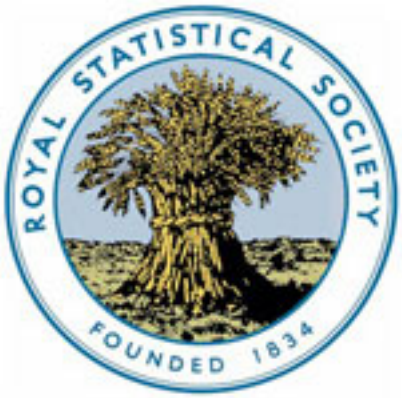

J. R. Statist. Soc. B (1988)
50, *No.* 3, *pp.* 338–354

A Review of Bootstrap Confidence Intervals

By THOMAS J. DICICCIO and JOSEPH P. ROMANO†

Stanford University, USA

[*Read before the* Royal Statistical Society *on Wednesday, March 16th, 1988 at a meeting organized by the* Birmingham Group, *Professor* J. B. Copas *in the Chair*]

SUMMARY

A survey of bootstrap procedures for constructing confidence regions is given. In particular, several distinct bootstrap methods are considered, with emphasis on the mathematical correctness of these procedures. The percentile, bias-corrected percentile and accelerated bias-corrected percentile methods, developed by Efron, are reviewed in both parametric and nonparametric situations. A procedure related to the accelerated bias-corrected method, which avoids explicit calculation of the analytical corrections required in Efron's method, is also introduced. In the context of a functional approach for the construction of confidence regions, the bootstrap is motivated as a method to estimate the distributions of approximate pivots. Finally, iterative bootstrap methods are discussed as means to improve coverage accuracy.

Keywords: ASYMPTOTIC THEORY; BOOTSTRAP; CONFIDENCE REGIONS; PERCENTILE METHOD; BIAS-CORRECTED PERCENTILE METHOD; ACCELERATED BIAS-CORRECTED PERCENTILE METHOD; PERCENTILE-*t*; PIVOT; LEAST FAVOURABLE FAMILY; ORTHOGONAL PARAMETERS; SECOND-ORDER ACCURACY; PREPIVOTING

1. INTRODUCTION

This paper is a survey of bootstrap procedures for constructing confidence regions for parameters of interest. Such procedures rely on estimating the sampling distribution of a statistic or an approximate pivot. In general, bootstrap methods consist of estimating a characteristic of the unknown population by simulating the characteristic when the true population is replaced by an estimated one. The appeal of this approach is its wide applicability to complex data structures in both parametric and nonparametric problems. Several distinct bootstrap methods will be reviewed, with the emphasis on the mathematical correctness of these procedures.

In Section 2, Efron's percentile, bias-corrected (BC) percentile and accelerated bias-corrected (BC_a) percentile methods are developed in both parametric and nonparametric situations. These procedures arise from transformation theory considerations, and they have the property of invariance under reparameterization. The most promising of these techniques is the BC_a, which depends on the calculation of the acceleration constant. A general formula is given for this analytical adjustment for situations other than maximum likelihood estimation. A procedure related to the accelerated bias-corrected method is also introduced which avoids explicit calculation of analytical corrections required in Efron's method. These methods are compared and the percentile-*t* method is considered in some numerical examples.

In Section 3, some bootstrap methods, including the percentile-*t*, are motivated as a functional approach to the construction of confidence regions by using the bootstrap to estimate the distribution of approximate pivots. The mathematical analysis of these

† *Address for correspondence*: Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305–4065, USA.

 0035–9246/88/50338 $2.00

methods is outlined, including conditions required to justify them and a discussion of examples where they lead to inconsistencies.

Section 4 focuses on refined bootstrap methods with the goal being improved coverage accuracy. In particular, iterative methods proposed by Beran, Loh and Hall are discussed. Beran's method of prepivoting and Loh's calibrated confidence sets are seen to be equivalent. These methods offer the potential for increased accuracy of coverage for confidence sets, but their computational feasibility needs to be established for them to be considered viable general approaches.

In any given situation, the choice of bootstrap procedure depends on available theoretical results, computational considerations, the level of accuracy desired, simulation results and experience with similar problems. For example, both the BC_a and the percentile-t are second order correct; however, the BC_a requires knowledge of an analytical constant while the percentile-t requires a stable estimate of variance. Given the diversity of criteria in choosing a procedure, it is unlikely that a single procedure will emerge as a preferred method in all problems.

A common feature of all the procedures considered is the use of simulation to approximate a sampling distribution by treating an estimate of the population as the true one. The computational aspects of this problem are not treated here, but the reader is referred to section 3 of Hinkley (1988) and Johns (1988).

2. PERCENTILE METHOD AND RELATED PROCEDURES

2.1. *Introduction*

In a series of articles, Efron (1981, 1982, 1985, 1987) has introduced and refined the percentile method of using bootstrap calculations to set approximate confidence limits for scalar parameters. These refinements of the percentile method are the bias-corrected (BC) percentile method and the accelerated bias-corrected (BC_a) percentile method. Efron's approach is to first develop these procedures in the simple context of a parametric model indexed by a scalar parameter, for which there are no nuisance parameters present, and then to adapt them for application in multiparameter families and nonparametric situations. This development relies on transformation theory, and the resulting procedures have the desirable property of invariance under reparameterization.

For a review of the percentile method in the simplest case, suppose that $x_n = (X_1, \ldots, X_n)$ is a sample from a distribution having probability density function f_θ which depends upon the scalar parameter of interest θ. Let $\hat{\theta}$ be an estimator of θ based on x_n with distribution function $G_\theta(s) = P_\theta(\hat{\theta} \leqslant s)$. The exact upper $1-\alpha$ confidence limit for θ is taken to be that value $\theta[1-\alpha]$ satisfying $G_{\theta[1-\alpha]}(\hat{\theta}) = \alpha$, and the bootstrap distribution for $\hat{\theta}$ is $G_{\hat{\theta}}$. Now suppose there exists a monotonically increasing transformation g and a constant τ such that for all values of θ

$$\tau\{g(\hat{\theta}) - g(\theta)\} \sim Z, \tag{2.1}$$

where Z is symmetrically distributed about zero with distribution function H. Then

$$G_\theta(s) = H[\tau\{g(s) - g(\theta)\}];$$

hence, in terms of the α quantile $z^{(\alpha)} = H^{-1}(\alpha)$,

$$\theta[1-\alpha] = g^{-1}(g(\hat{\theta}) + z^{(1-\alpha)}/\tau)$$

and

$$G_\theta^{-1}(\alpha) = g^{-1}(g(\hat{\theta}) + z^{(\alpha)}/\tau).$$

Since $\theta[1-\alpha] = G_\theta^{-1}(1-\alpha)$, the confidence limits for θ can be obtained directly from the bootstrap distribution for $\hat{\theta}$ without explicit knowledge of the transformation g.

The quantity $G_{\hat{\theta}}^{-1}(1-\alpha)$ is called the percentile bootstrap confidence limit. Typically, no transformation g exists for which (2.1) obtains exactly, and the difference between $G_{\hat{\theta}}^{-1}(1-\alpha)$ and $\theta[1-\alpha]$ is $O_p(n^{-1})$. Thus the percentile method can give poor approximations in small sample situations. For example, consider $n = 8$ observations from a bivariate normal distribution with known means and variances and unknown correlation coefficient θ. For each of the values 0, 0.3 and 0.8 of the usual estimator r, the exact upper and lower 97.5% confidence limits for θ are compared with the corresponding percentile bootstrap limits in Table 1.

