_1, the $(B_1 + 1)$st to $(B_1 + B_2)$th elements equal x_2 and so on. Suppose that the B resamples $\mathscr{X}_b^*$, $b = 1, \ldots, B$, are obtained by first randomly permuting $\mathscr{X}^\flat$ and then assigning the $(nb + 1)$st to $(nb + n)$th elements in the reordered set of $\mathscr{X}_b^*$, $b = 1, \ldots, B$. In this section we consider a method for selecting the B_i's which leads to an efficient approximation to the bootstrap distribution function (2.1). More specifically we suppose that $B_i \approx B_i'$, where $B_i' = nBp_i$; the approximation stemming from the fact that each B_i is an integer. There are a number of ways of choosing the B_i. In our approach, we first choose $\tilde{B}_i = [nBp_i]$ for $i = 1, \ldots, n$, where $[\cdot]$ denotes the integer part; next, we obtain $r = nB - \sum_{i=1}^n \tilde{B}_i$ and $d_i = nBp_i - \tilde{B}_i$ for $i = 1, \ldots, n$. Note that $0 \le r \le n$. We then add 1 to those $\tilde{B}_i$ for which the corresponding d_i is at least rth largest in the set $\{d_1, \ldots, d_n\}$, and write $B_1, \ldots, B_n$ for the resulting positive integers (which sum to nB). We note that this construction implies that B_i is of the same order as B for $i = 1, \ldots, n$.

Let $E_\theta\{\cdot|\mathscr{X}, B_i, i = 1, \ldots, n\}$ denote expectation under the resampling rule just described. Then it is shown by Hall (1990b) that, for each $b = 1, \ldots, B$,

$$E_\theta\{I(T_b^* \le t)e^{-n\{\theta\bar{\varepsilon}_b^* - \kappa(\theta)\}}|\mathscr{X}, B_i\ i = 1, \ldots, n\} = \hat{Q}(t) + O(B^{-1}) \tag{3.4}$$

provided $B \gg n$ (e.g., $B = n^2$). In the balanced uniform case $\theta = 0$, the bias

term in (3.4) disappears altogether. Thus, a second family of approximants for $\hat{Q}(t)$ indexed by θ has elements $\hat{Q}_{bi}(t;\theta)$ of the same form as (3.2) but where T_b^* and $\bar{\varepsilon}_b^*$ are computed using the bth *balanced* importance resample. The effect of balancing the resamples is to introduce a negative correlation $\rho = -(B-1)^{-1} + o(B^{-1})$ between the resampling estimates T_b^*, $b = 1, \ldots, B$.

We assume regularity conditions (see Section 5) such that $(T_b^*, T_{b'}^*)$, $b \neq b'$, has an approximate bivariate normal distribution with mean (θ, θ), unit variances and correlation ρ as $n \to \infty$ with $B \gg n$. Let $U_b^* = I(T_b^* \leq t)e^{-n\{\theta\bar{\varepsilon}_b^* - \kappa(\theta)\}}$, for $b = 1, \ldots, B$, and let (Y_1, Y_2) denote a standard bivariate normal variable with correlation coefficient ρ. Then the variance of $\hat{Q}_{bi}(t;\theta)$ is given by

$$
\begin{aligned}
&B\,\mathrm{Var}_\theta\{\hat{Q}_{bi}(t;\theta)|\mathscr{X}, B_i, i = 1, \ldots, n\} \\
&\quad = \mathrm{Var}_\theta\{U_b^*|\mathscr{X}, B_i, i = 1, \ldots, n\} \\
&\qquad + (B-1)\mathrm{Cov}_\theta\{U_b^*, U_{b'}^*|\mathscr{X}, B_i, i = 1, \ldots, n\} \\
&\quad \approx \mathrm{Var}\{I(Y_1 + \theta \leq t)e^{-\theta Y_1 - (1/2)\theta^2}\} \\
&\qquad + (B-1)\mathrm{Cov}\{I(Y_1 + \theta \leq t)e^{-\theta Y_1 - (1/2)\theta^2},\, I(Y_2 + \theta \leq t)e^{-\theta Y_2 - (1/2)\theta^2}\} \\
&\quad \approx \Phi(t+\theta)e^{\theta^2} - \Phi(t)^2 - \{\theta\Phi(t) + \phi(t)\}^2,
\end{aligned}
\tag{3.5}
$$

