Kernel Smoothing to Improve Bootstrap Confidence Intervals

by

Alan M. Polansky
Division of Statistics
Northern Illinois University
DeKalb, Illinois, 60115-2854

William R. Schucany
Department of Statistical Science
Southern Methodist University
Dallas, Texas 75275

Technical Report No. SMU/DS/TR-283

Kernel Smoothing to Improve Bootstrap Confidence Intervals

Alan M. Polansky and William R. Schucany*

SUMMARY

Some studies of the bootstrap have assessed the effect of smoothing the estimated distribution that is resampled, a process usually known as the smoothed bootstrap. Generally, the smoothed distribution for resampling is a kernel estimate and is often rescaled to retain certain characteristics of the empirical distribution. Typically the impact of such smoothing has been measured in terms of the mean squared error of bootstrap point estimates. The reports of these investigations have not been encouraging about the efficacy of smoothing. In this paper the effect of resampling a kernel smoothed distribution is evaluated through expansions for the coverage of bootstrap percentile confidence intervals. It is shown that, under the smooth function model, proper bandwidth selection can accomplish a first-order correction for the one-sided percentile method. With the objective of reducing coverage error, the appropriate bandwidth converges at a rate of $n^{-1/4}$ rather than the familiar $n^{-1/5}$ for kernel density estimation. Additionally, the bandwidth depends on moments of the smooth function model and not on derivatives of the underlying density of the data. The relationship of the method to both the accelerated bias correction and bootstrap-t methods provides some insight into the connections among three quite distinct approximate confidence intervals.

KEY WORDS: Bandwidth; Edgeworth expansion; Smooth function model; Accelerated bias correction; Bootstrap-t

* Alan M. Polansky is Assistant Professor of Statistics at Northern Illinois University, DeKalb, Illinois, 60115-2854. William R. Schucany is Professor of Statistical Science at Southern Methodist University, Dallas, Texas, 75275-0332. This research was completed while the first author was a graduate student at Southern Methodist University. The authors wish to thank Rudy Guerra, Suojin Wang, and Alastair Young for helpful comments and suggestions.

1. INTRODUCTION

Let $\mathcal{X} = \{X_1, ..., X_n\}$ be a set of independent and identically distributed random variables from a distribution F with density f and scalar parameter $\theta = \theta(F)$. The problem of primary interest here is the construction of confidence intervals for θ based on $\widehat{\theta} = \theta(\widehat{F}) = \widehat{\theta}(\mathcal{X})$, where $\widehat{F}$ is some estimate of F based on $\mathcal{X}$. Forming such intervals requires knowledge of the sampling distribution of $\widehat{\theta}$, denoted here by $H_n(\widehat{\theta}, F)$.

The bootstrap, a method developed by Efron (1979), estimates $H_n(\widehat{\theta}, F)$ by the sampling distribution of $\widehat{\theta}$ when the X_i are independent and identically distributed random variables from $\widehat{F}$. Thus, in the bootstrap methodology,

$$\widehat{H}_n = H_n(\widehat{\theta},\, \widehat{F})\,. \tag{1}$$

In general (1) will not have an explicit closed form. In such cases, $\widehat{H}_n$ can be approximated by simulation.

Some of the early bootstrap theory dealt with the selection of $\widehat{F}$. With the most common choice for $\widehat{F}$, the empirical distribution, $\widehat{F}_n$, the method is usually referred to as the standard or nonparametric bootstrap. When more is known about F, other alternatives have been suggested. Efron (1979) suggests that when a parametric form F_θ is known, one should use $\widehat{F} = F_{\widehat{\theta}}$, known as the parametric bootstrap. When symmetry can be assumed, Efron (1979) also explores the use of a symmetric version of $\widehat{F}$.

When no parametric assumptions can be made one might think it advantageous to choose $\widehat{F}$ to be smooth or continuous. A natural choice in the nonparametric setting is to estimate $\widehat{F}$ based on a kernel estimate of f,

$$\widehat{f}_{n,h}(x) = (nh)^{-1}\sum_{i=1}^{n} K[(x - X_i)/h],\; x \in \mathbb{R},\; h > 0,$$

where the kernel function, K, is commonly an absolutely continuous symmetric density with mean 0 and variance 1. The corresponding distribution is denoted by $\widehat{F}_{n,h}$. For the remainder of this paper K is the standard normal density because of its special set of cumulants. The parameter h is called the bandwidth or smoothing parameter and controls the smoothness of the resulting estimate. A complete review of kernel smoothing is given by Wand and Jones (1995). In the bootstrap setting, the pertinent problem is to choose h so that the resulting bootstrap estimates are optimal in some sense.

Evidence of substantial improvement in mean squared error due to resampling from $\widehat{F}_{n,h}$ has not been overwhelming. A review of the relevant ideas is given in De Angelis and Young (1992). Optimal choices for h based on mean squared error (MSE) are discussed by Silverman and Young (1987), Hall, DiCiccio, and Romano (1989) and Wang (1989). An example of smoothing the bootstrap with a criterion other than mean squared error has been considered by Banks (1988).

In this paper we use confidence interval coverage rather than MSE as the criterion for bandwidth selection. As far as we know, this is a new approach. Specifically, we derive a bandwidth that yields a first-order reduction of the coverage error of the standard percentile method confidence limit. The percentile method based on the smoothed bootstrap with this bandwidth becomes asymptotically as accurate as either the bootstrap-t or the accelerated bias correction (BC_a) methods.

Section 2 presents the model under which the results are obtained. Section 3 derives the bandwidth that produces the first-order correction of the percentile method. Section 4 shows some connections of this smoothed bootstrap method with the bootstrap-t and BC_a methods. Section 5 reports some small-sample simulations. Implications of the results, as well as some questions of the practicality of data-based bandwidths, are discussed in Section 6. The proofs of the results are relegated to the Appendix.

