The Annals of Statistics
1993, Vol. 21, No. 3, 1476–1485

AN IMPROVEMENT OF THE JACKKNIFE DISTRIBUTION FUNCTION ESTIMATOR

BY JAMES G. BOOTH AND PETER HALL

Australian National University, and Australian National University and CSIRO

In a recent paper, C. F. J. Wu showed that the jackknife estimator of a distribution function has optimal convergence rate $O(n^{-1/2})$, where n denotes the sample size. This rate is achieved by retaining $O(n)$ data values from the original sample during the jackknife algorithm. Wu's result is particularly important since it permits a direct comparison of jackknife and bootstrap methods for distribution estimation. In the present paper we show that a very simple, nonempirical modification of the jackknife estimator improves the convergence rate from $O(n^{-1/2})$ to $O(n^{-5/6})$, and that this rate may be achieved by retaining only $O(n^{2/3})$ data values from the original sample. Our technique consists of mixing the jackknife distribution estimator with the standard normal distribution in an appropriate proportion. The convergence rate of $O(n^{-5/6})$ makes the jackknife significantly more competitive with the bootstrap, which enjoys a convergence rate of $O(n^{-1})$ in this particular problem.

1. Introduction. In an interesting recent paper, Wu (1990) discussed the performance of the jackknife distribution estimator. Wu derived necessary and sufficient conditions for asymptotic normality of the estimator, and showed that the optimal convergence rate is $O(n^{-1/2})$, where n denotes the sample size. He demonstrated that this optimal rate may be achieved by retaining $O(n)$ of the data values during the "resampling without replacement" part of the jackknife algorithm.

Wu's contribution is particularly significant since it allows jackknife and bootstrap methods to be compared in the context of distribution estimation. It is known [e.g., Singh (1981) and Babu and Singh (1983, 1984)] that in a wide range of problems, the bootstrap can estimate distribution functions with an error of $O(n^{-1})$ in probability and $O\{n^{-1}(\log\log n)^{1/2}\}$ almost surely. The inability of the jackknife to do better than $O(n^{-1/2})$ must be seen as a disappointment. As Wu (1990) notes, this relatively large error is due to the fact that jackknife resampling is "without replacement," whereas the original sampling used to generate the data was "with replacement," since the effective population size was infinite. Bootstrap methods perform better than the jackknife because they preserve the "with replacement" sampling scheme.

Nevertheless, as we shall show in the present paper, all is not lost for the jackknife. A simple modification of Wu's procedure allows the convergence rate to be substantially improved, from $O(n^{-1/2})$ to $O(n^{-5/6})$. The latter rate is

Received October 1991; revised June 1992.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62D05.
Key words and phrases. Asymptotic normality, bootstrap, convergence rate, distribution estimation, Edgeworth expansion, jackknife, mixture, sampling without replacement.

available with probability 1, and is achieved by retaining $O(n^{2/3})$ of the original data values in each jackknife resample. To achieve the enhanced rate we take a jackknife estimator $\hat{G}$ of a distribution which is asymptotically $N(0,1)$, and which is computed by retaining r data values in each jackknife resample; and we correct it by mixing it with the limiting normal distribution Φ. Our final estimator is simply

$$\tilde{G} = f^{1/2}\hat{G} + (1 - f^{1/2})\Phi, \tag{1.1}$$

where $f = r/n$, and we suggest taking $r \sim \text{const.} \times n^{2/3}$. This estimator correctly allows for the $O(n^{-1/2})$ term in an Edgeworth expansion of the target distribution. In that sense it is second-order correct, and is relatively competitive with second-order correct bootstrap methods such as percentile-t and BC_a [Efron (1987) and Hall (1988)]. By way of contrast, the jackknife methods considered by Wu (1990) are only first-order correct.

