A SURVEY OF BOOTSTRAP CONFIDENCE INTERVALS

BY

THOMAS J. DICICCIO and BRADLEY EFRON

TECHNICAL REPORT NO. 181

DECEMBER 1995

PREPARED UNDER THE AUSPICES

OF

PUBLIC HEALTH SERVICE GRANT

5 R01 CA59039-20

DIVISION OF BIOSTATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
A Survey of Bootstrap Confidence Intervals

By

Thomas J. DiCiccio and Bradley Efron

Technical Report No. 181

December 1995

Prepared Under the Auspices

Of

Public Health Service Grant

5 R01 CA59039-20

Division of Biostatistics

Stanford University

Stanford, California
A SURVEY OF BOOTSTRAP CONFIDENCE INTERVALS

Thomas J. DiCiccio and Bradley Efron

Abstract

This article surveys bootstrap methods for producing good approximate confidence intervals. The goal is to improve by an order of magnitude upon the accuracy of the standard intervals \( \hat{\theta} \pm z^{(a)} \hat{\sigma} \), in a way that allows routine application even to very complicated problems. Both theory and examples are used to show how this is done. The first four sections provide a heuristic overview of four bootstrap confidence interval procedures, \( BC_a \), bootstrap-\( t \), ABC, and calibration. Sections 8 and 9 describe the theory behind these methods, and their close connection with the likelihood-based confidence interval theory developed by Barndorff-Nielsen, Cox and Reid, and others.

1. Introduction. Confidence intervals have become familiar friends in the applied statistician's collection of data-analytic tools. They combine point estimation and hypothesis testing into a single inferential statement of great intuitive appeal. Recent advances in statistical methodology allow the construction of highly accurate approximate confidence intervals, even for very complicated probability models and elaborate data structures. This article discusses bootstrap methods for constructing such intervals in a routine, automatic way.

Two distinct approaches have guided confidence interval construction since the 1930's. A small catalogue of exact intervals has been built up for special situations, like the ratio of normal means or a single binomial parameter. However most confidence intervals are approximate, with by far the favorite approximation being the standard interval

\[ \hat{\theta} \pm z^{(a)} \hat{\sigma}. \quad (1.1) \]

Here \( \hat{\theta} \) is a point estimate of the parameter of interest \( \theta \), \( \hat{\sigma} \) is an estimate of \( \hat{\theta}'s \) standard deviation, and \( z^{(a)} \) is the 100th percentile of a normal deviate, \( z^{(95)} = 1.645 \) etc. Often, and always in this paper, \( \hat{\theta} \) and \( \hat{\sigma} \) are obtained by maximum likelihood theory.

The standard intervals, as implemented by maximum likelihood theory, are a remarkably useful tool. The method is completely automatic: the statistician inputs the data, the class of possible probability models, and the parameter of interest; a computer algorithm outputs the intervals (1.1), with no further intervention required. This is in notable contrast to the construction of an exact interval, which requires clever thought on a problem-by-problem basis when it is possible at all.

The trouble with standard intervals is that they are based on an asymptotic approximation that can be quite inaccurate in practice. The example below illustrates what every applied statistician knows, that (1.1) can considerably differ from exact intervals in those cases where exact
intervals exist. Over the years statisticians have developed tricks for improving (1.1), involving bias-corrections and parameter transformations. The bootstrap confidence intervals that we will discuss here can be thought of as automatic algorithms for carrying out these improvements without human intervention. And of course they apply as well to situations so complicated that they lie beyond the power of traditional analysis.

We begin with a simple example, where we can compute the bootstrap methods with an exact interval. Figure 1 shows the cd4 data: 20 HIV-positive subjects received an experimental anti-viral drug; cd4 counts in hundreds were recorded for each subject at baseline and after one year of treatment, giving data say \( x_i = (B_i, A_i) \) for \( i = 1, 2, \cdots, 20 \). The two measurements are highly correlated, having sample correlation coefficient \( \hat{\theta} = 0.723 \).

![Figure 1](image.jpg)

Figure 1. The cd4 data; cd4 counts in hundreds for 20 subjects, at baseline and after one year of treatment with an experimental anti-viral drug. Numerical values appear in Table 1.

What if we wish to construct a confidence interval for the true correlation \( \theta \)? We can find an exact interval for \( \theta \) if we are willing to assume bivariate normality for the \((B_i, A_i)\) pairs,

\[
\begin{align*}
\begin{pmatrix} B_i \\ A_i \end{pmatrix} & \sim \text{i.i.d. } N_2(\lambda, \Gamma) \\
& \text{for } i = 1, 2, \cdots, 20,
\end{align*}
\]

(1.2)

\( \lambda \) and \( \Gamma \) being the unknown expectation vector and covariance matrix. The exact central 90% interval is

\[
(\hat{\theta}_{\text{EXACT}[.05]}, \hat{\theta}_{\text{EXACT}[.95]}) = (0.47, 0.86).
\]

(1.3)

This notation emphasizes that a two-sided interval is intended to give correct coverage at both endpoints, two .05 noncoverage probabilities in this case, not just an overall .10 noncoverage probability.
<table>
<thead>
<tr>
<th>subject</th>
<th>baseline</th>
<th>oneyear</th>
<th>subject</th>
<th>baseline</th>
<th>oneyear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.12</td>
<td>2.47</td>
<td>11</td>
<td>4.15</td>
<td>4.74</td>
</tr>
<tr>
<td>2</td>
<td>4.35</td>
<td>4.61</td>
<td>12</td>
<td>3.56</td>
<td>3.29</td>
</tr>
<tr>
<td>3</td>
<td>3.39</td>
<td>5.26</td>
<td>13</td>
<td>3.39</td>
<td>5.55</td>
</tr>
<tr>
<td>4</td>
<td>2.51</td>
<td>3.02</td>
<td>14</td>
<td>1.88</td>
<td>2.82</td>
</tr>
<tr>
<td>5</td>
<td>4.04</td>
<td>6.36</td>
<td>15</td>
<td>2.56</td>
<td>4.23</td>
</tr>
<tr>
<td>6</td>
<td>5.10</td>
<td>5.93</td>
<td>16</td>
<td>2.96</td>
<td>3.23</td>
</tr>
<tr>
<td>7</td>
<td>3.77</td>
<td>3.93</td>
<td>17</td>
<td>2.49</td>
<td>2.56</td>
</tr>
<tr>
<td>8</td>
<td>3.35</td>
<td>4.09</td>
<td>18</td>
<td>3.03</td>
<td>4.31</td>
</tr>
<tr>
<td>9</td>
<td>4.10</td>
<td>4.88</td>
<td>19</td>
<td>2.66</td>
<td>4.37</td>
</tr>
<tr>
<td>10</td>
<td>3.35</td>
<td>3.81</td>
<td>20</td>
<td>3.00</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Table 1. The cd4 data, as plotted in Figure 1.

The left panel of Table 2 shows the exact and standard intervals for the correlation coefficient of the cd4 data, assuming the normal model (1.2). Also shown are approximate confidence intervals based on three different (but closely related) bootstrap methods: $ABC, BC_a$, and Bootstrap–t. $ABC$ and $BC_a$ match the exact interval to two decimal places, and all of the bootstrap intervals are more accurate than the standard. The examples and theory that follow are intended to show that this is no accident. The bootstrap methods make computer-based adjustments to the standard interval endpoints that are guaranteed to improve the coverage accuracy by an order of magnitude, at least asymptotically.

\[
\text{NORMAL THEORY} \quad \text{NONPARAMETRIC}
\]

<table>
<thead>
<tr>
<th></th>
<th>exact</th>
<th>$ABC$</th>
<th>$BC_a$</th>
<th>boot-t</th>
<th>stand.</th>
<th>$ABC$</th>
<th>$BC_a$</th>
<th>boot-t</th>
<th>cal.</th>
<th>stand.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.47</td>
<td>0.47</td>
<td>0.47</td>
<td>0.45</td>
<td>0.55</td>
<td>0.56</td>
<td>0.55</td>
<td>0.51</td>
<td>0.56</td>
<td>0.59</td>
</tr>
<tr>
<td>0.95</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.90</td>
<td>0.83</td>
<td>0.85</td>
<td>0.86</td>
<td>0.83</td>
<td>0.85</td>
</tr>
<tr>
<td>length</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.42</td>
<td>0.35</td>
<td>0.27</td>
<td>0.30</td>
<td>0.35</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>shape</td>
<td>0.52</td>
<td>0.52</td>
<td>0.54</td>
<td>0.52</td>
<td>1.00</td>
<td>0.67</td>
<td>0.70</td>
<td>0.63</td>
<td>0.67</td>
<td>1.00</td>
</tr>
<tr>
<td>cov 05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cov 95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.97</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Exact and approximate confidence intervals for the correlation coefficient, cd4 data; $\hat{\theta} = .723$. The bootstrap methods $ABC, BC_a$, Bootstrap–t, and calibrated $ABC$ are explained in Sections 2–7. The $ABC$ and $BC_a$ intervals are close to exact in the normal theory situation, left panel; the standard interval errs badly at both endpoints, as can be seen from the coverage probabilities in the bottom rows.

The exact interval endpoints [0.47, 0.86] are defined by the fact that they “cover” the observed value $\hat{\theta} = .723$ with the appropriate probabilities,

\[
\text{Prob}_{\hat{\theta}=.47}\{\hat{\theta} > .723\} = .05 \quad \text{and} \quad \text{Prob}_{\hat{\theta}=.86}\{\hat{\theta} > .723\} = .95.
\]

(1.5)
Table 2 shows that the corresponding probabilities for the standard endpoints [.55,.90] are .12 and .99. The standard interval is far too liberal at its lower endpoint and far too cautious at its upper endpoint. This kind of error is particularly pernicious if the confidence interval is used to test a parameter value of interest like $\theta = 0$.