One approach for improvement of the percentile method is to account for bias in (2.1). Assume then that there exists a monotonically increasing transformation g and constants τ and z_0 such that for all θ

$$\tau\{g(\hat{\theta}) - g(\theta)\} + z_0 \sim Z,$$

where Z is as described above. In this case, $G_\theta(s) = H(\tau\{g(\hat{\theta}) - g(\theta)\} + z_0)$, from which it follows that

$$\theta[1-\alpha] = g^{-1}(g(\hat{\theta}) + (z^{(1-\alpha)} + z_0)/\tau)$$

and

$$G_\theta^{-1}(\alpha) = g^{-1}(g(\hat{\theta}) + (z^{(\alpha)} - z_0)/\tau).$$

Since $\theta[1-\alpha] = G_\theta^{-1}\{H(z^{(1-\alpha)} + 2z_0)\}$, the confidence limits for θ can be obtained from the bootstrap distribution for $\hat{\theta}$. The bias correction z_0 can be determined similarly by $z_0 = H^{-1}\{G_\theta(\hat{\theta})\}$. Efron (1982) uses the standard normal distribution function Φ for H, and he calls $G_\theta^{-1}\{\Phi(z^{(1-\alpha)} + 2z_0)\}$ the bias-corrected percentile bootstrap confidence limit. If $z_0 = 0$, then the BC method reduces to the percentile method. The upper and lower 97.5% BC limits are shown for the correlation coefficient example in Table 1.

The BC limits, like the percentile limits, typically differ from the exact ones by terms of order $O_p(n^{-1})$. In the correlation coefficient example the BC method produces fairly accurate approximations; however, for an example in which the BC method performs poorly, consider a sample of size $n = 5$ from the exponential distribution with mean θ, and take $\hat{\theta}$ to be the sample mean. Table 2 compares the exact upper and lower 97.5% confidence limits for θ with the corresponding approximations obtained by the percentile and BC methods. Although use of the bias correction produces an improvement over the percentile method, the BC limits are inadequate. This example is similar to one considered by Schenker (1985), which in part motivated the development of the BC_a method.

To improve the BC method, Efron (1987) supposes there exists a monotonically increasing transformation g and constants τ, z_0 and a such that for all values of θ

$$\tau\left\{\frac{g(\hat{\theta}) - g(\theta)}{1 + a\tau g(\theta)}\right\} + z_0 \sim Z, \tag{2.2}$$

TABLE 1
Upper and lower 97.5% confidence limits for the correlation coefficient ($n = 8$)

r	z_0	*Exact*		*Percentile*		*BC*		*(2.4)*		*Percentile-t*	
		Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
0	0	−0.666	0.666	−0.707	0.707	−0.707	0.707	−0.659	0.659	−1.412	1.412
		(2.50)	(2.50)	(1.66)	(1.66)	(1.66)	(1.66)	(2.68)	(2.68)		
0.3	−0.0611	−0.479	0.797	−0.495	0.838	−0.539	0.819	−0.475	0.792	−1.342	1.257
		(2.50)	(2.50)	(2.24)	(1.30)	(1.61)	(1.80)	(2.56)	(2.71)		
0.8	−0.1656	0.199	0.952	0.307	0.966	0.153	0.954	0.183	0.951	−0.107	0.996
		(2.50)	(2.50)	(4.39)	(0.91)	(1.96)	(2.19)	(2.30)	(2.59)	(0.47)	(0.00)

Below each lower limit θ_L and each upper limit θ_U is shown $1 - G_{\theta_L}(r)$ and $G_{\theta_U}(r)$, respectively, expressed as a percentage.

TABLE 2
Upper and lower 97.5% confidence limits for the exponential mean ($n = 5$)

Exact		*Percentile*		*BC*		BC_a	
Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
$0.488\hat{\theta}$	$3.080\hat{\theta}$	$0.325\hat{\theta}$	$2.048\hat{\theta}$	$0.390\hat{\theta}$	$2.270\hat{\theta}$	$0.488\hat{\theta}$	$3.083\hat{\theta}$
(2.50)	(2.50)	(0.06)	(10.11)	(0.43)	(7.28)	(2.50)	(2.49)

Below each confidence limit is shown its true error rate in coverage expressed as a percentage. $z_0 = 0.1497$, $a = 0.1491$.

where Z is as previously described. Under this assumption

$$G_\theta(s) = H\left[\tau\left\{g(s) - g(\theta)\right\} \Big/ \left\{1 + a\tau g(\theta)\right\} + z_0\right],$$

so that

$$\theta[1-\alpha] = g^{-1}\left[g(\hat{\theta}) + \frac{(z^{(1-\alpha)} + z_0)\{1 + a\tau g(\hat{\theta})\}}{\tau\{1 - a(z^{(1-\alpha)} - z_0)\}}\right]$$

and

$$G_\theta^{-1}(\alpha) = g^{-1}\left[g(\hat{\theta}) + \frac{(z^{(\alpha)} - z_0)\{1 + a\tau g(\hat{\theta})\}}{\tau}\right].$$

The confidence limits for θ can be obtained from the bootstrap distribution for $\hat{\theta}$, since

$$\theta[1-\alpha] = G_\theta^{-1}\left[H\left\{z_0 + \frac{(z^{(1-\alpha)} + z_0)}{1 - a(z^{(1-\alpha)} + z_0)}\right\}\right], \tag{2.3}$$

and $z_0 = H^{-1}\{G_\theta(\hat{\theta})\}$. Using the standard normal distribution function for H, Efron calls (2.3) the accelerated bias-corrected percentile bootstrap confidence limit. The BC_a method reduces to the BC method if $a = 0$.

Efron (1987) shows that for suitable choice of a the BC_a confidence limits are second order correct, which means that the BC_a limits differ from the corresponding

exact limits by terms of order $O_p(n^{-3/2})$. Unlike z_0, the acceleration adjustment a cannot be easily determined from the bootstrap distribution $G_{\hat{\theta}}$. If to error of order $O(n^{-1})$ the first three cumulants of $n^{1/2}(\hat{\theta}-\theta)$ are $n^{-1/2}\lambda_1(\theta)$, $\lambda_2(\theta)$ and $n^{-1/2}\lambda_3(\theta)$, and if

$$\gamma = n^{-1/2}\left\{3\frac{\lambda_2'}{\lambda_2^{1/2}} - 2\frac{\lambda_3}{\lambda_2^{3/2}}\right\}$$

and

$$\delta = n^{-1/2}\left\{\frac{\lambda_3}{\lambda_2^{3/2}} - 6\frac{\lambda_1}{\lambda_2^{1/2}}\right\},$$

where $\lambda_2' = \mathrm{d}\lambda_2(\theta)/\mathrm{d}\theta$, then a should be chosen to satisfy $a = \gamma(\hat{\theta})/6 + O_p(n^{-1})$, while $z_0 = \delta(\hat{\theta})/6 + O_p(n^{-1})$. When $\hat{\theta}$ is the maximum likelihood estimator, γ and δ are both equal to the skewness of the score function for θ based on x_n. DiCiccio and Tibshirani (1987) consider the construction of a transformation g which approximately satisfies (2.2).

As shown in Table 2, the $\mathrm{BC_a}$ method is very accurate in the exponential mean example. In this case, $\gamma = \delta = 2/n^{1/2}$. For the correlation coefficient example, these formulae give $\gamma = 0$ and $\delta = -3\theta/n^{1/2}$. Thus in this example, $a = 0$ and the BC and $\mathrm{BC_a}$ limits are equal.

2.2. *Multiparameter Families*

Suppose that $x_n = (X_1, \ldots, X_n)$ is a sample from a distribution having probability density function f_η which depends upon a vector parameter $\eta = (\eta^1, \ldots, \eta^p)$. Let $\hat{\eta}$ be an estimator of η based on x_n, and suppose that the scalar parameter $\theta = t(\eta)$ is of interest. Let the distribution function of the estimator $\hat{\theta} = t(\hat{\eta})$ be $G_\eta(s) = P_\eta(\hat{\theta} \leqslant s)$; then the bootstrap distribution of $\hat{\theta}$ is $G_{\hat{\eta}}$.