An intuitive explanation of our method may be given as follows. The usual jackknife method fails for two reasons. First, it uses an effective sample size of r, rather than n. This means that the jackknife's implicit second-order correction is of size $r^{-1/2}$ rather than $n^{-1/2}$. Second, the form of the implicit second-order correction is not quite right, being in error by terms of size $r/n + n^{-1/2}$. If we mix the jackknife and normal approximations in the proportion $(r/n)^{1/2} : \{1 - (r/n)^{1/2}\}$, we effectively replace $r^{-1/2}$ by $n^{-1/2}$ in the implicit second-order correction. And if we also ask that $r/n \to 0$, we ensure that the form of the second-order correction term is asymptotically right. By considering third-order terms we may show that the most appropriate size of r is $r \sim \text{const.} \times n^{2/3}$.

Of course, the enhanced jackknife convergence rate of $O(n^{-5/6})$ still falls short of the bootstrap rate, and the precise error of the jackknife method is affected by the choice of r. This tuning parameter must also be selected for Wu's (1990) uncorrected jackknife. For these reasons we tend to favour the bootstrap approach, although we agree that the jackknife algorithm has advantages [Wu (1990), pages 1450–1451).

It should be noted that our results are only significant when the second-order term in an Edgeworth expansion of the true distribution does not vanish. In the example of estimating the distribution of a mean, this is equivalent to asking that skewness (called κ_3 in Section 2) be nonzero. Otherwise, the jackknife approximation proposed by Wu (1990) is in error by only $O(n^{-1})$, and the accuracy $O(n^{-5/6})$ achieved by our recommendation is not the best possible.

Section 2 outlines the argument which motivates the corrected estimator defined at (1.1), and Section 3 summarizes a simulation study which demonstrates, in a particular problem where n is small, that our method improves on the uncorrected jackknife but falls short of bootstrap accuracy. Thus, the asymptotic theory appears to accurately predict small sample behaviour. Finally, in Section 4 we describe and illustrate a procedure for efficient approximation to quantiles of jackknife distribution estimators.

2. Methodology and main results. We treat initially the case of the Studentized mean, and then outline results in a more general context.

Let $\mathscr{X} = \{X_1, \dots, X_n\}$ denote a random sample drawn from a continuous, univariate distribution with finite fourth moment. Denote the population mean and variance by μ and σ^2, and let the sample mean and variance be

$$\bar{X} = n^{-1} \sum_{i=1}^{n} X_i, \qquad \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

We wish to estimate the distribution G of $T = n^{1/2}(\bar{X} - \mu)/\hat{\sigma}$:

$$\begin{aligned} G(x) = P(T \le x) &= \Phi(x) + n^{-1/2} p_1(x)\phi(x) \\ &\quad + n^{-1} p_2(x)\phi(x) + o(n^{-1}), \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} p_1(x) &= \tfrac{1}{6}\kappa_3(2x^2 + 1), \\ p_2(x) &= x\left\{\tfrac{1}{12}\kappa_4(x^2 - 3) - \tfrac{1}{18}\kappa_3^2(x^4 + 2x^2 - 3) - \tfrac{1}{4}(x^2 + 3)\right\} \end{aligned}$$

and κ_j denotes the jth cumulant of $(X_1 - \mu)/\sigma$.

The jackknife algorithm based on retaining r of the original n data values in each resample may be described as follows. Let $\mathscr{X}^* = \{X_1^*, \dots, X_r^*\}$ denote a resample of size $r < n$, drawn randomly *without* replacement from $\mathscr{X}$. Write $\bar{X}^* = r^{-1}\Sigma X_i^*$ and $\hat{\sigma}^{*2} = r^{-1}\Sigma(X_i^* - \bar{X}^*)^2$ for the mean and variance of $\mathscr{X}^*$. Put $f = r/n$, and observe that $E(\bar{X}^* | \mathscr{X}) = \bar{X}$, $\mathrm{var}(\bar{X}^* | \mathscr{X}) = (1 - f)(1 - n^{-1})^{-1} r^{-1} \hat{\sigma}^2$. Thus, $T^* = r^{1/2}(\bar{X}^* - \bar{X})/\{(1 - f)^{1/2}\hat{\sigma}^*\}$ is the version of T in the sampling-without-replacement problem. The jackknife estimator of G is

$$\hat{G}(x) = P(T^* \le x | \mathscr{X}).$$

Existing theory of Edgeworth expansion for sampling without replacement [e.g., Robinson (1978) and Babu and Singh (1985)] provides the formula