Table 2 describes the various confidence intervals in terms of their length and right-left asymmetry around the point estimate $\hat{\theta}$,

$$
\text{length} = \hat{\theta} [.95] - \hat{\theta} [.05], \quad \text{shape} = \frac{\hat{\theta} [.95] - \hat{\theta}}{\hat{\theta} - \hat{\theta} [.05]}. 
$$

(1.6)

The standard intervals always have shape= 1.00. It is in this way that they err most seriously. For example the exact normal-theory interval for Corr has shape= .52, extending twice as far to the left of $\hat{\theta} = .723$ as to the right. The standard interval is much too optimistic about ruling out values of $\theta$ below $\hat{\theta}$, and much too pessimistic about ruling out values above $\hat{\theta}$. This kind of error is automatically identified and corrected by all the bootstrap confidence interval methods.

There is no compelling reason to assume bivariate normality for the data in Figure 1. A nonparametric version of (1.2) assumes that the pairs $(B_i, A_i)$ are a random sample ("i.i.d.") from some unknown bivariate distribution $F$,

$$
\begin{pmatrix} B_i \\ A_i \end{pmatrix} \overset{i.i.d.}{\sim} F, \quad (i = 1, 2, \cdots, n), 
$$

(1.7)

$n = 20$, without assuming that $F$ belongs to any particular parametric family. Bootstrap-based confidence intervals such as ABC are available for nonparametric situations, as discussed in Section 6. In theory they enjoy the same second-order accuracy as in parametric problems. However in some nonparametric confidence interval problems that have been examined carefully the small-sample advantages of the bootstrap methods have been less striking than in parametric situations. Methods that give third-order accuracy, like the bootstrap calibration of an ABC interval seem to be more worthwhile in the nonparametric framework, see Section 6.

In most problems and for most parameters there will not exist exact confidence intervals. This great gray area has been the province of the standard intervals for at least seventy years. Bootstrap confidence intervals provide a better approximation to exactness in most situations. Table 3 refers to the parameter $\theta$ defined as the maximum eigenvalue of the covariance matrix of $(B, A)$ in the cd4 experiment,

$$
\theta = \text{maximum eigenvalue } \{\text{cov}(B, A)\}. 
$$

(1.8)

The maximum likelihood estimate (MLE) of $\theta$, assuming either model (1.2) or (1.7), is $\hat{\theta} = 1.68$. The bootstrap intervals extend further to the right than the left of $\hat{\theta}$ in this case, more than 2.5 times as far under the normal model. Even though we have no exact endpoint to serve as a gold standard here, the theory that follows strongly suggests the superiority of the bootstrap
intervals. Bootstrapping involves much more computation than the standard intervals, on the order of 1000 times more, but the algorithms are completely automatic, requiring no more thought for the maximum eigenvalue than the correlation coefficient, or for any other parameter.

One of the achievements of the theory discussed in Section 8 is to provide a reasonable theoretical "gold standard" for approximate confidence intervals. Comparison with this gold standard shows that the bootstrap intervals are not only asymptotically more accurate than the standard intervals, they are also more correct. "Accuracy" refers to the coverage errors: a one-sided bootstrap interval of intended coverage \( \alpha \) actually covers \( \theta \) with probability \( \alpha + O(1/n) \), \( n \) being the sample size. This is second-order accuracy, compared to the slower first-order accuracy of the standard intervals, with coverage probabilities \( \alpha + O(1/\sqrt{n}) \). However confidence intervals are supposed to be inferentially correct as well as accurate. Correctness is a harder property to pin down, but it is easy to give examples of incorrectness: if \( x_1, x_2, \ldots, x_n \) is a random sample from a normal distribution \( N(\theta, 1) \), then \( (\min(x_i), \max(x_i)) \) is an exactly accurate two-sided confidence interval for \( \theta \) of coverage probability \( 1 - 1/2^{n-1} \), but it is incorrect. The theory of Section 8 shows that all of our better confidence intervals are second-order correct as well as second-order accurate. We can see this improvement over the standard intervals on the left side of Table 2. The theory says that this improvement exists also in those cases like Table 3 where we cannot see it directly.

<table>
<thead>
<tr>
<th></th>
<th>NORMAL THEORY</th>
<th>NONPARAMETRIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( .05 )</td>
<td>1.11</td>
<td>1.10</td>
</tr>
<tr>
<td>( .95 )</td>
<td>3.25</td>
<td>3.18</td>
</tr>
<tr>
<td>length</td>
<td>2.13</td>
<td>2.08</td>
</tr>
<tr>
<td>shape</td>
<td>2.80</td>
<td>2.62</td>
</tr>
</tbody>
</table>

Table 3. Approximate 90% central confidence intervals for the maximum eigenvalue parameter (1.7), cd4 data; the bootstrap intervals extend much further to the right of the MLE \( \hat{\theta} = 1.68 \) than to the left.

2. The \( BC_a \) Intervals. The next six sections give an heuristic overview of bootstrap confidence intervals. More examples are presented, showing how bootstrap intervals can be routinely constructed even in very complicated and messy situations. Section 8 derives the second-order properties of the bootstrap intervals in terms of asymptotic expansions. Comparisons with likelihood-based methods are made in Section 9. The bootstrap can be thought of as a convenient way of executing the likelihood calculations in parametric exponential family situations and even in nonparametric problems.

The bootstrap was introduced as a nonparametric device for estimating standard errors and biases. Confidence intervals are inherently more delicate inference tools. A considerable amount of effort has gone into upgrading bootstrap methods to the level of precision required for confidence intervals.
The $BC_a$ method is an automatic algorithm for producing highly accurate confidence limits from a bootstrap distribution. Its effectiveness was demonstrated in Table 2. References include Efron (1987), Hall (1988), DiCiccio (1984), DiCiccio and Romano (1995), and Efron and Tibshirani (1993). A program written in the language $S$ is available, see the footnote in Section 4.

The goal of bootstrap confidence interval theory is to calculate dependable confidence limits for a parameter of interest $\theta$ from the bootstrap distribution of $\hat{\theta}$. Figure 2 shows two such bootstrap distributions relating to the maximum eigenvalue parameter $\theta$ for the cd4 data, (1.8). The nonparametric bootstrap distribution on the right, will be discussed in Section 6.

The left panel is the histogram of 2000 normal-theory bootstrap replications of $\hat{\theta}$. Each replication was obtained by drawing a bootstrap data set analogous to (1.2),

$$\left( \frac{B^*_i}{A^*_i} \right)^{i.i.d.} \sim N_2(\lambda, \Gamma) \quad (i = 1, 2, \cdots, 20),$$

and then computing $\hat{\theta}^*$, the maximum likelihood estimate (MLE) of $\theta$ based on the bootstrap data. In other words $\hat{\theta}^*$ was the maximum eigenvalue of the empirical covariance matrix of the 20 pairs $(B^*_i, A^*_i)$. The mean vector $\lambda$ and covariance matrix $\Gamma$ in (2.1) were the usual maximum likelihood estimates for $\lambda$ and $\Gamma$, based on the original data in Figure 1. (2.1) is a parametric bootstrap sample, obtained by sampling from a parametric MLE for the unknown distribution $F$. Section 5 discusses nonparametric bootstrap samples and confidence intervals.

The 2000 bootstrap replications $\hat{\theta}^*$ had standard deviation 0.52. This is the bootstrap estimate of standard error for $\hat{\theta}$, generally a more dependable standard error estimate than the usual parametric delta-method value, see Efron (1981). The mean of the 2000 values was 1.61 compared to $\hat{\theta} = 1.68$, indicating a small downward bias in the Maxeig statistic. In this case it is easy to see that the downward bias comes from dividing by $n$ instead of $n - 1$ in obtaining the MLE $\hat{\Gamma}$ of the covariance matrix. Bootstrap confidence intervals automatically correct for this kind of bias, whether or not we can see it.

2000 bootstrap replications is ten times too many for estimating a standard error, but not too many for the more delicate task of setting confidence intervals. These bootstrap sample size calculations appear in Section 9 of Efron (1987).

$BC_a$ is a method of setting approximate confidence intervals for $\theta$ from the percentiles of the bootstrap histogram. Suppose $\theta$ is a parameter of interest; $\hat{\theta}(x)$ an estimate of $\theta$ based on the observed data $x$; and $\hat{\theta}^* = \hat{\theta}(x^*)$ a bootstrap replication of $\hat{\theta}$ obtained by resampling $x^*$ from an estimate of the distribution governing $x$. Let $\hat{G}(c)$ be the cumulative distribution function (cdf) of $B$ bootstrap replications $\hat{\theta}^*(b)$,

$$\hat{G}(c) = \#\{\hat{\theta}^*(b) < c\} / B.$$  \hspace{1cm} (2.2)

$B = 2000$ in our case. The upper endpoint $\hat{\theta}_{BC_a}[\alpha]$ of a one-sided level-$\alpha$ $BC_a$ interval, $\theta \in (-\infty, \hat{\theta}_{BC_a}[\alpha])$ is defined in terms of $\hat{G}$ and two numerical parameters discussed below: The bias-
correction $z_0$ and the acceleration $a$. ($BC_a$ stands for "bias-corrected and accelerated"). By definition the $BC_a$ endpoint is

$$\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1} \Phi \left( z_0 + \frac{z_0 + z(\alpha)}{1 - a(z_0 + z(\alpha))} \right).$$  \hfill (2.3)

**Figure 2.** Bootstrap distributions for the maximum eigenvalue of the covariance matrix, cd4 data. Left: 2000 parametric bootstrap replications assuming a bivariate normal distribution. Right: 2000 nonparametric bootstrap replications, discussed in Section 6. The solid lines indicate the limits of the $BC_a .90$ central confidence intervals, compared to the standard intervals, dashed lines.