Having calculated the bootstrap distribution $G_{\hat{\eta}}$, the upper $1-\alpha$ percentile bootstrap confidence limit for θ is $G_{\hat{\eta}}^{-1}(1-\alpha)$, and the BC limit is $G_{\hat{\eta}}^{-1}\{\Phi(z^{(1-\alpha)} + 2z_0)\}$, where $z_0 = \Phi^{-1}\{G_{\hat{\eta}}(\hat{\theta})\}$. The implementation of the $\mathrm{BC_a}$ method is less straightforward in multiparameter situations because of complications arising in the calculation of the acceleration adjustment a. In this case, Efron (1987) restricts attention to maximum likelihood estimators.

To introduce the notation required for Efron's formulation of a, let $l(\eta; x_n)$ be the log-likelihood function for η based on x_n, and let $\kappa_{ij} = E\{l_i l_j\}$ and $\kappa_{ijk} = E\{l_i l_j l_k\}$, where $l_i = \partial l(\eta; x_n)/\partial \eta^i$. For brevity of notation in the expressions that follow, the usual convention is used whereby summation is understood over indices that appear as both subscripts and superscripts. Take $(\kappa^{ij}) = (\kappa_{ij})^{-1}$, $t_i = \partial t(\eta)/\partial \eta^i$, and set $\mu^i = \kappa^{ij} t_j$. Efron reduces the multiparameter family to a scalar parameter one by restricting attention to the line $\eta(\tau) = \hat{\eta} + \tau\hat{\mu}$, called the least favourable family, where $\hat{\mu} = (\hat{\mu}^1, \ldots, \hat{\mu}^p)$ and $\hat{\mu}^i = \mu^i(\hat{\eta})$. In analogy with the scalar parameter case, Efron takes $\gamma(\tau)$ to be the skewness of $\partial l(\eta(\tau); x_n)/\partial\tau$ evaluated at $\eta(\tau)$, so that $\gamma(0) = \hat{\kappa}_{ijk}\hat{\mu}^i\hat{\mu}^j\hat{\mu}^k/(\hat{\kappa}_{ij}\hat{\mu}^i\hat{\mu}^j)^{3/2}$, and then he recommends using $a = \gamma(0)/6$. Alternatively, observed information can be used in place of expected information for the calculation of a.

The difference between the true and nominal coverage levels of the percentile and BC confidence limits is typically $O(n^{-1/2})$, and for the $\mathrm{BC_a}$ method this difference is

$O(n^{-1})$. For multiparameter families, comparison of the approximate limits with exact ones is difficult, since a definition for exactness is not clear cut. See Bickel (1987) and Hall (1988) for further discussion concerning accuracy of the BC_a limits. Efron shows that, in certain circumstances, the BC_a limits agree to order $O_p(n^{-1})$ with limits derived by Cox (1980) and McCullagh (1984).

It is possible to obtain an expression for a which is appropriate for more general estimators. Let $\delta^i = n^{1/2}(\hat{\eta}^i - \eta^i)$, and suppose that to error of order $O(n^{-1})$, $E(\delta^i) = n^{-1/2}\lambda^i$, $\operatorname{cov}(\delta^i, \delta^j) = \lambda^{ij}$, and $\operatorname{cum}(\delta^i, \delta^j, \delta^k) = n^{-1/2}\lambda^{ijk}$, where $\operatorname{cum}(\delta^i, \delta^j, \delta^k)$ is the third order cumulant of δ^i, δ^j and δ^k. Let

$$\gamma(\eta) = n^{-1/2}\{(3\lambda_l^{ij}\lambda^{lk} - 2\lambda^{ijk})t_i t_j t_k\}/(\lambda^{ij}t_i t_j)^{3/2},$$

where $\lambda_l^{ij} = \partial\lambda^{ij}(\eta)/\partial\eta^l$; then a should be chosen to satisfy $a = \gamma(\hat{\eta})/6 + O_p(n^{-1})$.

For an example that illustrates the accuracy of the various approximate procedures, consider a sample from the bivariate distribution having probability density function $f(x, y) = \eta^1(\eta^2)^c \exp\{-(\eta^1 x + \eta^2 y)\}/\{y^{1-c}\Gamma(c)\}$ for $x > 0$, $y > 0$, and let $\theta = \eta^2/\eta^1$. In this case, $\gamma = 2(c-1)/\{n(c^2+c)\}^{1/2}$. Table 3 compares the upper and lower 97.5% percentile, BC and BC_a limits with the exact limits in the case $n = 5$ and $c = 0.4$. The BC_a method is very accurate in this situation. In the case $c = 1$, the percentile limits are exact.

2.3. *Related Procedures*

The BC_a method has been criticized because it is not fully automatic; that is, it requires the calculation of the analytic adjustment a. DiCiccio and Romano (1987) have considered related procedures which closely approximate the BC_a method, but do not require the explicit calculation of adjustments like z_0 and a.

Consider the simple case of a family indexed by a scalar parameter θ, and suppose there exists a transformation satisfying (2.2). The exact upper $1-\alpha$ confidence limit $\theta[1-\alpha]$ can be found by the formula

$$G_{\hat{\theta}}^{-1}\{G_{\theta_0'}(\theta_0)\}, \tag{2.4}$$

where θ_0 is any value of the parameter θ and $\theta_0' = G_{\theta_0}^{-1}(\alpha)$. Moreover, formula (2.4) is often exact when (2.2) is not exactly satisfied. As an example, this formula produces exact confidence limits for θ when $\hat{\theta}/\theta$ is pivotal; the BC_a method is not exact in this case.

In practice, a reasonable choice for the initial value θ_0 is the percentile limit $G_{\hat{\theta}}^{-1}(1-\alpha)$. Table 1 shows the approximate limits obtained by using this choice and

TABLE 3

Upper and lower 97.5% confidence limits for $\theta = \eta^2/\eta^1$ ($n = 5$)

Exact		*Percentile*		*BC*		*BC_a*	
Lower	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
$0.113\hat{\theta}$	$4.468\hat{\theta}$	$0.224\hat{\theta}$	$8.844\hat{\theta}$	$0.184\hat{\theta}$	$6.506\hat{\theta}$	$0.105\hat{\theta}$	$4.404\hat{\theta}$
(2.50)	(2.50)	(8.10)	(0.25)	(5.87)	(0.76)	(2.18)	(2.61)

Below each confidence limit is shown its true error rate in coverage expressed as a percentage. $z_0 = -0.1217$, $a = -0.1195$.

(2.4) for the correlation coefficient example. In the exponential mean example, (2.4) is exact. An important feature of this procedure is that it can be iterated according to $\theta_{i+1} = G_{\theta}^{-1}\{G_{\theta_i'}(\theta_i)\}(i = 0, 1, \ldots)$, where $\theta_i' = G_{\theta_i}^{-1}(\alpha)$. Under certain conditions the difference between θ_i and $\theta[1-\alpha]$ is $O_p(n^{-1-i/2})$. Moreover, the numbers of calculations involved in this iterative procedure are linear in i, unlike other iterative methods discussed in Section 4.

For the extension of this method to multiparameter families, suppose that the model has been parameterized so as to be indexed by (θ, ψ), where $\psi = (\psi^1, \ldots, \psi^{p-1})$ consists of nuisance parameters orthogonal to θ, i.e. $\operatorname{cov}\{n^{1/2}(\hat{\theta} - \theta), n^{1/2}(\hat{\psi}^i - \psi^i)\} = O(n^{-1})$. Cox and Reid (1987) give a detailed account of orthogonal parameters with reference to maximum likelihood estimation. To implement the method, commence with some initial value θ_0, perhaps the percentile limit $\theta_0 = G_{\hat{\theta},\hat{\psi}}^{-1}(1-\alpha)$, let $\theta_0' = G_{\theta_0,\hat{\psi}}^{-1}(\alpha)$, and then take $\theta_1 = G_{\hat{\theta},\hat{\psi}}^{-1}\{G_{\theta_0,\hat{\psi}}(\theta_0)\}$. The difference between the $\mathrm{BC_a}$ limit and θ_1 is $O_p(n^{-3/2})$, and the error in coverage of the approximate limit θ_1 is $O(n^{-1})$. However, further iteration in the multiparameter case does not improve coverage accuracy.