2. A MODEL FOR ERROR EXPANSIONS

The results here are based on the Edgeworth expansion theory used extensively by Hall (1988). To use this theory the parameter of interest, θ, its estimate, and the underlying distribution of the data must follow what is called the smooth function model. See Hall (1992, Section 2.4) for a complete discussion of this model. The model requires the data to be a sequence of independent and identically distributed random vectors $Y_1, Y_2, \ldots, Y_n$ of dimension d, from a sufficiently smooth distribution F_Y with mean μ. It is assumed that $\theta = g(\mu)$ for a smooth function g and $\widehat{\theta} = g(\overline{Y})$, where $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$. It is further assumed that for a smooth function v,

$$\sigma^2 = \mathrm{Var}[n^{1/2}\widehat{\theta}] = v^2(\mu).$$

This paper addresses the univariate case so that each random vector Y_i is assumed to be a function of a univariate random variable X_i, from a distribution F. Specifically, Y_i is assumed to have the form

$$Y_i = Y(X_i) = (\psi_1(X_i), \ldots, \psi_d(X_i))',$$

for $i = 1, \ldots, n$. The functions ψ_j are assumed to be polynomials of degree δ_j, for $j = 1, \ldots, d$. See Hall (1992, page 66) for a discussion of the validity of this restriction of the smooth function model. It is the sample $X_1, \ldots, X_n$ to which the kernel smoothing will be applied.

Suppose $X^* \sim \widehat{F}_n|\mathcal{X}$ and $\widetilde{X}^* \sim \widehat{F}_{n,h}|\mathcal{X}$. Let E^* be the appropriate bootstrap expectation conditional on $\mathcal{X}$ and define $\widehat{\mu} = E^*[Y(X^*)]$, $\widetilde{\mu} = E^*[Y(\widetilde{X}^*)]$, $\widehat{\sigma}^2 = v^2(\widehat{\mu})$ and $\widetilde{\sigma}^2 = v^2(\widetilde{\mu})$. Supposing that $h = O(n^{-1/4})$, it can be shown that the standard errors are related by

$$\widetilde{\sigma} = \widehat{\sigma} + \lambda_n h^2 + O_p(n^{-1}), \tag{2}$$

where λ_n is a sequence of constants which may depend on the moments of $\widehat{F}_n$. If λ_n depends continuously on the first k moments of X and $E[|X|^k] < \infty$, then λ_n will be consistent for its population analog which is denoted by Λ. We will further assume here that λ_n and $\widehat{\sigma}$ converge at a sufficiently fast rate, namely that $\lambda_n = \Lambda + O_p(n^{-1/2})$ and $\widehat{\sigma} = \sigma + O_p(n^{-1/2})$. Note that Equation (2) shows how λ_n relates the estimated standard error of $\widehat{\theta} = g(\widehat{\mu})$ to the estimated standard error of $\widetilde{\theta} = g(\widetilde{\mu})$. Under these additional assumptions, the smooth function model still holds for many univariate statistics such as means and variances.

Let $\widetilde{\theta} = \theta(\widehat{F}_{n,h})$ and $\widehat{\theta} = \theta(\widehat{F}_n)$. For many functionals θ there may be a considerable difference between $\widehat{\theta}$ and $\widetilde{\theta}$. We define a recentering factor, $\xi_n = \widehat{\theta} - \widetilde{\theta}$. The methodology here will require this factor to be added to the endpoints of bootstrap confidence intervals. Note that in some linear cases, e.g. when $\theta = E[X]$, $\xi_n = 0$ and no recentering is required. This restriction keeps any added bias introduced by $\widetilde{\theta}$ from entering the smoothed bootstrap confidence intervals. However, the key to the results in the next section is that ordinary kernel smoothing adds h^2 to Var$[X]$.

3. ASYMPTOTICALLY OPTIMAL BANDWIDTHS

Hall (1988) makes a detailed study of asymptotic expansions of coverage probabilities for various methods of constructing one-sided and two-sided bootstrap confidence intervals. In this paper we restrict attention to one-sided intervals, where the results illustrate the issues without additional complications. Furthermore, two-sided equal-tail intervals are already second-order correct. Our two-sided results also increase accuracy; see Polansky (1995). Let $\pi_{PM}(\alpha)$, $\pi_{BT}(\alpha)$, $\pi_{BC}(\alpha)$, denote the actual coverage of the percentile, bootstrap-t, and BC_a methods respectively. Hall (1988) shows that

$$\pi_{PM}(\alpha) = \alpha - n^{-1/2}\omega\phi(z_\alpha)(p_1(z_\alpha) + q_1(z_\alpha)) + O(n^{-1}), \tag{3}$$

$$\pi_{BT}(\alpha) = \alpha + O(n^{-1}),$$

and

$$\pi_{BC}(\alpha) = \alpha + O(n^{-1}),$$

as $n \to \infty$. The constant ω is 1 if the confidence limit is an upper limit, and is -1 otherwise. The function ϕ is the standard normal density and z_α its α^{th} percentile, so that $\Phi(z_\alpha) = \alpha$. The functions p_1 and q_1 are the polynomials from the first-order term in the Edgeworth expansions for $(\widehat{\theta} - \theta)/\sigma$ and $(\widehat{\theta} - \theta)/\widehat{\sigma}$, respectively. The coefficients of these polynomials depend on the moments of F. See Hall (1992) for more information about these polynomials. It is clear from the above expansions that the percentile method is asymptotically less accurate than the bootstrap-t and BC_a methods. In the terminology of Efron (1987), the percentile method is said to be first-order correct while the bootstrap-t and BC_a methods are second-order correct.