Here $\Phi$ is the standard normal cdf, with $z(\alpha) = \Phi^{-1}(\alpha)$ as before. The central .90 $BC_a$ interval is given by $(\hat{\theta}_{BC_a}[.05], \hat{\theta}_{BC_a}[.95])$. Formula (2.3) looks strange, but it is well-motivated by the transformation and asymptotic arguments that follow.

If $a$ and $z_0$ are zero then $\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1}(\alpha)$, the $100\alpha$th percentile of the bootstrap replications. In this case the .90 $BC_a$ interval is the interval between the 5th and 95th percentiles of the bootstrap replications. If in addition $\hat{G}$ is perfectly normal, then $\hat{\theta}_{BC_a}[\alpha] = \hat{\theta} + z(\alpha)\hat{\sigma}$, the standard interval endpoint. In general, formula (2.3) makes 3 distinct corrections to the standard intervals, improving their coverage accuracy from first to second order.

$\hat{G}$ is markedly long-tailed to the right on the normal-theory side of Figure 2. Also $a$ and $z_0$ are both estimated to be positive, $(\hat{a}, \hat{z}_0) = (.105, .226)$, further shifting $\hat{\theta}_{BC_a}[\alpha]$ to the right of $\hat{\theta}_{\text{STAN}}[\alpha] = \hat{\theta} + z(\alpha)\hat{\sigma}$. The .90 $BC_a$ interval for $\theta$ is

$$\left( \hat{G}^{-1}(.157), \hat{G}^{-1}(.995) \right) = (1.10, 3.18),$$ \hfill (2.4)

7
comparing to the standard interval $(0.80, 2.55)$.

The following argument motivates the $BC_a$ definition (2.3), as well as the parameters $a$ and $z_0$. Suppose that there exists a monotone increasing transformation $\phi = m(\theta)$ such that $\hat{\phi} = m(\hat{\theta})$ is normally distributed for every choice of $\theta$, but possibly with a bias and a non-constant variance,

$$\hat{\phi} \sim N(\phi - z_0 \sigma_\phi, \sigma_\phi^2), \quad \sigma_\phi = 1 + a\phi.$$  \hspace{1cm} (2.5)

Then formula (2.3) gives exactly accurate and correct confidence limits for $\theta$ having observed $\hat{\theta}$.

The argument in Section 3 of Efron (1987) shows that in situation (2.5) there is another monotone transformation, say $\xi = M(\theta)$ and $\tilde{\xi} = M(\tilde{\theta})$, such that $\tilde{\xi} = \xi + W$ for all values of $\xi$, with $W$ always having the same distribution. This is a translation problem so we know how to set confidence limits $\tilde{\xi}[\alpha]$ for $\xi$,

$$\tilde{\xi}[\alpha] = \xi - W(1-\alpha),$$  \hspace{1cm} (2.6)

$W(1-\alpha)$ being the $100 \cdot (1 - \alpha)$th percentile of $W$. The $BC_a$ interval (2.3) is exactly equivalent to the translation interval (2.6), and in this sense it is correct as well as accurate.

The bias-correction constant $z_0$ is easy to interpret in (2.5) since

$$\text{Prob}\{\hat{\phi} < \phi\} = \Phi(z_0).$$  \hspace{1cm} (2.7)

Then $\text{Prob}\{\hat{\theta} < \theta\} = \Phi(z_0)$ because of monotonicity. The $BC_a$ algorithm, in its simplest form, estimates $z_0$ by

$$z_0 = \Phi^{-1}\{\frac{\#\{\hat{\phi}^*(b) < \hat{\theta}\}}{B}\},$$  \hspace{1cm} (2.8)

$\Phi^{-1}$ of the proportion of the bootstrap replications less than $\hat{\theta}$. Of the 2000 normal-theory bootstrap replications $\hat{\phi}^*$ shown in the left panel of Figure 3, 1179 were less than $\hat{\theta} = 1.68$. This gave $z_0 = \Phi^{-1}(0.593) = .226$, a positive bias correction since $\hat{\phi}^*$ is biased downward relative to $\hat{\theta}$. An often more accurate method of estimating $z_0$ is described in Section 4.

The acceleration $a$ in (2.5) measures how quickly the standard error is changing on the normalized scale. The value $\tilde{a} = 0.105$ in (2.4), obtained from formula (3.9) of Section 4 is moderately large. Suppose we think we have moved $1.645$ standard errors to the right of $\hat{\phi}$, to

$$\tilde{\phi} = \hat{\phi} + 1.645\sigma_\phi.$$  

Actually though, with $a = .105$,

$$\sigma_\phi = (1 + 1.645a)\sigma_\phi = 1.173\sigma_\phi$$

according to (2.5). For the purpose of calculating a confidence level, $\tilde{\phi}$ is really only $1.645/1.173 = 1.40$ standard errors to the right of $\hat{\phi}$, considerably less than $1.645$. Formula (2.3) automatically
corrects for an accelerating standard error. The next section gives a geometrical interpretation of $a$, as well as the $BC_a$ formula (2.3).

The peculiar-looking formula (2.3) for the $BC_a$ endpoints is designed to give exactly the right answer in situation (2.5), and to give it automatically in terms of the bootstrap distribution of $\hat{\theta}^*$. Notice for instance that the normalizing transformation $\hat{\varphi} = m(\hat{\theta})$ is not required in (2.3). By comparison, the standard interval works perfectly only under the more restrictive assumption that

$$\hat{\theta} \sim N(\theta, \sigma^2),$$

(2.9)

with $\sigma^2$ constant. In practice we don't expect either (2.9) or (2.5) to hold exactly, but the broader assumptions (2.5) are likely to be a better approximation to the truth. They produce intervals that are an order of magnitude more accurate, as shown in Section 8.

Formula (2.5) generalizes (2.9) in three ways, by allowing bias, nonconstant standard error, and a normalizing transformation. These three extensions are necessary and sufficient to give second order accuracy,

$$\operatorname{Prob}\{\theta < \hat{\theta}_{BC_a} [\alpha]\} = \alpha + O(1/n),$$

(2.10)

compared with $\operatorname{Prob}\{\theta < \hat{\theta}_{STAN} [\alpha]\} = \alpha + O(1/\sqrt{n})$, $n$ being the sample size in an i.i.d. sampling situation. This result is stated more carefully in Section 8, which also shows the second order correctness of the $BC_a$ intervals. Hall (1988) was the first to establish (2.10).

The $BC_a$ intervals are transformation invariant. If we change the parameter of interest from $\theta$ to some monotone function of $\theta$, $\phi = m(\theta)$, likewise changing $\hat{\theta}$ to $\hat{\phi} = m(\hat{\theta})$ and $\hat{\theta}^*$ to $\hat{\phi}^* = m(\hat{\theta}^*)$, then the $\alpha$-level $BC_a$ endpoints change in the same way,

$$\hat{\phi}_{BC_a} [\alpha] = m(\hat{\theta}_{BC_a} [\alpha]).$$

(2.11)

The standard intervals are not transformation invariant, and this accounts for some of their practical difficulties. It is well known for instance that normal-theory standard intervals for the correlation coefficient are much more accurate if constructed on the scale $\phi = \tanh^{-1}(\theta)$ and then transformed back to give an interval for $\theta$ itself. Transformation invariance means that the $BC_a$ intervals can't be fooled by a bad choice of scale. To put it another way, the statistician does not have to search for a transformation like $\tanh^{-1}$ in applying the $BC_a$ method.

In summary, $BC_a$ produces confidence intervals for $\theta$ from the bootstrap distribution of $\hat{\theta}^*$, requiring on the order of 2000 bootstrap replications $\hat{\theta}^*$. These intervals are transformation invariant and exactly correct under the normal transformation model (2.5); in general they are second order accurate and correct.

3. The Acceleration $a$. The acceleration parameter $a$ appearing in the $BC_a$ formula (3.2) looks mysterious. Its definition in (2.5) involves an idealized transformation to normality which will not be known in practice. Fortunately $a$ enjoys a simple relationship with Fisher's score
function which makes it easy to estimate. This section describes the relationship in the context of
one-parameter families. In doing so it also allows us to better motivate the peculiar-looking $BC_a$
formula (2.3).

Suppose then that we have a one-parameter family of cdf’s $G_\theta(\hat{\theta})$ on the real line, $\hat{\theta}$ being an
estimate of $\theta$. In the relationships below we assume that $\hat{\theta}$ behaves asymptotically like a maximum
likelihood estimator, with respect to a notional sample size $n$, as made explicit in (5.3) of Efron
(1987). As a particular example, we will consider the case
\[ \hat{\theta} \sim \theta \frac{\text{Gamma}_n}{n} \quad (n = 10), \] (3.1)
Gamma indicating a standard gamma variate with density $t^{n-1} \exp(-t) / \Gamma(n)$ for $t > 0$.