In practice, it can be inconvenient to determine an orthogonal parameterization directly, and the preceding procedure can be sufficiently well approximated by using the least favourable family. In terms of the notation introduced previously for an estimator $\hat{\eta}$, consider the line $\eta(\tau) = \hat{\eta} + \tau\hat{\mu}$, where $\mu^i = \lambda^{ij}t_j$, $\hat{\mu}^i = \mu^i(\hat{\eta})$, and $\hat{\mu} = (\hat{\mu}^1, \ldots, \hat{\mu}^p)$. Let $G_\tau(s) = P_{\hat{\eta}+\tau\hat{\mu}}(\hat{\theta} \leqslant s)$, so that G_0 is the bootstrap distribution for $\hat{\theta} = t(\hat{\eta})$. Starting with an initial value θ_0, perhaps taken to be the percentile limit $G_0^{-1}(1-\alpha)$, the approximate limit is $G_0^{-1}\{G_{\tau_0}(\theta_0)\}$, where τ_0 is the value of τ such that $\theta_0 = t\{\eta(\tau)\}$, $\theta_0' = G_{\tau_0}^{-1}(\alpha)$, and τ_0' is the value of τ satisfying $\theta_0' = t\{\eta(\tau)\}$. This approximate limit differs from the one obtained using the orthogonal parameterization by terms of order $O_p(n^{-3/2})$.

For the multiparameter problem considered in Table 3, this procedure is exact. However, in the case of a sample drawn from the normal distribution with unknown mean μ and variance σ^2 where μ is the parameter of interest, the percentile method and related procedures perform poorly for small sample sizes. The approximate limits for μ given by these methods are the exact ones that would be obtained if the variance was known and equal to $\hat{\sigma}^2$.

A procedure which does give the correct limits in the normal mean example is the percentile-t method. This method, further discussed in Sections 3 and 4, makes use of the bootstrap distribution of an approximately pivotal quantity instead of using the bootstrap distribution for $\hat{\theta}$. In the multiparameter context, consider the approximate pivot $(\hat{\theta} - \theta)/(\hat{\lambda}^{ij}\hat{t}_i\hat{t}_j)^{1/2}$, and let K_η be its distribution function. Thus $K_\eta(s) = P_\eta\{(\hat{\theta} - \theta)/(\hat{\lambda}^{ij}\hat{t}_i\hat{t}_j)^{1/2} \leqslant s\}$, and the bootstrap distribution is $K_{\hat{\eta}}$. The percentile-t approximation to an upper $1-\alpha$ confidence limit for θ is $\hat{\theta} - (\hat{\lambda}^{ij}\hat{t}_i\hat{t}_j)^{1/2}K_{\hat{\eta}}^{-1}(\alpha)$. The difference between the true and nominal coverage levels of the percentile-t limits is $O(n^{-1})$. Although these limits are not invariant under reparameterization, the difference between the percentile-t and $\mathrm{BC_a}$ limits is $O_p(n^{-3/2})$. For the examples considered in Tables 2 and 3, the percentile-t method produces exact confidence limits. However, for the correlation coefficient example, with $\operatorname{var}(\hat{\theta}) = (1-\theta^2)^2/(n-1) + O(n^{-2})$, this method produces poor approximations as shown in Table 1.

The percentile-t method can be extended to the construction of approximate confidence regions for vector-valued parameters of interest, which has been discussed by Hall (1988); such an extension for the $\mathrm{BC_a}$ method has yet to be as fully developed.

For the case of maximum likelihood estimation, it may be appealing to Studentize by using observed rather than expected information in deriving the percentile-t limits. Beran (1987) considers the use of bootstrap distributions for log-likelihood ratio test statistics.

2.4. *Nonparametric Inference*

Suppose that $x_n = (X_1, \ldots, X_n)$ is a sample from an unknown distribution F, and suppose that $\theta = T(F)$ is the scalar parameter of interest. Let $\hat{F}_n$ be the empirical distribution function of x_n, and consider the estimator $\hat{\theta} = T(\hat{F}_n)$ of θ. Let $G_F(s) = P_F(\hat{\theta} \leqslant s)$ be the distribution function of $\hat{\theta}$ under F; then the bootstrap distribution for $\hat{\theta}$ is $G_{\hat{F}_n}$.

The percentile and BC limits are easily found in this situation. The approximate upper $1-\alpha$ confidence limits for θ obtained by the percentile and BC methods are $G_{\hat{F}_n}^{-1}(1-\alpha)$ and $G_{\hat{F}_n}^{-1}\{\Phi(z^{(1-\alpha)} + 2z_0)\}$, respectively, where $z_0 = \Phi^{-1}\{G_{\hat{F}_n}(\hat{\theta})\}$. The implementation of the $\mathrm{BC_a}$ method is less straightforward since it requires the calculation of the adjustment a.

To calculate a, Efron reduces the nonparametric situation to a multiparameter one by restricting attention to distribution functions with support on $X_1, \ldots, X_n$. Corresponding to each such distribution F is a vector $w = (w^1, \ldots, w^n)$, where w^i is the probability mass assigned to X_i. Efron considers the problem of setting confidence limits for $\theta = \theta(w)$ having made n draws from such an n-category multinomial distribution and observed each category to appear once; that is $\hat{w} = (n^{-1}, \ldots, n^{-1})$. The least favourable family in this situation is $w(\tau) = (w^1(\tau), \ldots, w^n(\tau))$, where

$$w^i(\tau) = \mathrm{e}^{\tau U_i} \Big/ \left\{ \sum_{j=1}^{n} \mathrm{e}^{\tau U_j} \right\}$$

and

$$U_i = \lim_{\Delta \to 0} \frac{T\{(1-\Delta)\hat{F}_n + \Delta\delta_i\} - T(\hat{F}_n)}{\Delta}$$

is the ith component of the empirical influence function with δ_i denoting a point mass at X_i. The formula for the acceleration adjustment is

$$a = \left(\sum_{i=1}^{n} U_i^3 \right) \Big/ \left\{ 6 \left(\sum_{i=1}^{n} U_i^2 \right)^{3/2} \right\}.$$

Efron suggests the U_is be calculated numerically, and he has used $\Delta = 0.001$ in practice.

Hall (1988) has considered these procedures in the nonparametric context of a 'smooth function' model for which the estimator can be expressed as a function of multivariate vector means. In such cases, the difference between the nominal and true coverage levels of the approximate limits is $O(n^{-1/2})$ for the percentile and BC methods and is $O(n^{-1})$ for the $\mathrm{BC_a}$ method. Hall (1988) also considers the percentile-t method for this model, and he shows that the percentile-t limits differ from the $\mathrm{BC_a}$ limits by terms of order $O_P(n^{-3/2})$. Hall (1987) considers the percentile-t method for vector-valued parameters in the nonparametric context.

The procedure given by expression (2.4) and extended to the multiparameter families in Section 2.3 can be applied in an obvious way to the nonparametric case by making use of the least favourable family $w(\tau)$.

3. FUNCTIONAL APPROACH

3.1. *Consistency*

In this section, the bootstrap is motivated as a natural functional approach to the construction of a confidence region. The development begins by focusing on the independent identically distributed case.