using the argument in Appendix A. An efficient Monte Carlo estimate of the theoretical bootstrap value $\hat{Q}(t)$ is obtained by selecting balanced resamples with B_i's determined by the value of θ which minimizes (3.5). As with importance resampling, a modification of this approach for the case $t > 0$ is suggested in the following section which takes account of the asymmetry about $t = 0$ of the minimized asymptotic variance formula (3.5).

3.3. *Efficiency comparisons.* Figure 1 depicts graphs of

$$\{\Phi(t)[1-\Phi(t)]\}/\min_\theta\{\Phi(t+\theta)e^{\theta^2} - \Phi(t)^2\},$$

$$\{\Phi(t)[1-\Phi(t)]\}/\min_\theta\{\Phi(t+\theta)e^{\theta^2} - \Phi(t)^2 - [\theta\Phi(t) + \phi(t)]^2\}$$

and

$$\{\Phi(t)[1-\Phi(t)]\}/\{\Phi(t)[1-\Phi(t)] - \phi(t)^2\}$$

versus t, corresponding to the asymptotic efficiencies of approximations to the theoretical bootstrap distribution based on random importance, balanced importance and balanced uniform resampling, respectively, relative to random uniform resampling. It is clear from the graphs that both random and balanced importance methods are more efficient for negative t. Hence, in these cases, when t is positive we first approximate $\hat{G}(t) = P(-T^* \leq -t|\mathscr{X})$ by either $\hat{G}_{ri}(t)$ or $\hat{G}_{bi}(t)$ say, and then set $\hat{Q}_{ri}(t) = 1 - \hat{G}_{ri}(t)$ and $\hat{Q}_{bi}(t) = 1 - \hat{G}_{bi}(t)$. The asymmetry of the random importance resampling efficiency curve is of course well known; see Johns (1988) and Hinkley and Shi (1989).

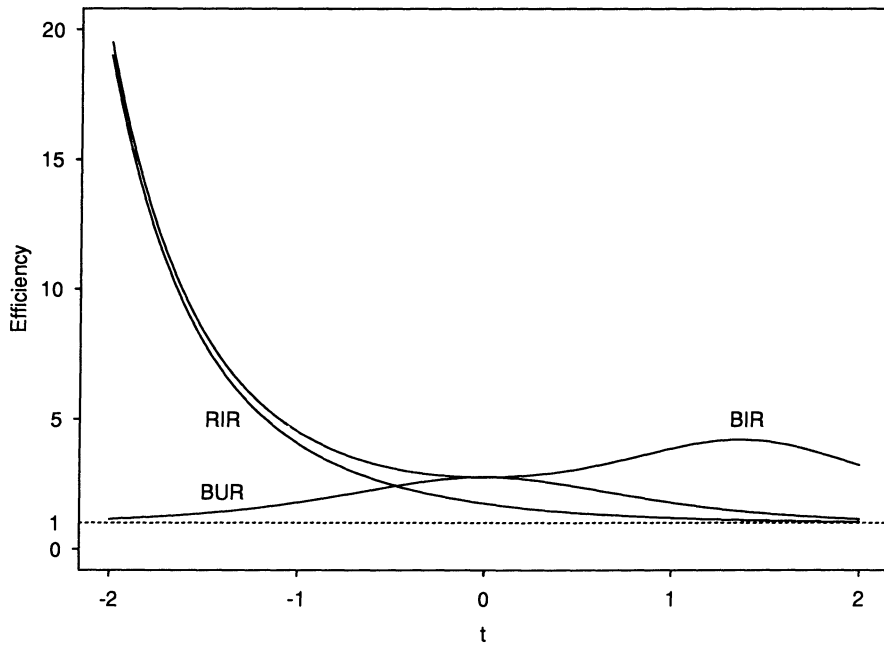

FIG. 1. *Asymptotic relative efficiencies.*

To construct confidence intervals of the form (1.3), for example, approximation of α-level quantiles of the theoretical bootstrap distribution is required. This may be accomplished as follows. First arrange the values $T_1^*, \ldots, T_B^*$ in ascending order. Let (b) denote the subscript of the bth smallest value. Then compute