Having observed $\hat{\theta}$, we wonder with what confidence we can reject a trial value $\theta_0$ of the
parameter $\hat{\theta}$. In the gamma example (3.1) we might have
\[ \hat{\theta} = 1 \quad \text{and} \quad \theta_0 = 1.5. \] (3.2)
The easy answer from the bootstrap point of view is given in terms of the bootstrap cdf $\hat{G}(c) =
G_\theta(c)$. We can define the bootstrap confidence value to be
\[ \hat{\alpha} = \hat{G}(\theta_0) = G_\theta(\theta_0). \] (3.3)
However this will usually not agree with the more familiar hypothesis-testing confidence level for a
one-parameter problem, say
\[ \bar{\alpha} = 1 - G_{\theta_0}(\hat{\theta}), \] (3.4)
the probability under $\theta_0$ of getting a less extreme observation than $\hat{\theta}$. (For convenience these
definitions assume $\hat{\theta} < \theta_0$.) In case (3.1), (3.2) we have $\bar{\alpha} = .930$ while $\hat{\alpha} = .863$.

The $BC_a$ formula (2.3) amounts to a rule for converting bootstrap confidence values $\hat{\alpha}$ into
hypothesis-testing confidence levels $\bar{\alpha}$. This becomes crucial as soon as we try to use the bootstrap
on problems more complicated than one-parameter families. Define
\[ \hat{z} = \Phi^{-1}(\hat{\alpha}) \quad \text{and} \quad \bar{z} = \Phi^{-1}(\bar{\alpha}). \] (3.5)
For a given value of $\theta_0$ and $\bar{\alpha}$ above, let $\alpha = \bar{\alpha}$ and $\hat{\theta}_{BC_a}[\alpha] = \theta_0$ in (2.3). If formula (2.3) works
perfectly then we have
\[ \Phi^{-1}G(\theta_0) = \hat{z} = z_0 + \frac{z_0 + \hat{z}}{1 - a(z_0 + \hat{z})}, \] (3.6)
or
\[ \hat{z} = \frac{\hat{z} - z_0}{1 + a(\hat{z} - z_0)} - z_0. \] (3.7)
The fact that the $BC_a$ intervals are second-order accurate implies that the conversion formula (3.7)
itself must be quite accurate.
In order to use (3.7), or (2.3), we first must estimate the two parameters $z_0$ and $a$. The bias-correction $z_0$ is estimated by

$$
\hat{z}_0 = \Phi^{-1}G(\hat{\theta}) = \Phi^{-1}G_{\hat{\theta}}(\hat{\theta})
$$

(3.8)
as in (2.8). The acceleration $a$ is estimated in terms of the skewness of the score function

$$
\hat{\ell}_a(\hat{\theta}) = \frac{\partial}{\partial \theta} \log \{g_a(\hat{\theta})\},
$$

(3.9)
g$\theta_a(\hat{\theta})$ being the density $\partial G_{\theta}(\hat{\theta})/\partial \hat{\theta}$. Section 10 of Efron (1987) shows that one-sixth the skewness of $\hat{\ell}_a(\hat{\theta})$ evaluated at $\theta = \hat{\theta}$,

$$
\hat{a} = \text{SKEW}_{\theta_0}(\hat{\ell}_a(\hat{\theta}))/6
$$

(3.10)
is an excellent estimate of $a$.

Both $z_0$ and $a$ are of order $O(1/\sqrt{n})$, with the estimates $\hat{z}_0$ and $\hat{a}$ erring by $O(1/n)$. For the gamma problem (3.1) it is easy to calculate that

$$
\hat{z}_0 = .106 \quad \text{and} \quad \hat{a} = .105.
$$

(3.11)
If $\hat{\theta}$ is the MLE in a one-parameter family (but not in general) then $\hat{z}_0$ and $\hat{a}$ are nearly the same, as is the case here.

The usable form of (3.7) is

$$
\hat{z} = \frac{\hat{z} - \hat{z}_0}{1 + \hat{a}(\hat{z} - \hat{z}_0)} - \hat{z}_0.
$$

(3.12)
We can list three important properties of the $(\hat{z}, \hat{z})$ curve (3.12) near $\hat{z} = \hat{z}_0$:

$$
(\hat{z}, \hat{z}) = (\hat{z}_0 - \hat{z}_0) \quad \text{at} \quad \hat{z} = \hat{z}_0
$$

(3.13)
$$
\frac{d\hat{z}}{d\hat{z}} = 1 \quad \text{at} \quad \hat{z} = \hat{z}_0,
$$

(3.14)
and

$$
\frac{d^2\hat{z}}{d\hat{z}^2} = -2\hat{a} \quad \text{at} \quad \hat{z} = \hat{z}_0.
$$

(3.15)
The last of these relationships is of special interest here. It says that the curvature of the $(\hat{z}, \hat{z})$ curve at $\hat{z}_0$ is directly proportional to the acceleration $\hat{a}$.

In any given one-parameter problem we can find the actual $(\hat{z}, \hat{z})$ curve, at least in theory. This is obtained by keeping $\hat{\theta}$ fixed and varying the trial point $\theta_0$ in (3.3)-(3.5). Figure 3 shows the $(\hat{z}, \hat{z})$ curve for the gamma problem, with $\hat{\theta}$ any fixed value, say $\hat{\theta} = 1$. In this case the $BC_a$ approximation formula (3.12) matches the actual $(\hat{z}, \hat{z})$ curve to three decimal places over most of the range of the graph. At $\hat{\theta} = 1$, $\theta_0 = 1.5$ for example, $\hat{z}$ equals 1.092 both actually and from (3.12).

The fact that the $BC_a$ formula (2.3) is second-order accurate implies that the conversion formula (3.12) errs only by $O(1/n)$. This means that relationships (3.13)-(3.15) must have the same order of accuracy, even in quite general problems. In particular, the curvature of the actual $(\hat{z}, \hat{z})$ plot, if it were possible to compute it, would nearly equal $-2\hat{a}$, with $\hat{a}$ given by the skewness definition (3.10).
Figure 3. Plot of $\hat{z}$ versus $\tilde{z}$ in the gamma problem (3.1); the $BC_a$ approximation (3.12), or (2.3), matches the actual curve to three decimal places. The central curvature of the $(\hat{z},\tilde{z})$ plot is proportional to the acceleration $\hat{a}$.

None of this is special to one-parameter families except for the skewness definition (3.10), which does not allow for nuisance parameters. The next section shows how to extend the skewness definition of $\hat{a}$ to multiparameter situations. This gives an estimate that is easy to evaluate, especially in exponential families, and that behaves well in practice. In fact $a$ is usually easier to estimate than $z_0$, despite the latter's simpler definition.

4. The ABC Method. We now leave one-parameter families and return to the more complicated situations that bootstrap methods are intended to deal with. In many such situations it is possible to approximate the $BC_a$ interval endpoints analytically, entirely dispensing with Monte Carlo simulations. This reduces the computational burden by an enormous factor, and also makes it easier to understand how $BC_a$ improves upon the standard intervals. ABC, standing for Approximate Bootstrap Confidence intervals, is an analytic version of $BC_a$ applying to smoothly defined parameters in exponential families. It also applies to smoothly defined nonparametric problems, as shown in Section 6. DiCiccio and Efron (1992) introduced the ABC method, which is also discussed in Efron and Tibshirani (1993).

The $BC_a$ endpoints (2.3) depend on the bootstrap c.d.f. $\hat{G}$ and estimates of the two parameters $a$ and $z_0$. ABC requires one further estimate, of the nonlinearity parameter $c_q$, but it does not involve $\hat{G}$.
The standard interval (1.1) depends only on the two quantities \((\hat{\theta}, \hat{\sigma})\). The ABC intervals depend on the five quantities \((\hat{\theta}, \hat{\sigma}, \hat{\alpha}, \hat{z}_0, \hat{c}_q)\). Each of the three extra numbers \((\hat{\alpha}, \hat{z}_0, \hat{c}_q)\) corrects a deficiency of the standard method, making the ABC intervals second order accurate as well as second order correct.

The ABC system applies within multiparameter exponential families, which are briefly reviewed below. This framework includes most familiar parametric situations: normal, binomial, Poisson, gamma, multinomial, ANOVA, logistic regression, contingency tables, log linear models, multivariate normal problems, Markov chains, and also nonparametric situations as discussed in Section 4.

The density function for a \(p\)-parameter exponential family can be written as

\[
g_\mu(x) = e^{\eta' y - \psi(\eta)} \tag{4.1}
\]

where \(x\) is the observed data and \(y = Y(x)\) is a \(p\)-dimensional vector of sufficient statistics; \(\eta\) is the \(p\)-dimensional natural parameter vector; \(\mu\) is the expectation parameter \(\mu = E_\mu(y)\); and \(\psi(\eta)\), the cumulant generating function, is a normalizing factor that makes \(g_\mu(x)\) integrate to one.

The vectors \(\mu\) and \(\eta\) are in one-to-one correspondence so that either can be used to index functions of interest. In (4.1) for example, we used \(\mu\) to index the densities \(g\), but \(\eta\) to index \(\psi\). The ABC algorithm involves the mapping from \(\eta\) to \(\mu\), say

\[
\mu = \mu \eta(\eta), \tag{4.2}
\]

which, fortunately, has a simple form in all of the common exponential families. Section 3 of DiCiccio and Efron (1992) gives function (4.2) for several families, as well as specifying the other inputs necessary for use of the ABC algorithm.