Let $x_n = (X_1, \ldots, X_n)$ be a sample of n random variables taking values in a sample space S and having unknown distribution F, where F is assumed to belong to a certain collection $\mathbf{F}$ of distributions. The collection $\mathbf{F}$ may be finite or infinite dimensional. The interest lies in constructing a confidence interval for some parameter $T(F)$, whose range $\{T(F)\colon F \in \mathbf{F}\}$ will be denoted $\mathbf{T}$. This leads to considering a root $R_n(x_n, T(F))$, which is just some functional depending on both x_n and $T(F)$. For example, an estimator T_n of a real-valued parameter $T(F)$ might be given so that a natural choice is $R_n(x_n, T(F)) = T_n - T(F)$, or alternatively, $R_n(x_n, T(F)) = [T_n - T(F)]/s_n$, where s_n is some estimate of the standard deviation of T_n.

When $\mathbf{F}$ is suitably large, a natural construction for an estimator T_n of $T(F)$ is $T_n = T(\hat{F}_n)$, where $\hat{F}_n$ is the empirical measure of $X_1, \ldots, X_n$. In regular parametric problems for which $\mathbf{F}$ is indexed by a parameter η belonging to a subset of $\mathbb{R}^p$, $T(F)$ can be described as a parameter $t(\eta)$, and hence T_n is often taken to be $T_n = t(\hat{\eta}_n)$, where $\hat{\eta}_n$ is some desirable estimate of η, such as a maximum likelihood estimate, a one-step maximum likelihood estimate, a minimum distance estimate, etc.

Let $J_n(F)$ be the law of $R_n(x_n, T(F))$ when $x_n = (X_1, \ldots, X_n)$ is a random sample from F, and let $J_n(x, F)$ be the corresponding cumulative distribution function. Also, let $J_n^{-1}(\alpha, F) = \inf\{x : J_n(x, F) \geqslant \alpha\}$ be an α quantile of the law $J_n(F)$. In order to construct a confidence region for $T(F)$, the sampling distribution or the appropriate quantiles of $J_n(F)$ must be known or estimated. The bootstrap procedure is to estimate $J_n(F)$ by $J_n(\hat{G}_n)$, where $\hat{G}_n$ is some estimate of F, and then estimate the appropriate quantiles of $J_n(F)$ by those of $J_n(\hat{G}_n)$. In nonparametric problems, $\hat{G}_n$ is typically (but not always) taken to be the empirical distribution $\hat{F}_n$; in parametric problems, $\hat{G}_n$ is usually $G_{\hat{\eta}_n}$. A resulting bootstrap confidence region for $T(F)$ takes the form

$$B_n(\alpha, x_n) = \{t \in \mathbf{T} : R_n(x_n, t) < J_n^{-1}(1-\alpha, \hat{G}_n)\}. \tag{3.1}$$

Note that when $J_n(F)$ is independent of F, the root $R_n(x_n, T(F))$ is said to be a pivot, in which case the bootstrap procedure is clearly valid. In general, it may not be possible to find an exact pivot. Some of the bootstrap literature refers to this procedure of estimating $J_n(F)$ by $J_n(\hat{G}_n)$ as the bootstrap pivotal method, but the term root is adapted here from Beran (1987) to distinguish a general root from a pivot defined in the classical sense.

How well does this bootstrap procedure work? Is it consistent, or perhaps optimal in any sense? Typically, one can show $J_n(F)$ converges weakly to a continuous limit law $J(F)$. In order for the bootstrap to be valid, $J_n(F)$ must be smooth in F. Smoothness in F can often be described in terms of a suitable metric d, depending on the choice of root R_n. Specifically, let d be a metric on $\mathbf{F}$ such that the estimate $\hat{G}_n$ of F satisfies $d(\hat{G}_n, F) \to 0$ in probability as $n \to \infty$. Then, a sufficient condition for the bootstrap to be consistent is that the convergence of $J_n(F)$ to $J(F)$ be locally uniform in F, where uniformity is described in terms of d. Formally, the following uniform weak convergence result (denoted TAC from Beran (1987) for triangular array convergence) must hold.

Triangular Array Convergence. If $\{F_n, F_n \in \mathbf{F}\}$ is any sequence of distributions in $\mathbf{F}$ satisfying $d(F_n, F) \to 0$, then $J_n(F_n)$ converges weakly to a continuous limit law $J(F)$, which depends only on F.

When TAC holds and $d(\hat{G}_n, F) \to 0$ in probability, it follows that

$$\sup_x |J_n(x, \hat{G}_n) - J_n(x, F)| \to 0 \text{ in probability.} \tag{3.2}$$

Moreover, the continuity of the limit law $J(F)$ entails the convergence in probability of $J_n^{-1}(\alpha, \hat{G}_n)$ to the α quantile, $J^{-1}(\alpha, F)$ of $J(F)$. It follows (see Beran (1984a), theorem 1) that a bootstrap confidence region B_n given by (3.1) satisfies, for any F in $\mathbf{F}$,

$$P_F\{T(F) \in B_n(\alpha, x_n)\} \to 1 - \alpha \quad \text{as} \quad n \to \infty. \tag{3.3}$$

Several papers in the bootstrap literature show that the TAC condition holds. See, for example, Babu and Singh (1983), Beran (1984a), Beran and Millar (1985), Beran and Srivastava (1985), Bickel and Freedman (1981), Ducharme *et al.* (1985), Freedman (1981), Romano (1987) and Singh (1981). Two well-studied examples are the following.

Example 3.1. Let $\mathbf{F}$ be the collection of distributions on the line with finite variance. The problem is to construct a confidence interval for $T(F) = \int x \, dF(x)$. Here,

$$R_n(x_n, F) = n^{1/2} |T(\hat{F}_n) - T(F)|,$$

where $\hat{F}_n$ is the empirical distribution so that $T(\hat{F}_n)$ is the sample mean. Then, TAC holds when d is, for example, Mallow's metric d_2 defined by: $d_2^2(F, G)$ is the infimum of $E(X - Y)^2$ over all joint distributions of X and Y whose fixed marginals are F and G, respectively. Convergence of $d(F_n, F)$ to 0 is equivalent to F_n converging weakly to F and $\mathrm{Var}(F_n) \to \mathrm{Var}(F)$. See Bickel and Freedman (1981).

Example 3.2. Let $\mathbf{F}$ be the collection of all distributions F on the line. The problem is to construct a confidence region for F based on the root

$$R_n(x_n, F) = n^{1/2} \sup_t |\hat{F}_n(t) - F(t)|,$$

where $\hat{F}_n$ is the empirical measure based on $x_n = (X_1, \ldots, X_n)$. Then, the TAC holds with

$$d(F, G) = \sup_t |F(t) - G(t)|,$$

the usual Kolmogorov distance. This example was first considered in Bickel and Freedman (1981) and has been generalized to constructing a confidence set for a measure in an arbitrary sample space S based on a supremum distance over a Vapnik–Cervonenkis class of sets; see Beran and Millar (1987).

Based on the validity of the bootstrap in example 2, it is natural to expect the asymptotic correctness of bootstrap confidence intervals for smooth functionals $T(F)$ of F, since the distribution of $n^{1/2}\{T(\hat{F}_n) - T(F)\}$ can often be approximated by some smooth functional of the empirical process. In particular, suppose T is a real-valued functional defined on a large class $\mathbf{F}$ of distribution functions on the real line. Assume that T is Fréchet differentiable; that is, for each fixed F, there exists a function ψ_F

such that

$$T(G) - T(F) = \int \psi_F \mathrm{d}(G - F) + o(|G - F|_\infty)$$

where $|G - F|_\infty$ is the supremum norm between distribution functions. Suppose further that

$$\int \psi_F^2 \, \mathrm{d}F < \infty \tag{3.4}$$

and

$$\int (\psi_G - \psi_F)^2 \mathrm{d}G = O(|G - F|_\infty). \tag{3.5}$$

Proposition 3.1. If T is Fréchet differentiable with derivative ψ_F satisfying (3.4) and (3.5), then TAC holds for the root

$$R_n(x_n, T(F)) = n^{1/2}\{T(\hat{F}_n) - T(F)\}$$

when $d(F, G) = |F - G|_\infty$. Hence, (3.2) and (3.3) hold when resampling from $\hat{G}_n = \hat{F}_n$, the empirical distribution.