$$P(T^* \le x | \mathscr{X}) = \Phi(x) + r^{-1/2}\hat{q}_1(x)\phi(x) + o(r^{-1/2}) \tag{2.2}$$

with probability 1, where

$$\hat{q}_1(x) = \tfrac{1}{6}\hat{\kappa}_3\left\{3x^2 - (1 - f)^{-1}(1 - 2f)(x^2 - 1)\right\}(1 - f)^{1/2} \tag{2.3}$$

and $\hat{\kappa}_3 = \hat{\sigma}^{-3} n^{-1} \Sigma(X_i - \bar{X})^3$. Indeed, the techniques of Babu and Singh (1985) permit (2.2) to be extended to a third term, so that

$$\begin{aligned} \hat{G}(x) = P(T^* \le x | \mathscr{X}) &= \Phi(x) + r^{-1/2}\hat{q}_1(x)\phi(x) \\ &\quad + r^{-1}\hat{q}_2(x)\phi(x) + o(r^{-1}) \end{aligned} \tag{2.4}$$

with probability 1. Sufficient regularity conditions are that $E|X_1|^{8+\varepsilon} < \infty$ for some $\varepsilon > 0$, that X_1 have a continuous distribution, and that $r, n \to \infty$ together in such a manner that f is bounded away from 1. The polynomial $\hat{q}_2$ is odd and of degree 5, and its coefficients are bounded with probability 1.

In the event that $f \to 0$, resampling without replacement converges to resampling with replacement. This is reflected in the fact that the polynomials

$\hat{q}_j$ in expansion (2.4) converge to their counterparts in (2.1). For example, we know from (2.3) that, since $\hat{\kappa}_3 = \kappa_3 + O(n^{-1/2+\varepsilon})$ with probability 1 for each $\varepsilon > 0$,

$$\hat{q}_1(x) = p_1(x) - \tfrac{1}{4}\kappa_3 f + O(n^{-1/2+\varepsilon} + f^2). \tag{2.5}$$

Similarly,

$$\hat{q}_2(x) = p_2(x) + o(1). \tag{2.6}$$

We may now deduce from (2.4) that the jackknife estimator $\hat{G}$ satisfies

$$\begin{aligned}\hat{G}(x) = \Phi(x) &+ r^{-1/2}p_1(x)\phi(x) + r^{-1}p_2(x)\phi(x)\\ &- r^{1/2}n^{-1}\tfrac{1}{4}\kappa_3\phi(x) + O(r^{-1/2}n^{-1/2+\varepsilon} + r^{3/2}n^{-2})\\ &+ o(r^{-1})\end{aligned} \tag{2.7}$$

with probability 1, for each $\varepsilon > 0$.

Combining (2.1) and (2.7) we see that the corrected jackknife estimator $\tilde{G}$, given by (1.1), satisfies

$$\begin{aligned}\tilde{G}(x) - G(x) = (rn)^{-1/2}p_2(x)\phi(x) &- rn^{-3/2}\tfrac{1}{4}\kappa_3\phi(x)\\ &+ O(n^{-1+\varepsilon} + r^2n^{-5/2}) + o\{(rn)^{-1/2}\}.\end{aligned} \tag{2.8}$$

Since (2.1) and (2.7) are available uniformly in x then so is (2.8).

The absolute value of the right-hand side of (2.8) is minimized by taking $r \sim cn^{2/3}$, where $c = c(x)$ is selected to minimize $|c^{-1/2}p_2(x) - (1/4)c\kappa_3|$ subject to $c > 0$. Depending on the signs of $p_2(x)$ and k_3 it may be theoretically possible to render this quantity equal to zero by a particularly judicious choice of c. However, such an approach requires careful empirical selection of r and is not very attractive. More generally, taking r to equal the integer part of $Cn^{2/3}$ for any fixed $C > 0$, we see from (2.8) that

$$\sup_{-\infty<x<\infty} |\tilde{G}(x) - G(x)| = O(n^{-5/6}) \tag{2.9}$$

with probability 1.