The goal of this paper is to show that a bandwidth exists so that a smooth bootstrap performs a first-order correction on the percentile method, making it second-order correct. Such an effect is the exception to the rule in Hall, DiCiccio and Romano (1989) that smoothing the bootstrap usually only has a second-order effect on performance. However, a more elaborate smoothing methodology by Wang (1995) has first-order improvements in MSE.

The following theorem establishes an asymptotic expansion for the coverage of the percentile method based upon the smoothed $\widehat{F}_{n,h}$ with $h = O(n^{-1/4})$. The confidence limit is denoted by $\widetilde{\theta}_{PM}(h, \alpha)$ and the associated coverage by $\widetilde{\pi}_{PM}(h, \alpha)$.

The proof, which relies heavily upon the techniques in Hall (1992), is outlined in the Appendix.

Our motive for the normal kernel, K, is that it affects only the second cumulant of the effective convolution with $\widehat{F}$. In other words, by this device the asymptotics for $\widehat{F}_{n,h}$ differ from those for $\widehat{F}_n$ by replacing Var[X^*] with Var[$\widetilde{X}^*$] = Var[X^*] + h^2. The result is that we may capitalize upon this very selective control over one ingredient in the error expansion to produce a strategic cancellation of the leading term of $\pi_{PM}(\alpha)$.

Theorem 1. Suppose $X_1,\ldots,X_n$ follow the restricted smooth function model with $\Lambda \neq 0$ and $E\|Y\|^l < \infty$ for an unspecified number $l > 3$ and that g has at least 5 continuous and bounded derivatives. Then, if $h = O(n^{-1/4})$, the coverage of a one-sided $100\alpha\%$ smoothed percentile method confidence limit has the asymptotic expansion

$$\begin{aligned} \widetilde{\pi}_{PM}(h,\alpha) &= \alpha - n^{-1/2}\omega\phi(z_\alpha)(p_1(z_\alpha) + q_1(z_\alpha)) \\ &\quad + h^2 z_\alpha \phi(z_\alpha)\Lambda\sigma^{-1} + O(n^{-1}). \end{aligned} \tag{4}$$

With expansion (4) it is simple to derive an expression for h that will eliminate the $O(n^{-1/2})$ term from $\widetilde{\pi}_{PM}(h,\alpha)$. The desired bandwidth is

$$h_c = n^{-1/4}[\omega\sigma(p_1(z_\alpha) + q_1(z_\alpha))/(\Lambda z_\alpha)]^{1/2}, \tag{5}$$

where it is assumed that $\omega\Lambda^{-1}(p_1(z_\alpha) + q_1(z_\alpha)) > 0$. We return to this condition later. Then it is immediate that for this choice of h, the coverage of the smoothed percentile method is

$$\widetilde{\pi}_{PM}(h_c,\alpha) = \alpha + O(n^{-1}).$$

So, for example, a smoothed bootstrap upper confidence limit would be $\widetilde{\theta}_{PM}(h_c,\alpha)$ + ξ_n, where ξ_n is the recentering factor. Thus, smoothing with bandwidth h_c performs a first order correction of the percentile method making it second-order correct. Note that

$h_c = O(n^{-1/4})$. This is consistent with the observation of De Angelis and Young (1992, Section 2), that appropriate bandwidths for the smoothed bootstrap are usually of smaller order than for density estimation.

Example. Consider the fundamental application to the mean functional, $\theta = E_F[X]$. Hall (1988) shows that this setup follows the smooth function model with $d = 2$, $Y(X_i) = (X_i, X_i^2)$, $g(Y) = Y^{(1)}$ and $v^2(Y) = Y^{(2)} - {Y^{(1)}}^2$, where $Y^{(i)}$ is the i^{th} element of the vector Y. Now $\sigma^2 = \mathrm{Var}[X_i]$ so that $\widetilde{\sigma}^{\,2} = \widehat{\sigma}^2 + h^2$. By a Taylor expansion argument, when $E\|Y\| < \infty$ and $h = O(n^{-1/4})$, it follows that

$$\widetilde{\sigma} = \widehat{\sigma} + h^2/(2\widehat{\sigma}) + O_p(n^{-1}).$$

Hence $\lambda_n = 1/(2\widehat{\sigma})$ and for this model $\Lambda = 1/(2\sigma) > 0$.

Hall (1988) shows that in this case

$$p_1(z_\alpha) + q_1(z_\alpha) = \gamma(z_\alpha^2 + 2)/(6\sigma^3),$$

where $\gamma = E[X - \theta]^3$. Hence, the bandwidth (5) becomes

$$h_c = n^{-1/4}\left[\omega\gamma(z_\alpha^2 + 2)/(3\sigma z_\alpha)\right]^{1/2}, \tag{6}$$

provided $\gamma > 0$. Note that h_c depends only on the variance and skewness of F. If F is symmetric, $h_c = 0$, and no smoothing is required. This is consistent with the observations of Hall (1988) on this same example. Note that it is relatively easy to construct a plug in estimate for h_c in this case. Let $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$ and $\widehat{\gamma} = n^{-1}\sum_{i=1}^n (X_i - \overline{X})^3$, then one can estimate h_c with

$$\widehat{h}_c = n^{-1/4}[\omega\widehat{\gamma}(z_\alpha^2 + 2)/(3\widehat{\sigma} z_\alpha)]^{1/2}, \tag{7}$$

provided $\widehat{\gamma} > 0$. Some properties of $\widehat{h}_c$ are investigated in Section 5.