The MLE of \(\mu\) in (3.1) is \(\hat{\mu} = y\), so that the MLE of a real-valued parameter of interest \(\theta = t(\mu)\) is

\[
\hat{\theta} = t(\hat{\mu}) = t(y). \tag{4.3}
\]

As an example consider the bivariate normal model (1.2). Here \(x = ((B_1, A_1), (B_2, A_2), \ldots, (B_{20}, A_{20}))\) and \(y = \sum_{i=1}^{20} (B_i, A_i, B_i^2, B_i A_i, A_i^2)' / 20\). The bivariate normal is a 5-parameter exponential family with

\[
\mu = (\lambda_1, \lambda_2, \lambda_1^2 + \Gamma_{11}, \lambda_1 \lambda_2 + \Gamma_{12}, \lambda_2^2 + \Gamma_{22})'. \tag{4.4}
\]

Thus the correlation coefficient is the function \(t(\mu)\) given by

\[
\theta = \frac{\mu_4 - \mu_1 \mu_2}{[(\mu_3 - \mu_1^2)(\mu_5 - \mu_2^2)]^{1/2}}; \tag{4.5}
\]

\(\hat{\theta} = t(\hat{\mu})\) is seen to be the usual sample correlation coefficient.
We denote the $p \times p$ covariance matrix of $y$ by $\Sigma(\mu) = \text{cov}_\mu \{y\}$, and let $\hat{\Sigma} = \hat{\Sigma}(\hat{\mu})$, the MLE of $\Sigma$. The delta-method estimate of standard error for $\hat{\theta} = t(\hat{\mu})$ depends on $\hat{\Sigma}$. Let $i$ denote the gradient vector of $\theta = t(\mu)$ at $\mu = \mu$,

$$i = (\cdots, \frac{\partial t}{\partial \mu_i}, \cdots)_{\mu = \hat{\mu}}.$$

Then

$$\hat{\sigma} = (i' \hat{\Sigma} i)^{1/2}$$

is the parametric delta-method estimate of standard error, and it is also the usual Fisher information standard error estimate.

The $\hat{\sigma}$ values for the standard intervals in Tables 2 and 3 were found by numerical differentiation, using

$$\frac{\partial t}{\partial \mu_i} \bigg|_{\hat{\mu}} = \frac{t(\hat{\mu} + \epsilon e_i) - t(\hat{\mu} - \epsilon e_i)}{2\epsilon}$$

for a small value of $\epsilon$, with $e_i$ the $i$th coordinate vector. The covariance matrix $\hat{\Sigma}$ is simple to calculate in most of the familiar examples, as shown in Section 3 of DiCiccio and Efron (1992), giving $\hat{\sigma}$ from (4.7). This assumes that $t(\mu)$ is differentiable. In fact we need $t(\mu)$ to be twice differentiable in order to carry out the ABC computations.

The ABC algorithm begins by computing $\hat{\sigma}$ from (4.7), (4.8). Then the constants $(a, z_0, c_0)$ are estimated by computing $p + 2$ numerical second derivatives. The first of these is

$$\hat{a} = \frac{\partial^2}{\partial \epsilon^2} [i' \mu (\hat{\eta} + \epsilon i)]_{\epsilon = 0} / 6 \hat{\sigma}^3,$$

when $\hat{\eta}$ is the MLE of the natural parameter vector $\eta$. This turns out to be the same as the skewness definition of $\hat{a}$, (3.10), in the one-parameter family obtained from Stein’s least favorable family construction, see (6.7) of Efron (1987). Formula (4.9) uses exponential family relationships to compute the skewness from a second derivation.

The second ABC numerical derivative is

$$\hat{c}_q = \frac{\partial^2}{\partial \epsilon^2} [i' \mu (\hat{\eta} + \epsilon i)]_{\epsilon = 0} / 2 \hat{\sigma};$$

$\hat{c}_q$ measures how nonlinear the parameter of interest $\theta$ is, as a function of $\mu$.

The final $p$ second derivatives are required for the bias-correction parameter $z_0$. The parametric delta-method estimate of bias for $\hat{\theta} = t(\hat{\mu})$ can be expressed as

$$\hat{b} = \frac{1}{2} \sum_{i=1}^{p} \frac{\partial^2}{\partial \epsilon^2} [i' \mu (\hat{\eta} + \epsilon d_i^{1/2} \gamma_i)]_{\epsilon = 0},$$

where $d_i$ is the $i$th eigenvalue and $\gamma_i$ is the $i$th eigenvector of $\hat{\Sigma}$. Then

$$\hat{z}_0 = \Phi^{-1}(2 \cdot \Phi(\hat{a}) \cdot \Phi(\hat{c}_q - \hat{b} / \hat{\sigma})) \div \hat{a} + \hat{c}_q - \hat{b} / \hat{\sigma}.$$

(4.12)
This involves terms other than \( \hat{\sigma} \) because \( z_0 \) relates to median bias. For the kind of smooth exponential family problems considered here, formula (4.12) is usually more accurate than the direct estimate (2.8).

The simplest form of the ABC intervals, called ABC\text{-}quadratic or ABCq, gives the \( \alpha \)-level endpoint directly as a function of the five numbers \((\hat{\theta}, \hat{\sigma}, \hat{\lambda}, \hat{z}_0, \hat{c}_q)\):

\[
\alpha \rightarrow w \equiv \hat{z}_0 + z^{(\alpha)} \rightarrow \lambda \equiv \frac{w}{(1 - \hat{\sigma} w)^2} \rightarrow \xi \equiv \lambda + \hat{c}_q \lambda^2
\]

\[
\rightarrow \hat{\theta}_{\text{ABCq}}[\alpha] = \hat{\theta} + \hat{\sigma} \xi.
\]

(4.13)

The original ABC endpoint, denoted \( \hat{\theta}_{\text{ABC}}[\alpha] \), requires one more recomputation of the function \( t(\cdot) \):

\[
\alpha \rightarrow w = \hat{z}_0 + z^{(\alpha)} \rightarrow \lambda = \frac{w}{(1 - \hat{\sigma} w)^2}
\]

\[
\rightarrow \hat{\theta}_{\text{ABC}}[\alpha] = t(\hat{\mu} + \lambda s t/\hat{\sigma}).
\]

(4.14)

Notice that \( \hat{c}_q \) is still required here, to estimate \( \hat{z}_0 \) in (4.12).

Formula (4.14) is the one used in Tables 2 and 3. It has the advantage of being transformation invariant, (2.11), and is sometimes more accurate than (4.13). However (4.13) is local, all of the recomputations of \( t(\mu) \) involved in (4.8)-(4.13) taking place infinitesimally near \( \hat{\mu} = y \). In this sense ABCq is like the standard method. Non-locality occasionally causes computational difficulties with boundary violations. In fact (4.13) is a simple quadratic approximation to (4.14), so ABC and ABCq usually agree reasonably well.

The main point of this article is that highly accurate approximate confidence intervals can now be calculated on a routine basis. The ABC intervals are implemented by a short computer algorithm.* There are five inputs to the algorithm: \( \hat{\mu}, \hat{\theta}, \hat{\sigma} \) and the functions \( t(\cdot) \) and \( \mu t(\cdot) \). The outputs include \( \hat{\theta}_{\text{STAN}}[\alpha], \hat{\theta}_{\text{ABC}}[\alpha] \) and \( \hat{\theta}_{\text{ABCq}}[\alpha] \). Computational effort for the ABC intervals is two or three times that required for the standard intervals.

The ABC intervals can be useful even in very simple situations. Suppose that the data consists of a single observation \( x \) from a Poisson distribution with unknown expectation \( \theta \). In this case \( \hat{\theta} = t(x) = x, \) and \( \hat{\sigma} = \sqrt{\hat{\theta}} \). Carrying through definitions (4.9)-(4.14) gives \( \hat{\alpha} = \hat{z}_0 = 1/(\hat{\theta}^{1/2}), \) \( \hat{c}_q = 0, \) and so

\[
\hat{\theta}_{\text{ABC}}[\alpha] = \hat{\theta} + \frac{w}{(1 - \hat{\sigma} w)^2} \sqrt{\hat{\theta} \left( w = \hat{z}_0 + z^{(\alpha)} \right)}.
\]

For \( x = 7 \), the interval \((\hat{\theta}_{\text{ABC}}[.05], \hat{\theta}_{\text{ABC}}[.95])\) equals \((3.54, 12.67)\). This compares with the exact interval \((3.57, 12.58)\) for \( \theta \), splitting the atom of probability at \( x = 7 \), and with the standard interval \((2.65, 11.35)\).

* The ABC intervals in Tables 2 and 3 were produced by the parametric and nonparametric ABC algorithms “abcpar” and “abcnon”. These and the \( BC_a \) program are available in the language S: send electronic mail to statlib@lib.stat.cmu.edu with the one-line message send bootstrap.funs from S.
Here is a more realistic example of the ABC algorithm, used in a logistic regression context. Table 4 shows the data from an experiment concerning mammalian cell growth. The goal of this experiment was to quantify the effects of two factors on the success of a culture. Factor "r" measures the ratio of two key constituents of the culture plate, while factor "d" measures how many days were allowed for culture maturation. A total of 1843 independent cultures were prepared, investigating 25 different \((r_i, d_j)\) combinations. The table lists \(s_{ij}\) and \(n_{ij}\) for each combination, the number of successful cultures, compared to the number attempted.

We suppose that the number of successful cultures is a binomial variate,

\[
s_{ij} \sim \text{binomial}(n_{ij}, \pi_{ij}) \quad i, j = 1, 2, 3, 4, 5, \tag{4.15}\]

with an additive logistic regression model for the unknown probabilities \(\pi_{ij},\)

\[
\log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \mu + \alpha_i + \beta_j \quad (\Sigma_i \alpha_i = \Sigma_j \beta_j = 0). \tag{4.16}\]

For the example here we take the parameter of interest to be

\[
\theta = \frac{\pi_{15}}{\pi_{51}} \tag{4.17}\]

the success probability for the lowest \(r\) and highest \(d\) divided by the success probability for the highest \(r\) and lowest \(d\). This typifies the kind of problem traditionally handled by the standard method.