$$S_r = \frac{1}{B} \sum_{b=1}^{r} \exp\bigl[-n\{\theta \bar{\varepsilon}_{(b)}^* - \kappa(\theta)\}\bigr] \tag{3.6}$$

as in Johns (1988) for increasing values of r, starting with $r = 1$, until S_r first exceeds α (at $r = R + 1$, say). An approximant for $\hat{t}_\alpha$ is then found by interpolating between $T_{(R)}^*$ and $T_{(R+1)}^*$. If $R = 0$, approximate $\hat{t}_\alpha$ by $T_{(1)}^*$. The value of the "tilting" parameter θ in (3.6), determining the way in which the importance resamples are selected, should be chosen to minimize the relevant asymptotic variance formula, (3.3) or (3.5), with $t = z_\alpha$, the α-level quantile of the standard normal distribution. Some of these "optimal" θ values are tabulated in Table 1 for a range of values of α together with the corresponding asymptotic efficiencies. It is interesting to note how close the optimal θ values are to the relevant quantiles of the standard normal distribution. This is particularly striking in the balanced case. In practice, substituting z_α for the optimal θ is likely to have only a marginal effect on efficiency. An intuitive interpretation of this phenomenon is that, since $\kappa'(\theta) = \theta$ asymptotically, the optimal tilted distribution is approximately centered over z_α, the asymptotic

TABLE 1
Optimal "tilting" parameters and asymptotic efficiencies of random and balanced importance resampling schemes versus uniform resampling for various values of α

		Balanced		Random	
α	z_α	θ	eff.	θ	eff.
0.025	−1.960	−1.959	18.03	−2.178	17.53
0.050	−1.645	−1.613	10.41	−1.894	9.98
0.100	−1.282	−1.206	6.20	−1.575	5.77
0.250	−0.675	−0.555	3.45	−1.078	2.89
0.500	−0.000	−0.000	2.76	−0.612	1.75

value of the quantile of interest. Quantiles close to the mode of a distribution can generally be estimated more accurately than those in the tails which explains why importance resampling leads to greater efficiency for quantile estimation.

4. A simulation study. In this section we compare the efficiencies of the four resampling methods described in Sections 2 and 3 in the context of a particular parametric model. Specifically, suppose that $x_1, \ldots, x_n$ is a random sample from a Weibull distribution with unknown scale parameter ω, $\omega > 0$, and index ν, $\nu > 0$, also unknown. That is, $x_1, \ldots, x_n$ is a sample from the distribution with density

$$f(x; \omega, \nu) = \omega \nu x^{\nu-1} e^{-\omega x^\nu}, \qquad x > 0. \tag{4.1}$$

Estimates of ω and ν based on the sample may be obtained by solving the maximum likelihood equations which are of the form (1.1) with $p = 2$ and

$$\psi_1(x_i; \omega, \nu) = \frac{1}{\nu} + \log(x_i) - \omega x_i^\nu \log(x_i),$$

$$\psi_2(x_i; \omega, \nu) = \frac{1}{\omega} - x_i^\nu.$$

In this case, solving (1.1) reduces to finding $\hat{\nu}$ which satisfies

$$\frac{1}{\hat{\nu}} + \frac{1}{n}\sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^n x_i^{\hat{\nu}} \log(x_i)}{\sum_{i=1}^n x_i^{\hat{\nu}}} = 0. \tag{4.2}$$

Then, $\hat{\omega}^{-1} = n^{-1}\sum_{i=1}^n x_i^{\hat{\nu}}$. It can be shown that (4.2) has a unique solution in the range $[\{\log(x_{(n)}) - n^{-1}\sum_{i=1}^n \log(x_i)\}^{-1}, \infty)$, where $x_{(n)} = \max\{x_1, \ldots, x_n\}$.