The condition $\omega\Lambda^{-1}(p_1(z_\alpha) + q_1(z_\alpha)) > 0$ implies that this type of correction by standard kernel smoothing is not always possible. The correction relies on an adjustment of the variance of F, achieved by kernel smoothing. For example, to remove some of the bias from the percentile method, may require the resampled $\widehat{F}$ to have a larger variance than $\widehat{F}_n$. By using a kernel estimate $\widehat{F}_{n,h}$, we increase the variance by h^2. This corresponds to the situation where $\omega\Lambda^{-1}(p_1(z_\alpha)+q_1(z_\alpha))>0$. In some situations a reduction in variance may be necessary, or $\omega\Lambda^{-1}(p_1(z_\alpha)+q_1(z_\alpha))<0$. It is not possible to achieve such a reduction by kernel smoothing alone. Such a reduction is possible, if one rescales the original data to have smaller variance. Kernel density estimation can then reinflate the variance to the desired amount. One such method is discussed by Polansky (1995). The effect of this procedure is to change the sign of Λ. This method is similar to the rescaling method of Wang (1995), although we are not presently interested in the optimal selection of the rescaling parameter. Note that if $p_1(z_\alpha)+q_1(z_\alpha)=0$, then $h_c=0$, which corresponds to no smoothing. In this case, from (3), the percentile method is already second-order correct and no adjustment is necessary.

4. CONNECTIONS WITH OTHER METHODS

The correction made by the smoothed bootstrap of Section 3 provides an adjustment that results in asymptotic coverage which is second-order accurate, the same order of accuracy as the BC_a and bootstrap-t methods. This section establishes connections between the smoothed bootstrap with bandwidth h_c and these two methods. The proofs of the two corollaries are in the Appendix.

To have a data-driven bandwidth to consider, but not to recommend, there is the obvious plug-in estimate of h_c,

$$\widehat{h}_c = n^{-1/4}[\omega\widehat{\sigma}(\widehat{p}_1(z_\alpha)+\widehat{q}_1(z_\alpha))/(\lambda_n z_\alpha)]^{1/2}, \tag{8}$$

where we assume that $\omega {\lambda_n}^{-1}(\widehat{p}_1(z_\alpha) + \widehat{q}_1(z_\alpha)) > 0$. Here, $\widehat{p}_1$ and $\widehat{q}_1$ are the sample versions of p_1 and q_1: the moments of F are replaced by the moments of $\widehat{F}_n$. Similarly, we treat λ_n as the sample version of Λ. The following corollary shows that smoothing with $\widehat{h}_c$ is asymptotically third-order equivalent to the bootstrap-t and thus to BC_a.

Corollary 1. Let $\widehat{\theta}_{BT}(\alpha), \widehat{\theta}_{BC}(\alpha)$, and $\widetilde{\theta}_{PM}(h, \alpha)$ be confidence limits of the bootstrap-t, BC_a, and the smoothed percentile method with bandwidth h, respectively. Then, under the conditions of Theorem 1,

$$\widetilde{\theta}_{PM}(\widehat{h}_c, \alpha) = \widehat{\theta}_{BT}(\alpha) + O_p(n^{-3/2})$$

$$= \widehat{\theta}_{BC}(\alpha) + O_p(n^{-3/2}).$$

Thus the smoothed percentile method with plug-in estimates for h has confidence limits that are asymptotically close to the bootstrap-t and BC_a confidence limits. Hence, for large samples one may expect similar performance from the bootstrap-t, BC_a, and an estimated smoothed percentile method. Note that for many cases the smoothed percentile method will be computationally less intensive than the bootstrap-t. Implementation of the bootstrap-t requires that an estimate of the standard error of $\widehat{\theta}$ be calculated with each bootstrap replication. In the smoothed percentile method a standard error estimate of $\widehat{\theta}$ is needed only for the calculation of $\widehat{h}_c$. Further, many authors, e.g. Efron and Tibshirani (1994), have noted numerical instability of bootstrap-t estimates for small sample sizes. It is possible that the smoothed percentile method is more stable in some cases.

The form of the bandwidth can also be related to the BC_a method. Let a and z_0 be the acceleration constant and the bias correction, respectively. See Efron (1987) for a complete description of this method.

Corollary 2. Under the conditions of Theorem 1,

$$h_c = [\omega\sigma(2z_0 + az_\alpha^2)/(\Lambda z_\alpha)]^{1/2} + O(n^{-1/2}),$$

provided $\omega\Lambda^{-1}(p_1(z_\alpha) + q_1(z_\alpha)) > 0$ and $\omega\Lambda^{-1}(2z_0 + az_\alpha^2) > 0$.

The bandwidth is clearly related to the acceleration constant and the bias correction used by Efron (1987) to achieve similar performance. The $O(n^{-1/2})$ error term is due to an approximation associated with z_0. The remaining dependence of the bandwidth is on F and $\widehat{\theta}$ through the standard error of $\widehat{\theta}$ and Λ, the relationship between the smoothed and nonsmoothed standard errors. Hence, the main part of the correction is due to the median bias of $\widehat{\theta}$ and the skewness of $H_n(\widehat{\theta}, \widehat{F})$. Corollary 2 suggests an obvious bandwidth estimation technique that relies upon routines for $\widehat{a}$ and $\widehat{z}_0$. We have experimented with this, but do not pursue that topic here.

5. A SIMULATION STUDY

To assess the finite-sample performance of the correction due to h_c a small simulation was performed. The data consisted of 1000 samples of sizes 10 and 20 from a Gamma distribution with density

$$f(x) = \begin{cases} \Gamma(\theta)^{-1}x^{\theta-1}e^{-x} & \theta > 0,\, x \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where the parameter θ is generally called the shape parameter. The standardized skewness of this distribution is $2\theta^{-1/2}$. Samples were generated using $\theta = 0.25$, 0.5, and 1.0 to study the effect of various amounts of skewness. The samples were simulated and analyzed using S-Plus. The parameter of interest is the mean θ. The skewness makes this a challenging interval estimation problem for many approximations in small samples. For each sample, upper 90% confidence limits for θ were calculated using various bootstrap, smoothed bootstrap, and traditional methods. For each method the coverage probability was estimated by the proportion of times the method covered θ in the 1000 samples. We studied lower confidence limits, as well. While there are some differences there, the overall lessons are the same.