A logistic regression program calculated maximum likelihood estimates \(\hat{\mu}, \hat{\alpha}, \hat{\beta}\), from which we obtained

\[
\hat{\theta} = \frac{1 + e^{-(\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1)}}{1 + e^{-(\hat{\mu} + \hat{\alpha}_5 + \hat{\beta}_5)}} = 4.16. \tag{4.18}\]

The output of the logistic regression program provided \(\hat{\mu}, \hat{\alpha}, \hat{\beta}\) for the ABC algorithm. Section 3 of DiCiccio and Efron gives the exact specification for an ABC analysis of a logistic regression problem. Applied here, the algorithm gave standard and ABC .90 central intervals for \(\theta,\)

\[
\begin{array}{cccccc}
 & d_1 & d_2 & d_3 & d_4 & d_5 & \text{Total} \\
 r_1 & 5/31 & 3/28 & 20/45 & 24/47 & 29/35 & 81/186 \\
 r_2 & 15/77 & 36/78 & 43/71 & 56/71 & 66/74 & 216/371 \\
 r_3 & 48/126 & 68/116 & 145/171 & 98/119 & 114/129 & 473/661 \\
 r_4 & 29/92 & 35/52 & 57/85 & 38/50 & 72/77 & 231/356 \\
 r_5 & 11/53 & 20/52 & 20/48 & 40/55 & 52/61 & 143/269 \\
 \hline 
 \text{Total} & 108/379 & 162/326 & 285/420 & 256/342 & 333/376 & 1144/1843 \\
\end{array}
\]

**Table 4.** Cell data. 1843 cell cultures were prepared, varying two factors, "r" the ratio of two key constituents, and "d" the number of days of culturing. Data shown are \(s_{ij}\) and \(n_{ij}\), the number of successful cultures and the number of cultures attempted, at the \(i\)th level of \(r\) and the \(j\)th level of \(d\).
\[(\hat{\theta}_{STAN}[.05], \hat{\theta}_{STAN}[.95]) = (3.06, 5.26)\]
\[(\hat{\theta}_{ABC}[.05], \hat{\theta}_{ABC}[.95]) = (3.20, 5.43)\]

(4.19)

The ABC limits are shifted moderately upwards relative to the standard limits, enough to make the shape (1.6) equal 1.32. The standard intervals are not too bad in this case, though better performance might have been expected with \(n = 1843\) data points. In fact it is very difficult to guess apriori what constitutes a large enough sample size for adequate standard-interval performance.

The ABC formulas (4.13), (4.14) were derived as second-order approximations to the \(BC_a\) endpoints by DiCiccio and Efron (1992). They showed that these formulas give second-order accuracy as in (2.10), and also second-order correctness. Section 8 reviews some of these results. There are many other expressions for ABC-like interval endpoints that enjoy equivalent second-order properties in theory, though they may be less dependable in practice. A particularly simple formula is

\[
\hat{\theta}_{ABC}[\alpha] = \hat{\theta}_{STAN}[\alpha] + \hat{\sigma}(\hat{z}_0 + (2\hat{a} + \hat{c}_q)z^2(\alpha)^2).
\]

(4.20)

This shows that the ABC endpoints are not just a translation of \(\hat{\theta}_{STAN}[\alpha]\).

In repeated sampling situations the estimated constants \((\hat{a}, \hat{z}_0, \hat{c}_q)\) are of stochastic order \(1/\sqrt{n}\) in the sample size, the same as \(\hat{\sigma}\). They multiply \(\hat{\sigma}\) in (4.20), resulting in corrections of order \(\hat{\sigma}/\sqrt{n}\) to \(\hat{\theta}_{STAN}[\alpha]\). If there were only \(1/4\) as much cell data, \(n = 461\), but with the same proportion of successes in every cell of Table 4, then \((\hat{a}, \hat{z}_0, \hat{c}_q)\) would be twice as large. This would double the relative difference \((\hat{\theta}_{ABC}[\alpha] - \hat{\theta}_{STAN}[\alpha])/\hat{\sigma}\) according to (4.20), rendering \(\hat{\theta}_{STAN}[\alpha]\) quite inaccurate.

Both \(\hat{a}\) and \(\hat{z}_0\) are transformation invariant, retaining the same numerical value under monotone parameter transformations \(\phi = m(\theta)\). The nonlinearity constant \(\hat{c}_q\) is not invariant, and it can be reduced by transformations that make \(\phi\) more linear as a function of \(\mu\). Changing parameters from \(\theta = \pi_{15}/\pi_{51}\) to \(\phi = \log(\theta)\) changes \((\hat{a}, \hat{z}_0, \hat{c}_q)\) from \((-0.06, -0.025, 0.105)\) to \((-0.06, -0.025, 0.025)\) for the cell data. The standard intervals are nearly correct on the \(\phi\) scale. ABC and \(BC_a\) automate this kind of data-analytic trick.

We can visualize the relationship between the \(BC_a\) and \(ABC\) intervals in terms of Figure 3. \(BC_a\) uses Monte Carlo bootstrapping to find \(\hat{z}\), as in (3.3), (3.5), and then maps \(\hat{z}\) into an appropriate hypothesis-testing value \(\hat{\xi}\) via formula (3.7). \(ABC\) also uses formula (3.7) (or, equivalently, (2.3)), but in order to avoid Monte Carlo computations it makes one further analytic approximation: \(\hat{\xi}\) itself, the point on the orizontal axis in Figure 3, is estimated from an Edgeworth expansion. The information needed for the Edgeworth expansion is obtained from the second derivatives (4.9)-(4.11).
5. Bootstrap—t Intervals. The $BC_a$ formula strikes some people as complicated, and also "unbootstraplike" since the estimate $\hat{a}$ is not obtained directly from bootstrap replications. The bootstrap—t method, another bootstrap algorithm for setting confidence intervals, is conceptually simpler than $BC_a$. The method was suggested in Efron (1979), but some poor numerical results reduced its appeal. Hall's (1988) paper showing the bootstrap—t's good second-order properties has revived interest in its use. Singh gave the first proof of second-order accuracy for the bootstrap—t in (1981).

Suppose that a data set $x$ gives an estimate $\hat{\theta}(x)$ for a parameter of interest $\theta$, and also an estimate $\hat{\sigma}(x)$ for the standard deviation of $\hat{\theta}$. By analogy with student’s $t$ statistic, we define

$$T = \frac{\hat{\theta} - \theta}{\hat{\sigma}},$$

and let $T^{(a)}$ indicate the 100$a$th percentile of $T$. The upper endpoint of an $a$-level one-sided confidence interval for $\theta$ is

$$\hat{\theta} - \hat{\sigma}T^{(1-a)}.$$  

(5.2)

This assumes we know the $T$ percentiles, as in the usual student’s $t$ case where $T^{(a)}$ is the percentile of a $t$ distribution. However the $T$ percentiles are unknown in most situations.

The idea of the bootstrap—t is to estimate the percentiles of $T$ by bootstrapping. First, the distribution governing $x$ is estimated and the bootstrap data sets $x^*$ are drawn from the estimated distribution, as in (2.1). Each $x^*$ gives both a $\hat{\theta}^*$ and a $\hat{\sigma}^*$, yielding

$$T^* = \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*},$$

(5.3)

a bootstrap replication of (5.1). A large number $B$ of independent replications gives estimated percentiles

$$\hat{T}^{(a)} = B \cdot a^{\text{th}} \text{ ordered value of } \{T^*(b), b = 1, 2, \ldots, B\}.$$ 

(5.4)

(So if $B = 2000$ and $a = .95$ then $\hat{T}^{(a)}$ is the 1900th ordered $T^*(b)$.) The 100$a$th bootstrap—t confidence endpoint $\hat{\theta}_T[a]$ is defined to be

$$\hat{\theta}_T[a] = \hat{\theta} - \hat{\sigma}T^{(1-a)},$$

(5.5)

following (5.2).

Figure 4 relates to the correlation coefficient for the cd4 data. The left panel shows 2000 normal-theory bootstrap replications of
\[ T = \frac{\hat{\theta} - \theta}{\hat{\sigma}}, \quad \hat{\sigma} = (1 - \hat{\theta}^2)/\sqrt{20}. \] (5.6)

Each replication required drawing \([(B_1^*, A_1^*), \ldots, (B_{20}^*, A_{20}^*)]\) as in (2.1), computing \(\hat{\sigma}^*\) and \(\hat{\sigma}^*\), and then calculating the bootstrap–\(t\) replication \(T^* = (\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*\). The precentiles \((\hat{T}^{(.05)}, \hat{T}^{(.95)})\) equalled \((-1.38, 2.62)\), giving a .90 central bootstrap–\(t\) interval of \((.45, .87)\). This compares nicely with the exact interval \((.47, .86)\) in Table 2.

Hall (1988) showed that the bootstrap–\(t\) limits are second order accurate, as in (2.10). DiCiccio and Efron (1992) showed that they are also second order correct, see Section 8.

Definition (2.17) uses the fact that \((1 - \hat{\theta}^2)/\sqrt{n}\) is a reasonable normal-theory estimate of standard error for \(\hat{\theta}\). In most situations \(\hat{\sigma}^*\) must be numerically computed for each bootstrap data set \(x^*\), perhaps using the delta method. This multiplies the bootstrap computations by a factor of at least \(p + 1\), where \(p\) is the number of parameters in the probability model for \(x\). The nonparametric bootstrap–\(t\) distribution on the right side of Figure 4 used \(\hat{\sigma}^*\) equal to the nonparametric delta-method estimates. The main disadvantage of both \(BC_a\) and Bootstrap–\(t\) is the large computational burden. This does not make much difference for the correlation coefficient, but it can become crucial for more complicated situations. The ABC method is particularly useful in complicated problems.