Under similar assumptions [specifically, that the distribution of X_1 has finite $(8 + \varepsilon)$th moment for some $\varepsilon > 0$, and is continuous], the usual percentile-t bootstrap approximation to the distribution G is in error by $O(n^{-1})$ in probability and $O\{n^{-1}(\log\log n)^{1/2}\}$ with probability 1; see for example Singh (1981) and Babu and Singh (1983, 1984). Therefore, the approximation at (2.9) is not quite as good as the bootstrap approximation, but it is nevertheless a substantial improvement on the best possible approximation obtainable by directly applying the jackknife approximation $\hat{G}$: for $\kappa_3 \neq 0$,

$$\inf_{1\le r\le n-1}\ \sup_{-\infty<x<\infty} |\tilde{G}(x) - G(x)| \sim c_0 n^{-1/2}, \tag{2.10}$$

where

$$\begin{aligned}c_0 = \tfrac{1}{6}|\kappa_3| \inf_{0<f<1}\ \sup_{-\infty<x<\infty} &|(2x^2+1) + (f^{-1}-1)^{1/2}\\ &\times\{(1-f)^{-1}(1-2f)(x^2-1) - 3x^2\}|\phi(x).\end{aligned} \tag{2.11}$$

The infimum in (2.10) is achieved asymptotically by taking $r \sim nf_0$, where f_0 attains the infimum in (2.11).

In this discussion we confined attention to the Studentized case, since this is often useful in applications. For example, it may be employed directly to construct confidence intervals and hypothesis tests. There is an analogue of our technique in the non-Studentized case, which more closely parallels the line of development taken by Wu (1990). There, we define $T = n^{1/2}(\bar{X} - \mu)/\sigma$ and $T^* = r^{1/2}(\bar{X}^* - \bar{X})/\{(1-f)^{1/2}\hat{\sigma}\}$. As before, but with this new notation, define $G(x) = P(T \le x)$ and $\hat{G}(x) = P(T^* \le x | \mathscr{X})$, and let $\tilde{G}$ be given by (1.1). Each of (2.7)–(2.9) has a direct analogue in this setting, albeit with different polynomials $p_1, p_2, \hat{q}_1, \hat{q}_2$. For example, the new version of p_1 is $p_1(x) = (1/6)\kappa_3(1 - x^2)$. Again, (2.9) holds if $r \sim Cn^{2/3}$ for some $C > 0$.

Many other contexts are identical in principle to that of the Studentized mean, differing only in the form of the polynomials $p_1, p_2, \hat{q}_1, \hat{q}_2$. We shall consider the so-called "smooth function model" for valid Edgeworth expansions, first brought to prominence by Bhattacharya and Ghosh (1978). There it is assumed that the sample $\mathscr{X} = \{X_1, \dots, X_n\}$ is d-variate, that the univariate parameter $\theta = g(\mu)$ is a smooth function of the vector mean $\mu = E(X_1)$, and that the estimator $\hat{\theta} = g(\bar{X})$ is the same function of the sample mean. Assuming nondegeneracy and sufficiently many moments of the sampling distribution, $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically normal $N(0, \sigma^2)$, where $\sigma^2 = h(\mu)$ can also be expressed as a smooth function of μ. (This may require an extension of each vector X_j, to incorporate products of earlier components.) Take $\hat{\sigma}^2 = h(\bar{X})$ and $T = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$. As Bhattacharya and Ghosh (1978) noted, expansion (2.1) is valid for quite general statistics T of this form, provided we assume adequate moment and smoothness conditions on the underlying population. The polynomial p_j is of degree $3j - 1$ and of opposite parity to j.