The three standard bootstrap procedures are the percentile method (BP), the bootstrap-t (BT), and the BC_a (BC) method. The acceleration constant for the BC_a method was estimated using the positive jackknife method as in Frangos and Schucany (1990).

Two smoothed bootstrap methods using the bandwidth selection result of Theorem 1 were evaluated. Consider samples $X_1, \ldots, X_n$ following the distribution (9). From the Example in Section 3, when the mean is the parameter of interest, the bandwidth (5) becomes

$$h_c = n^{-1/4}[\gamma(z_\alpha^2 + 2)/(3\sigma z_\alpha)]^{1/2}, \tag{10}$$

where $\sigma^2 = E[X - \theta]^2$ and $\gamma = E[X - \theta]^3$. For this example, $\sigma^2 = \theta$ and $\gamma = 2\theta$, so that

$$h_c = n^{-1/4}[2\theta^{1/2}(z_\alpha^2 + 2)/(3z_\alpha)]^{1/2}. \tag{11}$$

The first smoothed bootstrap method (ST) considered uses (11), where θ is set to its actual theoretical value. This allows us to evaluate the effect of the true smoothing correction on the coverage. The second method (SP) uses the plug-in estimator (7),

$$\widehat{h}_c = n^{-1/4}[\widehat{\gamma}(z_\alpha^2 + 2)/(3\widehat{\sigma} z_\alpha)]^{1/2}.$$

Whenever $\widehat{\gamma} < 0$, we used a rescaling algorithm described in Polansky (1995) for the smoothing.

Some classical limits are also evaluated. The first is a normal theory (NT) approximation $\overline{X} + (\widehat{\sigma}/n^{1/2})t_{\alpha,n-1}$, where $t_{\alpha,n-1}$ is the α^{th} percentile of the Student-t

distribution with $n-1$ degrees of freedom. For the special case in which an exact pivot exists at $\theta = 1$, we evaluated the parametric limit (PL), namely $2(\sum_{i=1}^{n} X_i)/\chi^2_{\alpha,2n}$, where $\chi^2_{\alpha,2n}$ is the α^{th} percentile of a chi-squared distribution with $2n$ degrees of freedom.

Some results of the simulations are summarized in Tables 1 and 2 and Figure 1. Table 1 gives the estimated coverages for each method, sample size, and skewness. With 1000 repetitions the marginal standard errors of these estimates are around .01. The columns are based on independent runs. Comparing each of the seven methods with an exact version of McNemar's test by blocking on each repetition, the coverages are pairwise statistically significantly different (at the .01 level) for all but a few exceptions noted below. See Frangos and Schucany (1990) for an explanation of a similar application of this methodology from Lehmann (1975, pages 268-269).

Table 1. Estimated coverages for approximate upper 90% confidence limits. Abbreviations for the methods are given in Section 5.

Method	$n = 10$			$n = 20$		
	$\theta = 0.25$	$\theta = 0.50$	$\theta = 1.0$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 1.00$
NT	.718	.761	.795	.769	.814	.824
PL			.886			.891
BP	.702	.755	.763	.766	.810	.817
BT	.852	.836	.856	.873	.877	.866
BC	.755	.779	.802	.829	.851	.844
ST	.991	.918	.889	.928	.907	.888
SP	.737	.768	.790	.800	.835	.835

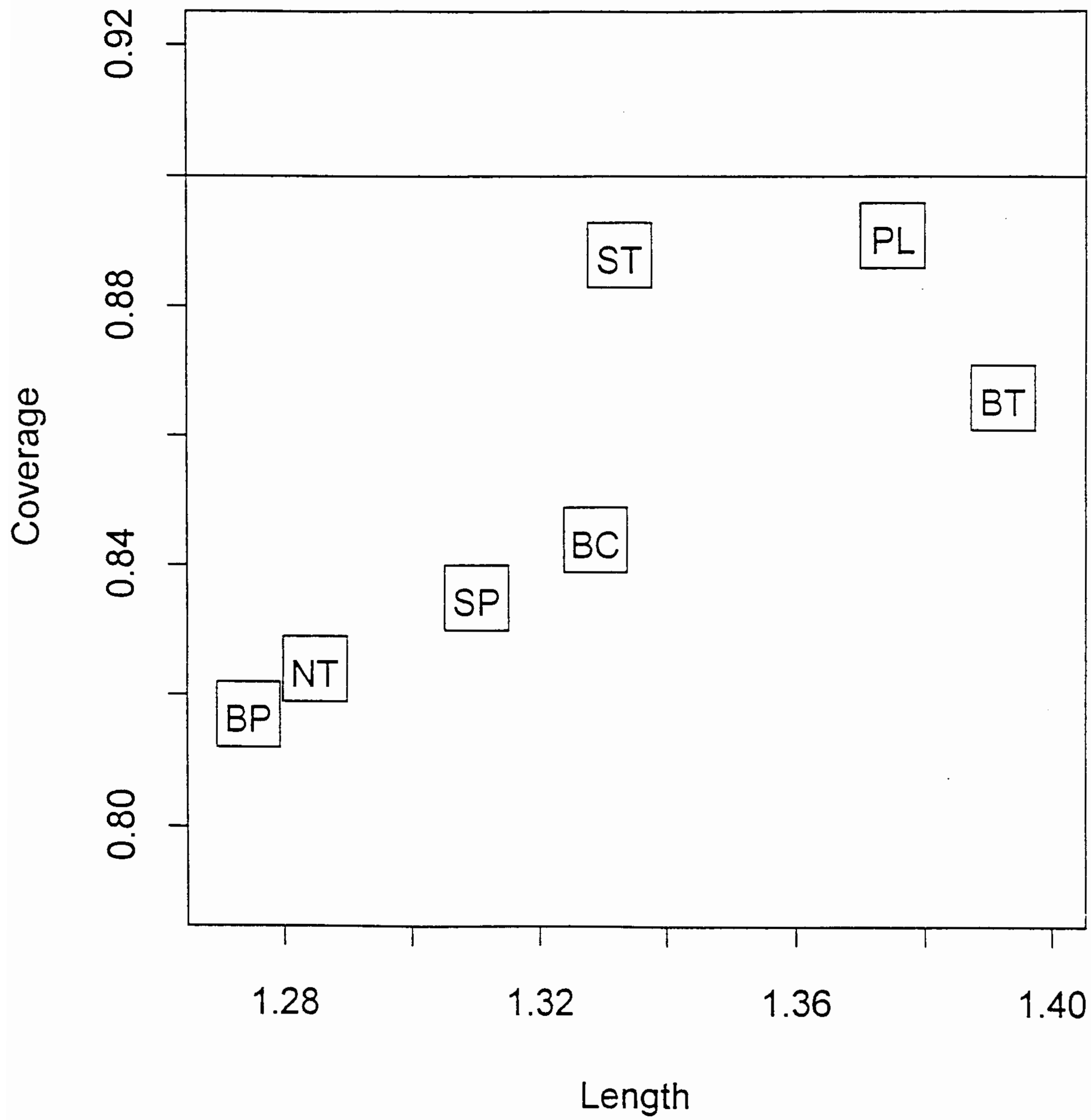
The NT and BP methods undercover consistently for all sample sizes and choices of θ. When $n = 20$, there was no significant difference in the two methods. The coverages improve for distributions with less skewness (larger θ) and for larger values of n. Both the BC and BT methods reduce this coverage error, the BT method being significantly better. The ST method provides a quite accurate correction, and was most successful in reducing the coverage error. The extreme case where $\theta = .25$ and $n = 10$ did cause the ST method to overcorrect. These impressive gains are lost for the plug-in (SP) method, which performs only slightly worse than the BC method, but is still significantly better than the BP method. In less skewed cases there was no significant difference between the BC and SP methods. A study of the average estimated bandwidths in Table 2 show that $\widehat{h}_c$ suffers from a large downward bias. It may be this bias that destroys much of the correction. If this method is to be of any practical value, a better bandwidth estimation technique will certainly be required. This may be difficult.