More seriously, the bootstrap–\(t\) algorithm can be numerically unstable, resulting in very long confidence intervals. This is a particular danger in nonparametric situations. As a rough rule of thumb, the \(BC_a\) intervals are more conservative than bootstrap–\(t\), tending to stay, if anything, too close to the standard intervals as opposed to deviating too much.

Bootstrap–\(t\) intervals are not transformation invariant. The method seems to work better if \(\theta\) is a translation parameter, like a median or an expectation. A successful application of the type appears in Section 9 of Efron (1981b). Tibshirani (1988) proposed an algorithm for transforming \(\theta\) to a more translation-like parameter \(\phi = m(\theta)\), before applying the bootstrap–\(t\) method. Then the resulting interval is transformed back to the \(\theta\) scale via \(\theta = m^{-1}(\phi)\). See Section 2.b of DiCiccio and Romano, or Section 12.6 of Efron and Tibshirani (1993).

The bootstrap–\(t\) and \(BC_a\) methods look completely different. However, surprisingly, the ABC method connects them.
Figure 4. Bootstrap-\(t\) distributions relating to \(\theta\) the \text{cd}4 data correlation. Left: 2000 normal-theory bootstrap replicates of \(T\) using \(\hat{\sigma}^* = (1 - \hat{\theta}^*)^2/\sqrt{20}\). Right: 2000 nonparametric bootstrap replicates of \(T\) using \(\hat{\sigma}^*\) given by the nonparametric delta method. Dashed lines show 5th and 95th percentiles.

ABC was introduced as a non-Monte Carlo approximation to \(BC_a\), but it can also be thought of as an approximation to the bootstrap-\(t\) method. The relationships in (4.13) can be reversed to give the attained significance level (ASL) \(\alpha\) for any observed data set. That is, we can find \(\alpha\) such that \(\hat{\theta}_{ABCq}[\alpha]\) equals the true \(\theta\):

\[
\begin{align*}
\theta \rightarrow \xi = \frac{\theta - \hat{\theta}}{\hat{\sigma}} \Rightarrow \lambda = \frac{2\xi}{1 + (1 + 4\hat{\sigma}^2\xi^2)^{1/2}} \Rightarrow w = \frac{2\lambda}{(1 + 2\hat{\sigma}^2\lambda) + (1 + 4\hat{\sigma}^2\lambda)^{1/2}} \Rightarrow \alpha = \Phi(w - \hat{\lambda}_0).
\end{align*}
\]

(5.7)

If the \(ABCq\) method works perfectly then the ASL as defined by (5.7) will be uniformly distributed over [0, 1], so

\[
Z = \Phi^{-1}(\alpha)
\]

(5.8)

will be distributed as a \(N(0, 1)\) variate.

Notice that \(T\) in (5.1) equals \(-\xi\) in (5.7). The \(ABCq\) method amounts to assuming that

\[
h_{\hat{\sigma}, \hat{\theta}}(T) \sim N(0, 1)
\]

(5.9)

for the transformation defined by (5.7), (5.8). In other words, \(ABCq\) uses an estimated transformation of \(T\) to get a pivotal quantity. The bootstrap-\(t\) method assumes that \(T\) itself is pivotal, but
then finds the pivotal distribution by bootstrapping. The calibration method discussed in Section 7 uses both an estimated transformation and bootstrapping, with the result being still more accurate intervals.

6. Nonparametric Confidence Intervals. The $BC_a$, bootstrap-$t$, and ABC methods can be applied to the construction of nonparametric confidence intervals. Here we will discuss the one-sample nonparametric situation where the observed data $x = (x_1, x_2, \cdots, x_n)$ are a random sample from an arbitrary probability distribution $F$,

$$x_1, x_2, \cdots, x_n \sim F.$$  \hspace{1cm} (6.1)

The sample space $\mathcal{X}$ of the distribution can be anything at all. $\mathcal{X}$ is the two-dimensional Euclidean space $\mathbb{R}^2$ in (1.7) and on the right side of Table 1, and is an extended version of $\mathbb{R}^3$ in the missing-value example below. Multi-sample nonparametric problems are mentioned briefly at the end of this section.

The empirical distribution $\hat{F}$ puts probability $1/n$ on each sample point $x_i$ in $x$. A real-valued parameter of interest $\theta = t(F)$ has the nonparametric estimate

$$\hat{\theta} = t(\hat{F}),$$  \hspace{1cm} (6.2)

also called the nonparametric maximum likelihood estimate. A nonparametric bootstrap sample $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$ is a random sample of size $n$ drawn from $\hat{F}$,

$$x_1^*, x_2^*, \cdots, x_n^* \sim \hat{F}.$$  \hspace{1cm} (6.3)

In other words, $x^*$ equals $(x_{j_1}, x_{j_2}, \cdots, x_{j_n})$ where $j_1, j_2, \cdots, j_n$ is a random sample drawn with replacement from $\{1, 2, \cdots, n\}$. Each bootstrap sample gives a nonparametric bootstrap replication of $\hat{\theta}$,

$$\hat{\theta}^* = t(\hat{F}^*),$$  \hspace{1cm} (6.4)

where $\hat{F}^*$ is the empirical distribution of $x^*$.

Nonparametric $BC_a$ confidence intervals for $\theta$ are constructed the same way as the parametric intervals of Section 2. A large number of independent bootstrap replications $\hat{\theta}^*(1), \hat{\theta}^*(2), \cdots, \hat{\theta}^*(B)$ are drawn according to (4.3), (4.4), $B \approx 2000$, giving a bootstrap cumulative distribution function $\hat{G}(c) = \#(\hat{\theta}^*(b) < c)/B$. The $BC_a$ endpoints $\hat{\theta}_{BC_a}^*[a]$ are then calculated from formula (2.3), plugging in nonparametric estimates of $z_0$ and $a$.

Formula (2.8) gives $z_0$, which can also be obtained from a nonparametric version of (4.12). The acceleration $a$ is estimated using the empirical influence function of the statistic $\hat{\theta} = t(\hat{F})$,

$$U_i = \lim_{\epsilon \to 0} \frac{t((1 - \epsilon)\hat{F} + \epsilon \delta_i)}{\epsilon}, \quad i = 1, 2, \cdots, n.$$  \hspace{1cm} (6.5)
Here \( \delta_i \) is a point mass on \( x_i \), so \((1 - \epsilon)\hat{F} + \epsilon\delta_i\) is a version of \( \hat{F} \) putting extra weight on \( x_i \) and less weight on the other points. The usual nonparametric delta method estimate of standard error is \( (\sum U_i^2 / n^2)^{1/2} \), this being the value used in our examples of the standard interval (1.1).

The estimate of \( a \) is

\[
\hat{a} = 1 \frac{\sum_{i=1}^{n} U_i^3}{6 (\sum_{i=1}^{n} U_i^2)^{3/2}}. \tag{6.6}
\]

This looks completely different than (4.9), but in fact it is the same formula, applied here in a multinomial framework appropriate to the nonparametric situation. The similarity of formula (6.6) to a skewness reflects the relationship of \( \hat{a} \) to the skewness of the score function, (3.10). The connection of nonparametric confidence intervals with multinomial estimation problems appears in Section 7 and 8 of Efron (1989).

There is a simpler way to calculate the \( U_i \) and \( \hat{a} \). Instead of (6.5) we can use the jackknife influence function

\[
U_i = (n - 1)(\hat{\theta} - \hat{\theta}_{(i)}) \tag{6.7}
\]

in (6.6) where \( \hat{\theta}_{(i)} \) is the estimate of \( \theta \) based on the reduced data set \( x_{(i)} = (x_1, x_2, \cdots, x_{n-1}, x_{n+1}, \cdots, x_n) \). This makes it a little easier to calculate the \( BC_a \) limits since the statistic \( \hat{\theta}(x) \) does not have to be reprogrammed in the functional form \( \hat{\theta} = t(F) \).

The nonparametric \( BC_a \) method is unfazed by complicated sample spaces. Table 5 shows an artificial missing-data example discussed in Efron (1993). Twenty-two students have each taken 5 exams labelled A, B, C, D, E, but some of the A and E scores, marked "?", are missing. If there were no missing data we would consider the rows of the matrix to be a random sample of size \( n = 22 \) from an unknown 5-dimensional distribution \( F \). Our goal is to estimate

\[
\theta = \text{maximum eigenvalue of } \Sigma, \tag{6.8}
\]

where \( \Sigma \) is the covariance matrix of \( F \).