Let η be the mean of the distribution (4.1). Then, in terms of ω and ν,

$$\eta = g(\omega, \nu) = \left(\frac{1}{\omega}\right)^{1/\nu} \Gamma\left(\frac{1+\nu}{\nu}\right). \tag{4.3}$$

An estimate of the mean is therefore given by $\hat{\eta} = g(\hat{\omega}, \hat{\nu})$, where g is the "smooth" function given in (4.3).

TABLE 2
Relative efficiencies of various approximants for quantiles of a theoretical bootstrap distribution versus random uniform resampling

α	$\hat{t}_\alpha$	RIR[1]	BIR[2]	BUR[3]
0.025	−3.67	7.46	11.03	1.04
0.050	−2.33	5.13	6.10	0.97
0.100	−1.59	3.69	4.49	1.13
0.250	−0.743	3.69	3.88	1.55
0.500	−0.0401	2.17	3.37	3.37
0.750	0.656	1.77	1.97	1.70
0.900	1.25	5.13	4.97	1.84
0.950	1.64	6.49	6.78	1.59
0.975	1.98	13.82	17.09	1.34

[1]RIR = random importance resampling.
[2]BIR = balanced importance resampling.
[3]BUR = balanced uniform resampling.

Let $T = n^{1/2}(\hat{\eta} - \eta)/\hat{\sigma}$, where $\hat{\sigma}$ is given by (1.2). Table 2 summarizes the results of a simulation study comparing four approximations to the theoretical bootstrap estimate of the distribution function of T given in (2.1). The sample $\mathscr{X}$ used in the study consisted of the $n = 10$ values

$$3.13, 2.81, 1.36, 0.79, 2.25, 0.34, 1.29, 0.80, 0.28, 0.64$$

generated from an exponential distribution with mean 1 corresponding to the Weibull model (4.1) with $\omega = \nu = 1$. The values in Table 2 were obtained as follows. First, we obtained "exact" quantiles of the theoretical bootstrap distribution based on 10,000 random uniform resamples from $\mathscr{X}$. These values are denoted by $\hat{t}_\alpha$, where α is the quantile level. Then, using each method, 100 independent estimates of $\hat{t}_\alpha$ were obtained, each based on $B = 100$ resamples. Denoting these estimates by $\hat{t}_\alpha^{(i)}$, $i = 1, \ldots, 100$, the relative efficiencies of the various resampling techniques were computed as the ratio of their mean squared errors to that of random importance resampling, where

$$\text{mse} = \frac{1}{100} \sum_{i=1}^{100} \left(\hat{t}_\alpha^{(i)} - \hat{t}_\alpha\right)^2.$$

The "tilting" parameters θ for the two importance resampling methods were selected by minimizing the asymptotic variance formulae given in Sections 3.1 and 3.2 with $t = z_\alpha$, the α-level quantile of the standard normal distribution. The modification described in Section 3.3 for the case $t > 0$ was also used. In practice, if the aim is to construct a confidence interval of a prespecified level, the "optimal" value of θ may be read from Table 1. However, the minimization process is almost instantaneous on a computer and hence this is not an important consideration. An algorithm for generating balanced resamples is given in Appendix B. The algorithm appears to be both simpler and more efficient than algorithm BB3 of Gleason (1988).

There are several things of note in this example. First, there is a distinct asymmetry in the results. The theoretical bootstrap distribution is highly skewed to the left and the approximants for quantiles in the lower tail are much less accurate than those for quantiles in the upper tail. This applies to all four resampling methods. Second, the relative efficiencies of the four methods qualitatively follow those predicted by the asymptotic theory. In particular, random uniform resampling is uniformly less efficient and balanced importance sampling is uniformly more efficient than the other methods. Moreover, balancing yields large improvements towards the center of the distribution, whereas importance resampling yields large improvements in the tails. Third, there was a significant bias in the balanced importance resampling estimates in both tails of the distribution. However, despite this bias, balanced importance resampling is substantially more efficient than random importance resampling. This is an important saving when one considers that the difference in computation between producing random resamples and balanced resamples is negligible compared with the computations involved in solving the estimating equations. Finally, we note that the efficiency of balanced importance resampling relative to random importance resampling is substantially greater than that predicted by the asymptotic theory. Specifically, there is a 24% improvement when $\alpha = 0.975$ and a 48% improvement when $\alpha = 0.025$ compared with the predicted values of 2.9%. Here, for example, the predicted value was calculated using Table 1 as $100(18.03/17.53 - 1)$.