Let $\mathscr{X}^* = \{X_1^*, \dots, X_r^*\}$ denote a resample of size $r < n$, drawn randomly but without replacement from $\mathscr{X}$. Put $\bar{X}^* = r^{-1}\Sigma X_i^*$, $\hat{\theta}^* = g(\bar{X}^*)$, $\hat{\sigma}^{*2} = h(\bar{X}^*)$ and $f = r/n$. Then $E(\bar{X}^*|\mathscr{X}) = \bar{X}$ and

$$\begin{aligned}\operatorname{var}(\bar{X}^*|\mathscr{X}) &= (1-f)(1-n^{-1})(nr)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T \\ &= (1-f)r^{-1}\{\operatorname{var}(X_1) + O_p(n^{-1/2})\},\end{aligned}$$

where all sides of the displayed identities are $d \times d$ matrices. From these results, and the fact that $g(\bar{X}^*) - g(\bar{X})$ is (by Taylor expansion) approximately linear in $\bar{X}^* - \bar{X}$, we see that conditional on $\mathscr{X}$, $T^* = r^{1/2}(\hat{\theta}^* - \hat{\theta})/\{(1-f)^{1/2}\hat{\sigma}^*\}$ is asymptotically $N(0,1)$. The conditional distribution function $\hat{G}$ of T^* admits the expansion (2.4), in which the polynomials $\hat{q}_1$ and $\hat{q}_2$ satisfy analogues of (2.5) and (2.6):

$$\begin{aligned}\hat{q}_1(x) &= p_1(x) + fp(x) + O(n^{-1/2+\varepsilon} + f^2), \\ \hat{q}_2(x) &= p_2(x) + o(1),\end{aligned}$$

where p is an even polynomial of degree 2.

It is now straightforward to develop versions of (2.7)–(2.9). In particular, if the corrected estimator $\tilde{G}$ is defined by (1.1) and if $r \sim Cn^{2/3}$ then (2.9) holds.

3. A simulation study. Table 1 summarizes the results of a simulation study designed to compare the performances of the uncorrected and corrected jackknife distribution function estimators described in Sections 1 and 2 in a particular problem. Results are also given for the analogous percentile-t bootstrap estimator. [See Hall (1992), page 15] for an introduction to the percentile-t bootstrap.] The problem considered is that of estimating quantiles of the distribution of the Studentized mean, $T = n^{1/2}(\bar{X} - \mu)/\hat{\sigma}$, of a random sample of size $n = 16$ from a chi-squared distribution with $\mu = 1$ degree of freedom. Quantile estimates are required, for example, to construct confidence intervals for μ and may be obtained by numerically inverting the appropriate distribution function estimate. For example, the corrected jackknife α-quantile estimate is given by $\tilde{t}_\alpha = \tilde{G}^{-1}(\alpha)$.

TABLE 1

Means, mean squared errors and empirical error rates of uncorrected- and corrected-jackknife and percentile-t bootstrap quantile estimators

		Left tail			Right tail		
α	r	mean	mse	error rate	mean	mse	error rate
0.025	8	−6.20	15.0	0.024	1.65	0.0621	0.013
		−5.41	9.65	0.033	1.73	0.0532	0.010
	9	−5.36	7.80	0.030	1.62	0.0581	0.013
		−4.85	5.36	0.040	1.69	0.0468	0.010
	10	−5.12	6.83	0.036	1.58	0.0628	0.015
		−4.68	5.62	0.044	1.65	0.0487	0.013
	11	−4.55	4.13	0.045	1.54	0.0689	0.018
		−4.29	3.80	0.051	1.59	0.0534	0.015
	12	−4.47	4.61	0.048	1.47	0.0845	0.025
		−4.33	4.48	0.049	1.51	0.0685	0.020
	Bstrap	−4.42	3.81	0.047	1.73	0.0493	0.009
	Exact*	−4.49		0.024	1.58		0.022
0.05	8	−4.40	5.72	0.050	1.43	0.0364	0.033
		−3.83	3.70	0.062	1.48	0.0375	0.024
	9	−4.23	5.28	0.056	1.41	0.0338	0.035
		−3.53	3.03	0.073	1.45	0.0338	0.027
	10	−3.69	3.02	0.068	1.40	0.0359	0.041
		−3.38	2.56	0.075	1.43	0.0344	0.033
	11	−3.63	3.29	0.072	1.37	0.0384	0.039
		−3.42	3.01	0.078	1.40	0.0355	0.036
	12	−3.48	3.17	0.081	1.33	0.0431	0.043
		−3.06	2.24	0.091	1.35	0.0393	0.041
	Bstrap	−3.26	1.83	0.078	1.45	0.0280	0.034
	Exact*	−3.40		0.045	1.34		0.047

*Based on 10^6 simulated values of $T = n^{1/2}(\bar{X} - \mu)/\hat{\sigma}$.