Table 2. Average estimated bandwidths for the SP confidence limits (estimated standard error).

Method	$n = 10$			$n = 20$		
	$\theta = 0.25$	$\theta = 0.50$	$\theta = 1.0$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 1.00$
true h_c	.55	.65	.77	.46	.55	.65
plug-in $\widehat{h}_c$	.28 (.007)	.36 (.008)	.43 (.010)	.30 (.006)	.38 (.006)	.50 (.007)

Figure 1 is a plot of the estimated coverage versus the average confidence limit for each method when $\theta = 1$ and $n = 20$. The boxes around each point have heights and widths approximately equal to one standard error for both the estimated coverage and the average upper confidence limit, respectively. These boxes are not meant to represent confidence regions, since the necessary independence assumptions do not hold. We see that the PL and ST methods performed equally well in terms of coverage; neither is significantly below 90%. But the added knowledge of the true moments used in the ST method make it appear more efficient in the sense of interval length. Of the estimated methods, the smaller coverage error of the BT method is at the price of longer intervals, on average. The remaining methods seem to follow a rule that a small increase in coverage is accompanied by a small increase in the average "length".

Figure 1. Estimated Coverage vs. Average Upper Confidence Limit for the mean of a Gamma Distribution with $n = 20$ and $\theta = 1$. The abbreviations for the methods are the same as in Table 1.



6. DISCUSSION

We have seen that the smoothed bootstrap can accomplish corrections to the percentile method similar to that of the BC_a method. Why does the change in performance criterion from MSE to confidence interval coverage change the effect of smoothing so dramatically? The smoothed bootstrap has been shown to have only second-order effects in estimating linear functionals of the underlying distribution; see Silverman and Young (1987). Confidence interval coverage, however, can be linked to the bias in the confidence limits, which heavily depend on the estimation of quantiles of the underlying distribution. It has been shown by Hall, DiCiccio and Romano (1989), that when estimates rely heavily on local properties of the underlying distribution, such as the bias of quantile estimation, dramatic effects can been realized in smoothing the bootstrap.

Hall (1992, Section 4.1) also attributes these second-order effects to the estimation of moments in Edgeworth expansions. Since, in the nonparametric case, these moments cannot be estimated better than at a rate of $O_p(n^{-1/2})$, smoothing will not have a first-order effect on the bootstrap estimates. The arguments in the Appendix show that correction of (5) is a result of an adjustment of these moments. In this case we are not inherently interested in good estimation of these moments, or even of $H_n(\widehat{\theta}, F)$. Instead, we knowingly introduce bias into these estimates so that certain corrections are realized.

The correction proposed here relies on an adjustment of the moments of the resampling distribution and not its continuity. Hence the improvement in the performance of the smoothed bootstrap in this study is not due to the continuity of the resampling distribution. Note that a purely asymptotic analysis of bootstrap performance may never account for this induced continuity. A study of the number of atoms in the bootstrap distribution, such as given in Fisher and Hall (1991), shows that the bootstrap estimate becomes "approximately" continuous at a very fast rate. Thus the best place to look for the effect of continuity on the resampling distribution of the bootstrap is where the sample size is very small, when asymptotics cannot be guaranteed. Additionally, such an analysis of the added continuity of smoothing can never use the existing Edgeworth expansion theory. The empirical resampling distribution and a continuous resampling distribution with the same moments will have the same asymptotic expansion.

This paper presents results for smoothing univariate data with a scalar parameter of interest. The case of multivariate data and a scalar parameter should also follow from a slight extension of these results. The bandwidth for two-sided intervals and an example with $\theta = \mathrm{Var}[X]$ is in Polansky (1995).