5. Summary of technical details. In this section we outline an argument which shows how results for the "smooth function model" described in Hall (1990a, b) extend to the more general setting of this paper. A similar approach can also be used to show second-order correctness, as defined by Hall (1988), of the percentile-t bootstrap in this context. The key step is to approximate the distribution of the statistic T by polynomials in the means of independent random variables, plus negligible remainder terms. Although this facility may not be immediately apparent in the example of Section 4, it is nevertheless available. For example, we may write

$$n^{-1}\sum_{i=1}^{n} x_i^{\hat{\nu}}(\log x_i)^r = \sum_{j=0}^{\infty}(j!)^{-1}(\hat{\nu}-\nu)^j n^{-1}\sum_{i=1}^{n} x_i^{\nu}(\log x_i)^{j+r}$$

$$= \sum_{j=0}^{k-1}(j!)^{-1}(\hat{\nu}-\nu)^j n^{-1}\sum_{i=1}^{n} x_i^{\nu}(\log x_i)^{j+r} + R_k,$$

where R_k denotes a generic random variable satisfying

$$P\bigl(|R_k| > n^{-(1/2)k+\varepsilon}\bigr) = O(n^{-\lambda}) \tag{5.1}$$

for all $\varepsilon, \lambda > 0$. Arguing in this way we see that the solution $\hat{\nu}$ of (4.2) may be

expressed in the form

$$\hat{\nu} - \nu = S_1 + \cdots + S_{k-1} + R_k, \tag{5.2}$$

where S_j is a polynomial of pure degree j in centered means of independent and identically distributed random variables, and R_k satisfies (5.1).

Analogous arguments allow bootstrap versions of expansions such as (5.2), having forms such as

$$\hat{\nu}^* - \hat{\nu} = S_1^* + \cdots + S_{k-1}^* + R_k^*,$$

where

$$P\big(|R_k^*| > n^{-(1/2)k+\varepsilon} | \mathscr{X}\big) = O_p(n^{-\lambda})$$

for all ε, $\lambda > 0$. Provided B increases no faster than a power of n then, by choosing k sufficiently large in expansions such as (5.2), we see that it suffices to establish the efficacy of efficient bootstrap methods (such as balanced importance resampling) for statistics which may be expressed in the form

$$S = S_1 + \cdots + S_{k-1}.$$

More generally still, the efficacy of efficient bootstrap methods may be proved for statistics S which are smooth functions of vector means. For example, using arguments similar to those of Hall (1990a), it is shown in Hall (1990b) that result (3.5) holds if S, regarded as a function of a vector of sample means, has at least three bounded derivatives of all types in a neighbourhood of the true population mean, if sufficiently many moments of the sampling distribution are finite, if the sampling distribution satisfies Cramér's continuity condition and if $B = B(n)$ increases at a rate between n^δ and n^λ where $3/2 < \delta \le \lambda < \infty$.

APPENDIX A

Let (Y_1, Y_2) be a standard bivariate normal variable with correlation coefficient ρ and let $U_i = I(Y_i + \theta \le t)\exp\{-\theta Y_i - (1/2)\theta^2\}$ for $i = 1, 2$. Let $\Phi_2(t_1, t_2; \rho) = \Pr(Y_1 \le t_1,\ Y_2 \le t_2)$. Then some straightforward algebra shows that

$$\operatorname{Cov}\{U_1, U_2\} = e^{\rho\theta^2}\Phi_2(t + \rho\theta, t + \rho\theta; \rho) - \Phi(t)^2. \tag{A.1}$$

Expanding (A.1) in a Taylor series about $\rho = 0$ we obtain

$$\operatorname{Cov}\{U_1, U_2\} = \rho\{\theta\Phi(t) + \phi(t)\}^2 + O(\rho^2), \tag{A.2}$$

where we have used the fact that $\int_{-\infty}^{t} x\phi(x)\,dx = -\phi(t)$.