In this particular study 1000 random samples of size $n = 16$ were generated. Uncorrected and corrected quantile estimates were obtained from each sample based on 1000 without replacement resamples of sizes ranging from $r = 8$ to $r = 12$. Percentile-t bootstrap estimates were also computed for each sample based on 1000 with replacement resamples of size 16.

The "left tail" values in Table 1 are the means, mean squared errors and empirical error rates of the various α-quantile estimates. The "right tail" values are the analogous results for the $(1 - \alpha)$-quantile estimates. Results are given for $\alpha = 0.025$ and $\alpha = 0.05$. Upper and lower values for each value of r correspond to uncorrected and corrected jackknife estimates, respectively. The empirical error rates in the left and right portions of the table are, respectively, the proportion of times T fell below and above the quantile estimate in 1000 simulations.

It can be seen from Table 1 that the corrected jackknife estimator is almost uniformly more accurate than its uncorrected counterpart. This holds for all resample sizes r; the only exception being in the right tail when $\alpha = 0.05$ and $r = 8$. However, the accuracy of the corrected jackknife estimator typically falls short of that of the percentile-t bootstrap estimator, as predicted by the asymptotic theory. In this example, the performances of both the uncorrected and corrected jackknife methods vary considerably with the resample size r. We note, however, that the accuracy of jackknife methods based on half-samples ($r = 8$) is substantially less than optimal.

4. Efficient jackknife resampling. In this section we propose a method for reducing the number of without-replacement resamples required for accurate approximation to jackknife quantile estimators. The method is closely related to importance resampling for the bootstrap, introduced by Johns (1988) and Davison (1988). However, complications arise in the jackknife situation because, unlike bootstrap resamples, the components of jackknife resamples are dependent.

Exact evaluation of the jackknife distribution function estimate $\hat{G}$ in Section 2 requires computation of the statistic T^* for all $\binom{n}{r}$ possible resamples, a task which may often be impractical. Thus, in practice we approximate $\hat{G}$ by

$$\hat{G}_B(t) = \frac{1}{B} \sum_{b=1}^{B} I(T_b^* \le t),$$

where T_b^*, $b = 1, \dots, B$, are the values of T^* computed from B independently selected simple random samples of size r from $\mathscr{X}$. Notice that $\hat{G}_B(t)$ is an unbiased estimator of $\hat{G}(t)$ for all t.

One way of obtaining a simple random resample of size r from $\mathscr{X}$ is via an r-stage sequential procedure in which a value X_{i_j}, say, is selected from $\mathscr{X}$ at the jth stage whose subscript differs those selected at each of the preceding stages. At the first stage a value is chosen at random from $\mathscr{X}$ in such a way that the ith element has probability $\pi_{1i}^{\dagger} = n^{-1}$ of being the one selected. At the jth stage, $j = 2, \dots, r$, an element is chosen from $\mathscr{X} \setminus \{X_{i_1}, \dots, X_{i_{j-1}}\}$

with a remaining value X_i, $i \notin \{i_1, \dots, i_{j-1}\}$ having probability $\pi_{ji}^\dagger = (n-j+1)^{-1}$ of being selected. Under this scheme, each element of $\mathscr{X}$ has probability $f = r/n$ of being in any resample. A modified form of the above scheme in which elements of $\mathscr{X}$ have unequal probabilities of selection is obtained by changing the selection probabilities at the first stage to $\pi_{1i}^\dagger = \pi_i$, $i = 1, \dots, n$, where the π_i's are positive and sum to 1, and by setting the selection probabilities at the jth subsequent stage equal to

$$\pi_{ji}^\dagger = \frac{\pi_i}{1 - \sum_{k=1}^{j-1} \pi_{i_k}}$$

if $i \notin \{i_1, \dots, i_{j-1}\}$ and $\pi_{ji}^\dagger = 0$ otherwise. Notice that the modified resampling scheme reduces to simple random sampling if $\pi_i = n^{-1}$ for all i. Also, for unequal π_i's the selection probabilities $\pi_{ji}^\dagger$ at the jth stage are random variables.