The primary theme of this paper is a re-evaluation of the smoothed bootstrap in terms of confidence interval coverage rather than MSE. With this new objective it is shown that asymptotically the smoothed bootstrap can have a first-order effect on one-sided confidence interval coverage. This correction is closely related to the correction of the BC_a method. This method, bootstrap-t, and BC_a methods agree to third order, when one estimates the bandwidth with plug-in estimates for population moments. A small simulation study demonstrates that the one-sided smoothed bootstrap correction has some potential in reducing the finite-sample coverage error of the percentile method, as well. The difficulty of developing a data-based bandwidth selector, that is ready for wide practical use, should not be underestimated.

APPENDIX: PROOFS

Proof of Theorem 1

Under the stated conditions, the upper confidence limit of the smoothed percentile method, with $h = O(n^{-1/4})$, has the asymptotic expansion

$$\widetilde{\theta}_{PM}(h,\alpha) = \widetilde{\theta} + \xi + n^{-1/2}\widetilde{\sigma}\left[z_\alpha - n^{-1/2}\widetilde{p}_1(z_\alpha) + n^{-1}\widetilde{p}_{21}(z_\alpha)\right] + O_p(n^{-3/2}),$$

where $\widetilde{p}_1$ and $\widetilde{p}_{21}$ are versions of p_1 and p_{21} whose coefficients depend on $\widehat{F}_{n,h}$. The polynomials p_{21} is specified in Equation (1.5) of Hall (1988). The details of the proofs are in Polansky (1995). For $h = O(n^{-1/4})$,

$$\widetilde{p}_1(z_\alpha) = \widehat{p}_1(z_\alpha) + O_p(n^{-1/2}),$$

and

$$\widetilde{p}_{21}(z_\alpha) = \widehat{p}_{21}(z_\alpha) + O_p(n^{-1/2}).$$

Hence, from (2),

$$\widetilde{\theta}_{PM}(h,\alpha) = \widehat{\theta}_{PM}(\alpha) + n^{-1/2}\lambda_n h^2 z_\alpha + O_p(n^{-3/2}). \tag{12}$$

First, consider the case of a one-sided upper smoothed-percentile method interval $(-\infty, \widetilde{\theta}_{PM}(h,\alpha)]$. The coverage of this interval, proceeding as in Hall (1988), is given by

$$\begin{aligned} \widetilde{\pi}_{PM}(h,\alpha) &= P\Big[\theta \le \widetilde{\theta}_{PM}(h,\alpha)\Big] \\ &= P[S_n + \Delta_n \ge -z_\alpha + \delta_n], \end{aligned}$$

where

$$\delta_n = n^{-1/2}p_1(z_\alpha) - n^{-1}p_{21}(z_\alpha) - \lambda_n h^2 z_\alpha \widehat{\sigma}^{-1} + O_p(n^{-1}),$$

$\Delta_n = n^{-\frac{1}{2}}[p_1(z_\alpha) - \widehat{p}_1(z_\alpha)]$, and $S_n = n^{1/2}\widehat{\sigma}^{-1}(\widehat{\theta} - \theta)$. Under the conditions of the theorem, $\widehat{\sigma} = \sigma + O_p(n^{-1/2})$ and $\lambda_n = \Lambda + O_p(n^{-1/2})$. Since $h = O(n^{-1/4})$, it follows that

$$\lambda_n h^2 z_\alpha \widehat{\sigma}^{-1} = \Lambda h^2 z_\alpha \sigma^{-1} + O_p(n^{-1}).$$

Applying Equation (3.28) of Hall (1992), the desired coverage may be rewritten as

$$\tilde{\pi}_{PM}(h,\alpha) = P[S_n + \Delta_n \geq -z_\alpha + \delta] + O(n^{-1}),$$

where

$$\delta = n^{-1/2}p_1(z_\alpha) - n^{-1}p_{21}(z_\alpha) - \Lambda h^2 z_\alpha \sigma^{-1}.$$

Equation (3.36) of Hall (1992) gives an expansion for

$$P[S_n + \Delta_n \leq x] = P[S_n \leq x] - n^{-1}ux\phi(x) + O(n^{-3/2}), \tag{13}$$

where u is a constant satisfying $E[S_n\Delta_n] = u + O(n^{-1})$. Using a standard Edgeworth expansion,

$$P[S_n \leq x] = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + O(n^{-3/2}). \tag{14}$$

Consider (13) evaluated at $x = -z_\alpha + \delta$. Using Taylor expansions,

$$\begin{aligned}\Phi(-z_\alpha + \delta) = {} & 1 - \alpha + n^{-1/2}p_1(z_\alpha)\phi(z_\alpha) \\ & - \Lambda h^2 z_\alpha \phi(z_\alpha)\sigma^{-1} + O(n^{-1})\end{aligned}$$

and the two functions in the second term of (14) are expanded separately to give

$$\begin{aligned}\phi(-z_\alpha + \delta) = {} & \phi(z_\alpha) - n^{-1/2}p_1(z_\alpha)\phi'(z_\alpha) \\ & + \Lambda h^2 z_\alpha \phi'(z_\alpha)\sigma^{-1} + O(n^{-1})\end{aligned} \tag{15}$$

and

$$\begin{aligned}-q_1(-z_\alpha + \delta) = {} & -q_1(z_\alpha) - n^{-1/2}B_{32}z_\alpha p_1(z_\alpha)/3 \\ & + B_{32}\Lambda h^2 z_\alpha^2 \sigma^{-1}/3 + O(n^{-1}).\end{aligned} \tag{16}$$

The constant B_{32} is a coefficient of the polynomial q_1, see Hall (1988) for more details. Since $h^2 = O(n^{-1/2})$, we can combine (15) and (16) and keep terms of order $n^{-1/2}$ to get

$$n^{-1/2}q_1(z_\alpha + \delta)\phi(z_\alpha + \delta) = n^{-1/2}q_1(z_\alpha)\phi(z_\alpha) + O(n^{-1}).$$

Substitution in (14) and because u is a constant, yields the coverage of the interval stated in the Theorem,

$$\begin{aligned}\widetilde{\pi}_{PM}(h,\alpha) &= \alpha - n^{-1/2}\omega\phi(z_\alpha)(p_1(z_\alpha)+q_1(z_\alpha))\,. \\ &\quad + h^2 z_\alpha \phi(z_\alpha)\Lambda\sigma^{-1} + O(n^{-1}).\end{aligned}$$

The arguments for the lower confidence limit proceed similarly.