An argument for this proposition is essentially given in Bickel and Freedman (1981). In fact, the result holds under the weaker assumption that T is compactly differentiable with derivative ψ_F satisfying (3.4) and (3.5). See Liu *et al.* (1986) for a discussion of the negligibility of the remainder term, although the TAC condition is not made explicit. Proposition 3.1 applies to many M estimators, L estimators and U statistics, for example.

Extensions of the bootstrap outside of the independent and identically distributed framework are clearly possible. Given a root $R(x_n, T(F))$ based on data x_n with distribution F, let $J_n(F)$ be the distribution of the root. As before, the bootstrap procedure is to estimate $J_n(F)$ by $J_n(\hat{G})$ for some suitable estimate $\hat{G}$. Freedman (1981) discusses regression models, Freedman (1984) and Haycock (1986) discuss time series models, and Sugahara (1987) discusses Markov chain models. The simplest example outside of the independent and identically distributed context is confidence intervals for two-sample problems.

Example 3.3. Let x_n be a random sample of size n from F and let y_m be an independent random sample of size m from G. The problem is to construct a confidence interval for $T(F) - T(G)$ where T is some functional (such as the mean or a quantile). Let

$$R_{nm}(x_n, y_m, T(F), T(G)) = [T(\hat{F}_n) - T(\hat{G}_m)] - [T(F) - T(G)],$$

where $\hat{F}_n$ is an estimate of F based on x_n and $\hat{G}_m$ is an estimate of G based on y_m. Let $J_{nm}(F, G)$ be the law of $R_{nm}(x_n, y_m, T(F), T(G))$ when x_n is a sample of size n from F and y_m is a sample of size m from G. The bootstrap procedure approximates the appropriate quantiles of $J_{nm}(F, G)$ by those of $J_{nm}(\hat{F}_n, \hat{G}_m)$. As in proposition 3.1, smoothness assumptions on T yield consistency by taking $\hat{F}_n$ and $\hat{G}_m$ the empirical distributions corresponding to x_n and y_m. With respect to example 3.2, one may also construct a confidence band for the difference $F - G$; see Beran (1984a).

3.2. *Optimality*

We have seen that smoothness of $J_n(F)$ in F entails the consistency of the bootstrap. A natural question to ask is whether the bootstrap estimate $J_n(\hat{G}_n)$ of $J_n(F)$ is optimal in any sense. In nonparametric problems where $\mathbf{F}$ is the class of all distribution functions, it is well known that the empirical distribution function $\hat{F}_n$ is an optimal estimator of F in a local asymptotic minimax (LAM) sense. One would expect similar optimality results for smooth functionals of F, such as $J_n(F)$; such results are well established in the literature, with a unifying approach developed in Millar (1983). In a decision-theoretic framework, Beran (1982) establishes the LAM property for the bootstrap distribution estimate $J_n(\hat{F}_n)$, assuming a locally uniform first-order Edgeworth expansion for $J_n(F)$ (which entails a certain differentiability property in F for $J_n(F)$). Bootstrap estimates of bias, variance and skewness, which can be viewed as functionals of $J_n(F)$, are also typically LAM; see Beran (1984a) for details.

Having derived optimality properties for the bootstrap estimate $J_n(\hat{F}_n)$ of $J_n(F)$, a natural question is: do the resulting confidence sets possess any optimality property? In Beran and Millar (1985), a general asymptotic theory of optimal confidence sets is presented. In their framework, confidence sets should not only have the approximate stated level, but should also be properly centred and small in size. In particular, suppose the problem is to construct a confidence set for $T(F)$, where T takes values in a Banach space B with norm $|\cdot|_B$. If $\hat{C}_n$ is a confidence set for $T(F)$, the risk of $\hat{C}_n$ is given by

$$\mathbf{E}_F\{\sup\, g(|\, t - T(F)\,|_B) : t \in \hat{C}_n\}$$

for some increasing function g. The goal is to find a procedure $\hat{C}_n$, a ball in B specified by its centre and radius, that achieves the minimum LAM risk subject to the constraint on level. Uner suitable conditions, the confidence set

$$\hat{C}_n = \{t : |\, t - T(\hat{F}_n)\,|_B \leqslant \hat{r}_n\}$$

is LAM, where $\hat{r}_n$ is determined by bootstrapping the root $R_n(x_n, T(F)) = |\, T(\hat{F}_n) - T(F)\,|_B$. Their framework applies to both nonparametric and parametric problems.

3.3. *Inconsistency*

The bootstrap need not be consistent, even in the weakest sense (3.3). The problem is due to lack of uniformity in F of $J_n(F)$. Bickel and Freedman (1981) give two such examples. Other counterexamples may be found in Beran (1982) and Athreya (1987). Ghosh *et al.* (1984) discuss the bootstrap estimate of the variance of the sample median $T(\hat{F}_n)$, where $T(F) = \inf\{t : F(t) > 1/2\}$. They show it is possible for (3.3) to hold based on the root

$$R_n(x_n, T(F)) = n^{1/2}\{T(\hat{F}_n) - T(F)\};$$

however, the bootstrap estimate of variance of $R_n(x_n, F)$ is inconsistent. Although Studentization typically improves coverage accuracy, their results suggest that using a root based on Studentization can be inappropriate. The problem lies in the fact that the variance functional is not weakly continuous. The following example shows the problems that can occur when constructing bootstrap confidence intervals for functionals of a density.

Example 3.4. Given a sample $x_n = (X_1, \ldots, X_n)$ of n observations from a distribution F having density f, the problem is to construct a confidence interval for $T(F) = f(t)$ for some fixed t. Let $\hat{f}_{nh}(t)$ be a kernel density estimate of f given by

$$\hat{f}_{nh}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right)$$

for some appropriate smooth kernel K, and let $\hat{F}_{nh}$ be the corresponding distribution satisfying $\hat{F}_{nh}^{(1)} = \hat{f}_{nh}$. The density f is assumed to be smooth (say, having two continuous derivatives) but is otherwise unknown. Let

$$R_{nh}(x_n, T(F)) = (nh)^{1/2}\{\hat{f}_{nh}(t) - f(t)\}.$$

Let $J_{nh}(F)$ be the distribution of $R_{nh}(x_n, T(F))$ under F. The optimal rate at which $h = h_n$ should be chosen is well known to be $nh_n^5 \to q$ for some $q > 0$. For such a choice of h_n, the following uniformity result holds (which can easily be translated into the TAC condition). If G_n is a distribution having density g_n such that the first two derivatives of g converge uniformly to those of f in a neighbourhood of t, then $J_{nh_n}(G_n)$ converges weakly to a continuous limit law $J(F)$. Hence, if $nb_n^5/\log(n) \to \infty$ and $b_n \to 0$, then $J_{nh_n}(\hat{F}_{nb_n})$ is a consistent bootstrap estimate so that (3.2) and (3.3) hold. On the other hand, when $b_n = h_n$, the second derivative of the kernel density estimate $\hat{f}_{nh_n}$ is not a consistent estimate of $f^{(2)}$ and failure of the bootstrap results. Unfortunately, the appropriate class of distributions which one may resample from often depends heavily on the parameter of interest when this parameter is a local property of the distribution. This emphasizes the need to be careful in a naive application of the bootstrap when bootstrapping functionals of a density; see Romano (1988).