Now, if $T^\dagger = r^{1/2}(\hat{\theta}^\dagger - \hat{\theta})/\{(1-f)^{1/2}\hat{\sigma}^\dagger\}$ where $\hat{\theta}^\dagger$ and $\hat{\sigma}^\dagger$ are computed using a resample selected according to the modified scheme just described, then

$$E\left\{ I(T^\dagger \le t) \prod_{j=1}^{r} (n-j+1)\pi_{ji_j}^\dagger | \mathscr{X} \right\} = \hat{G}(t)$$

for all t. Thus, an entire family of unbiased approximations for $\hat{G}(t)$ has the form

$$\hat{G}_B(t;\pi) = \frac{1}{B} \sum_{b=1}^{B} I(T_b^\dagger \le t) \prod_{j=1}^{r} (n-j+1)\pi_{ji_{j(b)}}^\dagger, \tag{4.1}$$

where $i_{j(b)}$ is the subscript of the element of $\mathscr{X}$ selected at the jth stage in the bth resample.

Importance resampling for the bootstrap works by "tilting" the selection probabilities so that the mean value of the appropriate bootstrap version of $T^\dagger$ is approximately equal to the quantile of interest. An intuitive rationale for this approach is that quantiles close to the mode of a distribution can be estimated more accurately than those in the tails. In the jackknife setting this centering of the distribution of $T^\dagger$ may be achieved as follows. We confine ourselves to the "smooth function model" described in Section 2. Let Y_i^* be the indicator that the ith element of $\mathscr{X}$ is a member of a jackknife resample and let $g' = (g_1, \dots, g_p)^T$ where g_j is the derivative of g with respect to its jth component. Then, under the assumptions of Section 2, $T^* = \sum_{i=1}^n \varepsilon_i Y_i^* + O_p(r^{-1/2})$, where

$$\varepsilon_i = \{r(1-f)\hat{\sigma}^2\}^{-1/2} g'(\bar{X})^T (X_i - \bar{X}), \qquad i = 1, \dots, n.$$

Now, set $\pi_i = n^{-1}\exp\{\eta\varepsilon_i - K(\eta)\}$ where $K(\eta) = \log(n^{-1}\sum_{i=1}^n e^{\eta\varepsilon_i})$. Then it follows that $T^\dagger$ is approximately $N(\eta, 1)$ as $r \to \infty$. Since $T^* \sim N(0,1)$, the α-quantile of T^* is to first order equal to z_α, the α-quantile of the standard normal distribution. Thus, to approximate quantiles $\hat{t}_\alpha = \hat{G}^{-1}(\alpha)$ in the lower tail of the distribution of T^* we set $\eta = z_\alpha$ and solve the equation in t:

$\hat{G}_B(t;\pi) = \alpha$. On the other hand, to approximate corrected jackknife α-quantile estimators $\tilde{t}_\alpha = \tilde{G}^{-1}(\alpha)$ we set $\eta = z_\alpha$ and solve

$$f^{1/2}\hat{G}_B(t;\pi) + (1 - f^{1/2})\Phi(t) = \alpha.$$

A refinement of this approach is to choose η to minimize the asymptotic variance of $\hat{G}_B(z_\alpha;\pi)$. This leads to the same formula for the "optimal" η as in bootstrap importance resampling. In practice, however, the asymptotically optimal value of η differs little from z_α. See, for example, Booth, Hall and Wood (1993) for a simple discussion of importance resampling for bootstrap quantile estimation.

The asymmetry of the tilting procedure about $t = 0$ has the result that (4.1) is more efficient for approximating quantiles in the lower tail. To approximate uncorrected jackknife quantiles in the upper tail we first approximate $\hat{H}(t) = P(-T^* \le -t|\mathscr{X})$ by $\hat{H}_B(t;\pi)$, say, and then set $\hat{G}_B(t;\pi) = 1 - \hat{H}_B(t;\pi)$. Analogous modifications apply for the corrected jackknife.