Proof of Corollary 1

Consider the case of the upper confidence limit which, from (12), is

$$\widetilde{\theta}_{PM}(h,\alpha) = \widehat{\theta}_{PM}(\alpha) + n^{-1/2}\lambda_n h^2 z_\alpha + O_p(n^{-3/2}).$$

Using the estimated bandwidth, $\widehat{h}_c$, specified by (8), provided that

$$(\omega/\lambda_n)[\widehat{p}_1(z_\alpha) + \widehat{q}_1(z_\alpha)] \geq 0 \quad ,$$

it follows that the corresponding estimated upper limit of the interval is

$$\begin{aligned}\widetilde{\theta}_{PM}(\widehat{h}_c,\alpha) &= \widehat{\theta}_{PM}(\alpha) + n^{-1/2}\lambda_n \widehat{h}_c^2 z_\alpha + O_p(n^{-3/2}) \\ &= \widehat{\theta} + n^{-1/2}\widehat{\sigma}\times \\ &\quad \left[z_\alpha + n^{-1/2}\widehat{q}_1(z_\alpha) + n^{-1}\widehat{p}_{21}(z_\alpha)\right] + O_p(n^{-3/2}).\end{aligned}$$

From Hall (1988), the upper endpoint of the bootstrap-t interval has asymptotic expansion

$$\widehat{\theta}_{BT}(\alpha) = \widehat{\theta} + n^{-1/2}\widehat{\sigma}\left[z_\alpha + n^{-1/2}\widehat{q}_1(z_\alpha) + n^{-1}\widehat{q}_{21}(z_\alpha)\right] + O_p(n^{-2}).$$

Thus the kernel smoothed upper endpoint matches the bootstrap-t upper endpoint through terms of order n^{-1}. Hall also shows that $\widehat{\theta}_{BT}(\alpha) = \widehat{\theta}_{BC}(\alpha) + O_p(n^{-3/2})$, so that $\widetilde{\theta}_{PM}(\alpha) = \widehat{\theta}_{BC}(\alpha) + O_p(n^{-3/2})$. The arguments for lower confidence limits follow the same lines.

Proof of Corollary 2

Under the smooth function model, according to Hall (1992, page 133), the acceleration constant $a = n^{-1/2} A\sigma^{-3}/6$, where $A = -(A_{32} + B_{32}\sigma^{-3})$. Also, Equation (3.95) of Hall (1992) also gives an expansion for

$$z_0 = n^{-1/2} p_1(0) + O(n^{-1}).$$

Finally, Equation (3.86) of Hall (1992) gives

$$p_1(x) + q_1(x) = 2p_1(0) + A\sigma^{-3}x^2/6.$$

Combining the above equations,

$$p_1(z_\alpha) + q_1(z_\alpha) = \left(2z_0 + az_\alpha^2\right) + O(n^{-1/2}).$$

Hence,

$$\begin{aligned} h_c &= n^{-1/4}[\omega\sigma(p_1(z_\alpha) + q_1(z_\alpha))/(\Lambda z_\alpha)]^{1/2} \\ &= \left[\omega\sigma\left(2z_0 + az_\alpha^2\right)/(\Lambda z_\alpha)\right]^{1/2} + O(n^{-1/2}). \end{aligned}$$

REFERENCES

Banks, D. L. (1988) Histospline smoothing the bayesian bootstrap. *Biometrika*, **75**, 673-684.

De Angelis, D. and Young, G. A. (1992) Smoothing the bootstrap. *International Statistical Review*, **60**, 45-56.

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

Efron, B. (1987) Better bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association*, **82**, 171-200.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Fisher, N. I. and Hall, P. (1991) Bootstrap algorithms for small sample sizes. *Journal of Statistical Planning and Inference*, **27**, 157-169.

Frangos, C. C. and Schucany, W. R. (1990) Jackknife estimation of the bootstrap acceleration constant. *Computational Statistics and Data Analysis*, **9**, 271-281.

Hall, P. (1988) Theoretical comparison of bootstrap confidence intervals (with discussion). *Annals of Statistics*, **16**, 927-985.

Hall, P. (1992) *The Bootstrap and the Edgeworth Expansion.* New York: Springer-Verlag.

Hall, P., DiCiccio, T., and Romano, J. (1989) On smoothing and the bootstrap. *Annals of Statistics*, **17**, 692-704.

Lehmann, E. L. (1975) *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden Day.

Polansky, A. M. (1995) *Kernel Smoothing to Improve Bootstrap Confidence Intervals*, Ph.D. Thesis, Southern Methodist University, Department of Statistical Science.

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

Wand, M. P and Jones, M. C. (1995) *Kernel Smoothing*. New York: Chapman and Hall.

Wang, S. (1989) On the bootstrap and the smoothed bootstrap. *Communications in Statistics, Series A - Theory and Methods*, **18**, 3949-3962.

Wang, S. (1995) Optimizing the smoothed bootstrap. *Annals of the Institute of Statistical Mathematics*, **47**, 65-80.