Finally, we remark that bootstrap confidence intervals may be consistent in the sense that (3.3) holds for any fixed F, but the convergence is not uniform over F. That is, define the level of a confidence set $\hat{C}_n$ for $T(F)$ to be

$$\inf_F P_F\{T(F) \in \hat{C}_n\}.$$

For fixed F, one may have $P_F(T(F) \in \hat{C}_n)$ tends to $1 - \alpha$ (or at least $1 - \alpha$), but the level converges to a number less than $1 - \alpha$. The following example, adapted from Romano (1986) shows this can happen even in a parametric setting; see Romano (1986) for a nonparametric example.

Example 3.5. Let $x_n = (X_1, \ldots, X_n)$ be independent Bernoulli variables such that $P_\theta(X_i = 1) = \theta$. The problem is to construct a confidence interval for θ. Let

$$R_n(x_n, \theta) = n^{1/2}(\hat{\theta}_n - \theta)$$

where $\hat{\theta}_n$ is the sample mean. Let $J_n(\theta)$ be the distribution of $R_n(x_n, \theta)$ under θ, and let B_n be the bootstrap confidence set

$$B_n(\alpha, x_n) = \{\theta: R_n(x_n, \theta) \leqslant J_n^{-1}(1 - \alpha, \hat{\theta}_n)\}.$$

Fix any α in (0, 1), and β in $(\alpha, 1)$ and choose $\theta_n > 0$ so $(1 - \theta_n)^n \geqslant \beta$. If x_n is a sample from P_{θ_n}, the chance that all observations are zero is at least β. In such a case, the resulting bootstrap confidence set is just the set $\{0\}$ and does not contain the true

value θ_n. It follows that

$$\limsup_{n\to\infty} P_{\theta_n}\{\theta \in B_n(\alpha, x_n)\} \leqslant 1-\beta.$$

For fixed α, β was arbitrary, so in fact the level tends to zero.

4. ITERATIVE BOOTSTRAP REFINEMENTS

4.1. *Introduction and Studentized Roots*

Bootstrap methods discussed in Section 3 offer a widely applicable construction of confidence sets. In this Section, we review some refinements of the method, with the goal of improved coverage accuracy. Other methods, based in part on analytical corrections, are discussed in Abramovitch and Singh (1985), Hall (1983) and Withers (1983).

In order to construct a bootstrap confidence set (3.1), a choice of root must be specified. In particular, consider the root

$$R_n(x_n, T(F)) = n^{1/2}(T_n - T(F))$$

and the Studentized root

$$S_n(x_n, T(F)) = n^{1/2}(T_n - T(F))/s_n,$$

where T_n is some estimate of $T(F)$ and $n^{1/2}s_n$ is some (consistent) estimate of the (asymptotic) standard deviation of T_n. In large samples, the distribution of R_n under F, say $J_n(\cdot, F)$, is typically normal with mean 0 and variance $\sigma^2(F)$, which depends on F. On the other hand, the distribution of S_n under F, say $K_n(\cdot, F)$, is asymptotically standard normal, and so is at least asymptotically pivotal. This suggests the (finite sample) distribution of S_n is less dependent on F than that of R_n. Indeed (see Beran (1982) or Hall (1988), for example), in regular cases, bootstrapping a Studentized root results in improved coverage accuracy. In particular,

$$J_n(x, F) - J_n(x, \hat{F}_n) = O_F(n^{-1/2})$$

and

$$K_n(x, F) - K_n(x, \hat{F}_n) = O_F(n^{-1}),$$

and the same orders of coverage error apply to the corresponding one-sided bootstrap confidence intervals. It should be noted, however, that coverage error of $O_F(n^{-1/2})$ based on the root R_n is not really a fault of the bootstrap as Beran (1982) has shown that there does not exist an estimate of $J_n(\cdot, F)$ that can improve upon this rate of convergence. Rather, the problem lies in a perhaps poor choice of root.

Of course, the drawback of bootstrapping a Studentized root is that one must have a consistent estimate of $\sigma(F)$. One possibility is to use a bootstrap estimate of variance, but the resulting confidence procedure would involve nested bootstrap calculations, and so may be computationally undesirable. Moreover, as remarked in Section 3.3, bootstrap estimates of variance typically are, but need not be, consistent.

4.2. *Prepivoting*

As in Section 3, let $J_n(F)$ be the distribution function of a root $R_n(x_n, T(F))$. The resulting bootstrap confidence region $B_n(\alpha, x_n)$ is given by (3.1). Letting

$$R_{n1}(x_n, T(F)) = J_n\{R_n(x_n, T(F)), \hat{G}_n\},$$

we may write $B_n(\alpha, x_n)$ as

$$B_n(\alpha, x_n) = \{t \in \mathbf{T}: R_{n1}(x_n, t) < 1 - \alpha\}. \tag{4.1}$$

Let $J_{n1}(\cdot, F)$ be the distribution function of $R_{n1}(x_n, T(F))$ under F. In effect, confidence set (4.1) approximates the distribution $J_{n1}(\cdot, F)$ by the uniform distribution. Instead, analogous to (3.1), one may estimate the distribution $J_{n1}(\cdot, F)$ by $J_{n1}(\cdot, \hat{G}_n)$ to obtain a new confidence set:

$$B_{n1}(\alpha, x_n) = \{t \in \mathbf{T}: R_{n1}(x_n, t) < J_{n1}^{-1}(1 - \alpha, \hat{G}_n)\}. \tag{4.2}$$

This construction is Beran's (1987) method of prepivoting. Instead of bootstrapping the root $R_n(x_n, t)$, a new root, $R_{n1}(x_n, t)$, is formed (by using the bootstrapping distribution of $R_n(x_n, t)$!) and the method of Section 3 is applied to form a confidence set based on this new root. Beran argues that R_{n1} is more nearly pivotal than R_n; that is, the distribution $J_{n1}(\cdot, F)$ is less dependent on F than is $J_n(\cdot, F)$. This seems plausible because one typically begins by choosing an asymptotically normal root R_n whose asymptotic variance depends on F. The asymptotic distribution of R_{n1} is uniform and does not depend on F. Moreover, beginning with an asymptotically normal root

$$R_n(x_n, T(F)) = n^{1/2}\{T_n - T(F)\},$$

Beran argues that the prepivoting operation is asymptotically equivalent to Studentizing. The beauty of prepivoting is the generality of the approach.

The motivation for prepivoting may be derived from the point of view of Loh's (1987) calibrated confidence sets. In fact, we show these methods to be the same. In particular, consider the following refinement of the bootstrap confidence set (4.1). The exact coverage probability that $T(F)$ is contained in $B_n(\alpha, x_n)$ under F is given by $J_{n1}(1 - \alpha, F)$, and is unknown because F is unknown. We can estimate this coverage probability by $J_{n1}(1 - \alpha, \hat{G}_n)$. Now, choose α_1 so that this estimated coverage is $1 - \alpha$; that is, solve

$$J_{n1}(1 - \alpha_1, \hat{G}_n) = 1 - \alpha. \tag{4.3}$$

Compute a new confidence set as before, except use α_1 instead of α to obtain

$$\{t \in \mathbf{T}: R_{n1}(x_n, t) < 1 - \alpha_1\}. \tag{4.4}$$

By (4.3), $1 - \alpha_1 = J_{n1}^{-1}(1 - \alpha, \hat{G}_n)$, so that (4.2) and (4.4) agree.

The method of prepivoting may be iterated as follows. Given a root $R_{nj}(x_n, T(F))$, let $J_{nj}(\cdot, F)$ be its distribution function under F. Form a new root $R_{n,j+1}$ by

$$R_{n,j+1}(x_n, T(F)) = J_{nj}\{R_{nj}(x_n, T(F)), \hat{G}_n\}$$

and let

$$B_{nj}(\alpha, x_n) = \{t \in \mathbf{T}: J_{nj}(R_{n,j+1}(x_n, t), \hat{G}_n) < 1 - \alpha\}.$$

Beran argues that, when Edgeworth expansions for $J_n(\cdot, F)$ exist, the error level of $B_{nj}(\alpha, x_n)$ decreases as j increases.