APPENDIX B

Our algorithm for generating balanced importance resamples is now given.

BIR algorithm.

STEP 0. Put $m = nB + 1$.

STEP 1. Change m to $m - 1$, generate u from the Uniform(0, 1) distribution and put $c = mu$.

STEP 2. Find the k such that $\sum_{i=1}^{k-1} B_i \le c < \sum_{i=1}^{k} B_i$.

STEP 3. Put $I_m = k$ and change B_k to $B_k - 1$.

STEP 4. If $m > 1$, return to Step 1. Otherwise, stop.

The sets $\{I_1, \ldots, I_n\}, \{I_{n+1}, \ldots, I_{2n}\}, \ldots \{I_{(B-1)n+1}, \ldots, I_{Bn}\}$ contain the indicies of the B bootstrap samples of size n.

Note that, if we introduce the B_i in Step 2 in the order $B_1, B_2, \ldots$, then there should be a modest advantage in starting with the ordering $B_1 \ge B_2 \ldots \ge B_n$. Also, to avoid storing nB integers, we may return the indicies to the main program in blocks of n, provided that we keep track of the "current" values of $B_1, \ldots, B_n$.

Finally, we indicate the theoretical basis of algorithm BIR. Let $U_i = (U_{i1}, \ldots, U_{in})$, $i = 1, \ldots, nB$, denote vectors with each $U_{ij} = 0$ or 1, and $\sum_{j=1}^{n} U_{ij} = 1$ for each i. Write the multinomial distribution with probability vector $p = (p_1, \ldots, p_n)$ and index m as Mult(m, p). Suppose $U_1, \ldots, U_{nB}$ are i.i.d. Mult(1, p). Then $\sum_{i=1}^{nB} U_i$ is Mult(nB, p). First, note that

$$U_{nB} \Big| \sum_{i=1}^{nB} U_i = q \sim \text{Mult}\big(1, (nB)^{-1} q\big), \tag{B.1}$$

where $q = (B_1, \ldots, B_n)$ is given. Now

$$\begin{aligned} &\Pr(U_1 = u_1, \ldots, U_{nB} = u_{nB} | S_{nB} = q) \\ &\quad = f(u_{nB} | S_{nB} = q) \prod_{i=1}^{nB-1} f\big(u_{nB-i} | S_{nB-i} = q - \bar{S}_{nB-i+1}\big), \end{aligned} \tag{B.2}$$

where f is the generic symbol for probability mass function, $S_i = \sum_{j=1}^{i} U_j$ and $\bar{S}_i = \sum_{j=i}^{nB} u_j$. Sequential application of (B.1) to (B.2) leads to algorithm BIR.

REFERENCES

DAVISON, A. C. (1988). Discussion of paper by D. V. Hinkley. *J. Roy. Statist. Soc. Ser. B* **50** 356–357.

DAVISON, A. C., HINKLEY, D. V. and SCHECHTMAN, E. (1986). Efficient bootstrap simulation. *Biometrika* **73** 555–566.

GLEASON, J. R. (1988). Algorithms for balanced bootstrap simulations. *Amer. Statist.* **42** 263–266.

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.* **16** 927–985.

HALL, P. (1990a). Performance of the balanced bootstrap in distribution function and quantile problems. *Probab. Theory Related Fields* **85** 239–260.

HALL, P. (1990b). Balanced importance resampling for the bootstrap. Canberra Statistics Technical Report CSTR-022-90.

HINKLEY, D. V. and SHI, S. (1989). Importance resampling and the nested bootstrap. *Biometrika* **76** 435–446.

JOHNS, M. V. (1988). Importance resampling for bootstrap confidence intervals *J. Amer. Statist. Assoc.* **83** 709–714.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA ACT 2601
AUSTRALIA