For a numerical illustration of this procedure we use the following sample of $n = 25$ paired data values from Davison and Hinkley (1988):

X	1546	505	280	410	1450	517	738	2225	1660	505	680	145	224
Y	155	68	28	25	190	82	92	196	164	68	82	36	195
X	733	1957	287	473	473	1260	1958	4375	1499	245	828	6361	
Y	92	185	61	62	71	207	185	475	155	29	96	699	

The problem we consider is that of approximating the quantiles of the distribution of the Studentized ratio of means $\hat{\theta} = \hat{X}/\hat{Y}$. In this case,

$$\varepsilon_i = \left\{r(1-f)\hat{\sigma}^2\right\}^{-1/2}\left\{X_i - \left(\bar{X}/\bar{Y}\right)Y_i\right\}\bar{Y}^{-1}, \qquad i = 1,\dots,n$$

and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n\{X_i - (\bar{X}/\bar{Y})Y_i\}^2\bar{Y}^{-2}$.

Table 2 summarizes the results for the case $r = 15$. The values given are the means and mean squared errors of 100 independent approximations to

TABLE 2

Means and mean squared errors of approximations to corrected jackknife quantile estimators for a Studentized ratio. The case $\eta = 0$ corresponds to approximation based on simple random sampling, whereas $\eta = z_\alpha = \Phi^{-1}(\alpha)$ corresponds to importance resampling

		$\eta = 0$		$\eta = z_\alpha$	
α	Exact* $\tilde{t}_\alpha$	mean	100 × mse	mean	100 × mse
0.005	−2.26	−2.20	10.8	−2.14	2.51
0.01	−2.02	−2.01	6.67	−1.90	3.50
0.025	−1.68	−1.70	2.62	−1.60	3.35
0.05	−1.42	−1.41	1.52	−1.39	3.94
0.1	−1.14	−1.13	0.878	−1.14	1.73
0.9	2.28	2.30	11.6	2.31	2.74
0.95	3.10	3.09	15.1	3.10	3.89
0.975	3.77	3.73	20.4	3.76	4.86
0.99	4.43	4.45	28.7	4.46	4.91
0.995	4.85	4.81	33.4	4.85	4.83

*Based on 10^5 simulated values of T^*.

"exact" corrected jackknife quantiles obtained using simple random sampling and importance resampling. Rows correspond to different values of α, with $\alpha < 1/2$ indicating the left tail and $\alpha > 1/2$ the right tail. The performance of the ordinary approximation to the jackknife estimate, based on simple random sampling, is described in the column headed $\eta = 0$. The efficient jackknife method is depicted in the column headed $\eta = z_\alpha$. Each approximation is based on $B = 100$ resamples. The "exact" values were obtained by simulating T^* 10^5 times. The table indicates that in this example, importance resampling greatly increases the efficiency of the approximations to quantiles in the right tail of the distribution where the approximations based on simple random sampling are highly inaccurate. In the left tail, importance resampling is more efficient for small values of α and is slightly less efficient nearer the centre of the distribution, although the approximants are more accurate than their right tail counterparts.

Acknowledgments. We are grateful to two referees for their helpful comments.

REFERENCES

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BABU, G. J. and SINGH, K. (1984). On one-term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46** 219–232.

BABU, G. J. and SINGH, K. (1985). Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17** 261–278.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of formal Edgeworth expansions. *Ann. Statist.* **6** 434–451.

BOOTH, J. G., HALL, P. and WOOD, A. T. A. (1993). Balanced importance resampling for the bootstrap. *Ann. Statist.* **21** 286–298.

DAVISON, A. C. (1988). Discussion of paper by D. V. Hinkley. *J. Roy. Statist. Soc. Ser. B* **50** 256–257.

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

EFRON, B. (1987). Better bootstrap confidence intervals (with discussion). *J. Amer. Statist. Assoc.* **82** 171–185.

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

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

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

ROBINSON, J. (1978). An asymptotic expansion for samples from a finite population. *Ann. Statist.* **6** 1005–1011.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

WU, C. F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18** 1438–1452.

DEPARTMENT OF STATISTICS
FLOYD HALL
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

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