4.3. *Hall's Iterative Additive Correction*

Let $S_{n1}(x_n, t) = R_n(x_n, t) - J_n^{-1}(1-\alpha, \hat{G}_n)$. Rewrite confidence set (3.1) in the form

$$B_n(\alpha, x_n) = \{t \in \mathbf{T}: S_{n1}(x_n, t) < 0\}. \tag{4.5}$$

Let $L_{n1}(\cdot, F)$ be the distribution function of $S_{n1}(x_n, T(F))$ under F. The confidence set (4.5) pretends that $L_{n1}^{-1}(1-\alpha, F) = 0$. Instead one may estimate $L_{n1}^{-1}(1-\alpha, F)$ by $L_{n1}^{-1}(1-\alpha, \hat{G}_n)$ and form a new confidence set

$$A_{n1}(\alpha, x_n) = \{t \in \mathbf{T}: S_{n1}(x_n, t) < L_{n1}^{-1}(1-\alpha, \hat{G}_n)\}.$$

This is Hall's (1986a) additive correction. Like prepivoting, this procedure may be iterated as follows. Let

$$S_{n,j+1}(x_n, t) = S_{nj}(x_n, t) - L_{nj}^{-1}(1-\alpha, \hat{G}_n).$$

Let $L_{n,j+1}(\cdot, F)$ be the distribution function of $S_{n,j+1}(x_n, t)$ under F and form

$$A_{n,j+1}(\alpha, x_n) = \{t \in \mathbf{T}: S_{n,j+1}(x_n, t) < L_{n,j+1}^{-1}(1-\alpha, \hat{G}_n)\}.$$

Like prepivoting, each round of iteration reduces the level error in regular cases.

4.4. *Conclusions*

Bootstrap iterative methods offer the potential of high accuracy, at least in regular problems. Of course, for fixed sample size, iteration may not improve accuracy and can make the situation worse. It is important to note that the currently available supporting theory is asymptotic in the sample size. Prepivoted confidence sets, unlike Hall's method, offer the advantage of being invariant under monotone transformations of the root and reparameterizations of $T(F)$. This can, in part, be explained by the fact that Beran's roots R_{nj} obtained by prepivoting do not depend on the initial fixed choice of α whereas Hall's S_{nj} do depend on α. The current drawback with iterative methods is the inherent computational complexity. In some cases, however, prepivoting has analytical approximations; see Beran (1987). Otherwise, each round of iteration involves a nested bootstrap calculation, so that the number of computations is typically exponential in the number of iterations. In addition, even disregarding the computational problems, how many iterations are possible? Current asymptotics assume the number of iterations j is fixed while $n \to \infty$. Clearly, j cannot be comparable with n, but what is feasible? Two important open questions are the following. Can clever computational algorithms be designed to efficiently reduce the number of calculations as to make these methods viable? Second, do these iterative procedures converge and, if so, can one directly approximate the limiting algorithm?

REFERENCES

Abramovitch, L. and Singh, K. (1985) Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.*, **13**, 116–132.

Athreya, K. (1987) Bootstrap of the mean in the infinite variance case. *Ann. Statist.*, **15**, 724–731.

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

Beran, R. (1982) Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.*, **10**, 212–225.

———(1984a) Bootstrap methods in statistics. *Jber. Dtsch. Math-Ver.*, **86**, 14–30.

———(1984b) Jackknife approximations to bootstrap estimates. *Ann. Statist.*, **12**, 101–118.

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

Beran, R. and Millar, P. (1985) Asymptotic theory of confidence sets. In *Proc. Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer*, II (eds L. Le Cam and R. Olshen), pp. 865–886. Belmont: Wadsworth.

———(1986) Confidence sets for a multivariate distribution. *Ann. Statist.*, **14**, 431–443.

Beran, R. and Srivastava, M. (1985) Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.*, **13**, 95–115.

Bickel, P. J. (1987) Discussion on Efron. *J. Amer. Statist. Ass.*, **82**, 191.

Bickel, P. J. and Freedman, D. A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Cox, D. R. (1980) Local ancillarity. *Biometrika*, **67**, 273–278.

Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *J. R. Statist. Soc.* B, **49**, 1–39.

DiCiccio, T. and Romano, J. (1987) Accurate bootstrap confidence limits in parametric models. *Technical Report No. 281.* Department of Statistics, Stanford University.

DiCiccio, T. and Tibshirani, R. (1987) Bootstrap confidence intervals and bootstrap approximations. *J. Amer. Statist. Ass.*, **82**, 163–170.

Ducharme, G., Jhun, M., Romano, J. and Truong, K. (1985) Bootstrap confidence cones for directional data. *Biometrika*, **72**, 637–645.

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

———(1981) Non-parametric standard errors and confidence intervals. *Can. J. Statist.*, **9**, 139–172.

———(1982) The jackknife, the bootstrap, and other resampling plans. In *Regional Conference Series in Applied Mathematics*, No. 38. Philadelphia: SIAM.

———(1985) Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, **72**, 45–58.

———(1987) Better bootstrap confidence intervals. *J. Amer. Statist. Ass.*, **82**, 171–185.

Efron, B. and Tibshirani, R. (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, **1**, 54–77.

Freedman, D. A. (1981) Bootstrapping regression models. *Ann. Statist.*, **9**, 1218–1228.

Freedman, D. (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.*, **12**, 827–842.

Ghosh, M., Parr, W., Singh, K. and Babu, G. (1984) A note on bootstrapping the sample median. *Ann. Statist.*, **12**, 1130–1135.

Hall, P. (1983) Inverting an Edgeworth expansion. *Ann. Statist.*, **11**, 569–576.

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

———(1986b) On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.*, **4**, 1453–1462.

———(1987) On the bootstrap and likelihood-based confidence regions. *Biometrika*, **74**, 481–493.

———(1988) Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.*, to be published.

Haycock, K. (1986) Bootstrapping prediction error estimates in dynamic linear models. *PhD Dissertation.* Department of Statistics, University of California at Berkeley.

Hinkley, D. (1988) Bootstrap methods. *J. R. Statist. Soc.* B, **50**, 321–337.

Johns, V. (1988) Importance sampling for bootstrap confidence intervals. *J. Amer. Statist. Ass.*, to be published.

Liu, R., Singh, K. and Lo, S. (1986) On a representation related to the bootstrap. Preprint.

Loh, W. (1987) Calibrating confidence coefficients. *J. Amer. Statist. Ass.*, **82**, 155–162.

McCullagh, P. (1984) Local sufficiency. *Biometrika*, **71**, 233–244.

Millar, P. W. (1983) The minimax principle in asymptotic statistical theory. *Springer Lecture Notes*, **976**, 76–265.

Romano, J. (1986) A note on uniform convergence of the empirical measure with applications to simulation techniques. *Technical Report No. 263.* Department of Statistics, Stanford University.

———(1988) Bootstrapping the mode. *Ann. Inst. Statist. Math.*, to be published.

Schenker, N. (1985) Qualms about bootstrap confidence intervals. *J. Amer. Statist. Ass.*, **80**, 360–361.

———(1987) Discussion on Efron. *J. Amer. Statist. Ass.*, **82**, 192–194.

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

Sugahara, C. N. (1987) Bootstrapping Markov chains. Forthcoming *PhD Dissertation.* Department of Statistics, University of California at Berkeley.

Withers, C. (1983) Expansion for the distribution and quantiles of a regular functional of the empirical distribution with applications to non-parametric confidence intervals. *Ann. Statist.*, **11**, 577–587.