An easy way, though not necessarily the best way, to fill in Table 5 is to fit a standard two-way additive model \( \mu + \alpha_i \beta_j \) to the non-missing scores by least squares, and then to replace the missing values \( x_{ij} \) by

\[
\hat{x}_{ij} = \hat{\nu} + \hat{\alpha}_i + \hat{\beta}_j. \tag{6.9}
\]

The filled-in \( 22 \times 5 \) data matrix has rows \( \hat{x}_i, i = 1, 2, \cdots, 22, \) from which we can calculate

\[
\hat{F} = \frac{1}{22} \sum_{i=1}^{22} (\hat{\epsilon}_i - \hat{\mu}_i)(\hat{x}_i - \hat{\mu}_i)'; \quad (\hat{\mu} = \frac{1}{22} \sum_{i=1}^{22} \hat{x}_i) \tag{6.10}
\]

giving the point estimate

\[
\hat{\theta} = \text{maximum eigenvalue of } \hat{F} = 633.2. \tag{6.11}
\]

How accurate is \( \hat{\theta} \)?
<table>
<thead>
<tr>
<th>student</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>?</td>
<td>63</td>
<td>65</td>
<td>70</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>61</td>
<td>72</td>
<td>64</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>67</td>
<td>65</td>
<td>65</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>?</td>
<td>69</td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>?</td>
<td>69</td>
<td>61</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>?</td>
<td>49</td>
<td>62</td>
<td>63</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>61</td>
<td>52</td>
<td>62</td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>41</td>
<td>61</td>
<td>49</td>
<td>?</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>69</td>
<td>50</td>
<td>52</td>
<td>45</td>
</tr>
<tr>
<td>10</td>
<td>?</td>
<td>59</td>
<td>51</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>11</td>
<td>?</td>
<td>40</td>
<td>56</td>
<td>54</td>
<td>?</td>
</tr>
<tr>
<td>12</td>
<td>42</td>
<td>60</td>
<td>54</td>
<td>49</td>
<td>?</td>
</tr>
<tr>
<td>13</td>
<td>?</td>
<td>63</td>
<td>53</td>
<td>54</td>
<td>?</td>
</tr>
<tr>
<td>14</td>
<td>?</td>
<td>55</td>
<td>59</td>
<td>53</td>
<td>?</td>
</tr>
<tr>
<td>15</td>
<td>?</td>
<td>49</td>
<td>45</td>
<td>48</td>
<td>?</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>53</td>
<td>57</td>
<td>43</td>
<td>51</td>
</tr>
<tr>
<td>17</td>
<td>39</td>
<td>46</td>
<td>46</td>
<td>32</td>
<td>?</td>
</tr>
<tr>
<td>18</td>
<td>48</td>
<td>38</td>
<td>41</td>
<td>44</td>
<td>33</td>
</tr>
<tr>
<td>19</td>
<td>46</td>
<td>40</td>
<td>47</td>
<td>29</td>
<td>?</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>34</td>
<td>43</td>
<td>46</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>?</td>
<td>30</td>
<td>32</td>
<td>35</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>?</td>
<td>26</td>
<td>15</td>
<td>20</td>
<td>?</td>
</tr>
</tbody>
</table>

Table 5. 22 students have each taken 5 exams, labelled A, B, C, D, E. Some of the scores for A and E, indicated by "?" are missing. Original data set from Kent, Mardia and Bibby (1979).

It is easy to carry out a nonparametric $BC_a$ analysis. The "points" $x_i$ in the data set $x = (x_1, x_2, \ldots, x_n), n = 22$, are the rows of Table 5, for instance $x_{22} = (?, 26, 15, 20, ?)$. A bootstrap data set $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ is a $22 \times 5$ data matrix, each row of which has been randomly selected from the rows of Table 5. Having selected $x^*$, the bootstrap replication $\hat{\theta}^*$ is computed by following the same steps (4.9)-(4.11) that gave $\hat{\theta}$. Figure 5 is a histogram of 2200 bootstrap replications $\hat{\theta}^*$, the histogram being noticeably long-tailed toward the right. The .90 $BC_a$ confidence interval for $\theta$ is

$$(\hat{\theta}_{BC_a}(.05), \hat{\theta}_{BC_a}(.95)) = (379, 1164),$$

extending twice as far to the right of $\hat{\theta}$ as to the left.

It is easy to extend the ABC method of Section 4 to nonparametric problems, greatly reducing the computational burden of the $BC_a$ intervals. The formulas are basically the same as in (4.9)-(4.14), but they simplify somewhat in the nonparametric multinomial framework. The statistic is expressed in the functional form $\theta = t(F)$ and then reevaluated for values of $\theta$ very near $\hat{F}$, as in (6.5). The ABC limits require only $2n + 4$ reevaluation of the statistic. By comparison, the $BC_a$ method requires some 2000 evaluations $\hat{\theta}^* = t(\hat{F}^*)$, where $\hat{F}^*$ is a bootstrap empirical distribution.

The nonparametric ABC algorithm "abcon" was applied to the maximum eigenvalue statistic for the student score data. After 46 reevaluations of the statistic defined by (6.9)-(6.11), it gave .90 central confidence interval
Figure 5. Histogram of 2200 nonparametric bootstrap replications of the maximum eigenvalue statistic for the student score data; bootstrap standard error estimate $\hat{\sigma} = 212.0$. The histogram is long-tailed to the right, and so is the $BC_\alpha$ confidence interval (6.12).

\[
(\hat{\theta}_{ABC[.05]}, \hat{\theta}_{ABC[.95]}) = (379, 1172),
\]

(6.13) nearly the same as (6.12). The statlib program abcon used here appears in the appendix to Efron (1994). Efron (1993) also applied abcon to the full normal theory MLE of $\theta$, (6.5), rather than to the ad hoc estimator (6.9)-(6.11). The resulting ABC interval (353, 1307) was 20% longer than (6.13), perhaps undermining belief in the data’s normality.

So far we have only discussed one-sample nonparametric problems. The $K$-sample nonparametric problem has data

\[
x_{k1}, x_{k2}, \ldots, x_{kn_k} \overset{\text{i.i.d.}}{\sim} F_k \quad \text{for} \quad k = 1, 2, \ldots, K
\]

(6.14) for arbitrary probability distributions $F_k$ on possibly different sample spaces $X_k$. The nonparametric MLE of a real-valued parameter of interest $\theta = t(F_1, F_2, \ldots, F_K)$ is

\[
\hat{\theta} = t(\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_K),
\]

(6.15) where $\hat{F}_k$ is the empirical distribution corresponding to $x_k = (x_{k1}, x_{k2}, \ldots, x_{kn_k})$.

It turns out that $K$-sample nonparametric confidence intervals can easily be obtained from either abcon or bcanon, its nonparametric $BC_\alpha$ counterpart. How to do so is explained in remarks C and H of Efron (1994).
7. Calibration. Calibration is a bootstrap technique for improving the coverage accuracy of any system of approximate confidence intervals. Here we will apply it to the nonparametric ABC intervals in Tables 2 and 3. The general theory is reviewed in Efron and Tibshirani (1993), Sections 18.3 and 25.6, following ideas of Loh (1987), Beran (1987), Hall (1986), and Hall and Martin (1988).

Let \( \hat{\theta} [\alpha] \) be the upper endpoint of a one-sided level–\( \alpha \) approximate confidence interval for parameter \( \theta \). If the approximation is actually working perfectly then the true probability of coverage

\[
\beta(\alpha) \equiv \text{Prob}\{\theta < \hat{\theta} [\alpha]\}
\]

(7.1)

will equal \( \alpha \). If not, we could use the calibration curve \( \beta(\alpha) \) to improve the approximate confidence intervals. For example if \( \beta [0.03] = 0.05 \) and \( \beta [0.98] = 0.95 \), then we could use \((\hat{\theta} [0.03], \hat{\theta} [0.98])\) instead of \((\hat{\theta} [0.05], \hat{\theta} [0.95])\) as our approximate central .90 interval.

Of course we do not know the calibration curve \( \beta(\alpha) \). The interesting fact is that we can apply the bootstrap to estimate \( \beta(\alpha) \), and then use the estimate to improve our original approximate intervals. The estimated calibration curve is

\[
\hat{\beta}(\alpha) = \text{Prob}^*\{\hat{\theta} < \hat{\theta} [\alpha]^*\}.
\]

(7.2)

\( \text{Prob}^* \) indicates bootstrap sampling as in (2.1) or (6.3) (so \( \hat{\theta} \) is fixed) with \( \hat{\theta} [\alpha]^* \) being the upper \( \alpha \) endpoint based on the bootstrap data.

It looks like we have to do separate bootstrap calculations in (7.2) for every value of \( \alpha \), but that is unnecessary if \( \hat{\theta} [\alpha] \) is an increasing function of \( \alpha \), as it usually is. For a given bootstrap sample, let \( \hat{\alpha}^* \) be the value of \( \alpha \) that makes the upper endpoint equal \( \alpha \),

\[
\hat{\alpha}^* : \hat{\theta} [\hat{\alpha}^*]^* = \hat{\theta}.
\]

(7.3)

Then the event \( \{\hat{\alpha}^* < \alpha\} \) is equivalent to the event \( \{\hat{\theta} < \hat{\theta} [\alpha]^*\} \), so

\[
\hat{\beta}(\alpha) = \text{Prob}^*\{\hat{\alpha}^* < \alpha\}.
\]

(7.4)

In order to calibrate a system of approximate confidence intervals we generate \( B \) bootstrap samples and for each one we calculate \( \hat{\alpha}^* \). The estimated calibration curve is

\[
\hat{\beta}(\alpha) = \#\{\hat{\alpha}^*(b) < \alpha\}/B.
\]

(7.5)

In other words we estimate the cdf of \( \hat{\alpha}^* \). If the cdf is nearly uniform, \( \hat{\beta}(\alpha) \approx \alpha \), then this indicates accurate coverage for our system of intervals. If not, we can use \( \hat{\beta}(\alpha) \) to improve the original endpoints by calibration.
Figure 6. Estimated calibration curves for the nonparametric ABC method, cd4 data. Left panel: correlation coefficient as in Table 2. Right panel: maximum eigenvalue as in Table 3; each based on 2000 bootstrap replications.

This idea was applied to the nonparametric ABC intervals of Tables 2 and 3, the correlation coefficient and maximum eigenvalue statistic for the cd4 data. Figure 6 shows the result of $B = 2000$ bootstrap replications for each situation. The calibration shows good results for the correlation coefficient, with $\hat{\beta}(\alpha) \approx \alpha$ over the full range of $\alpha$. The story is less pleasant for the maximum eigenvalue. At the upper end of the scale we have $\hat{\beta}(\alpha) < \alpha$, indicating that we need to take $\alpha > .95$ to get actual 95\% coverage. According to Table 6, which shows the percentiles of the $\hat{\alpha}^*$ distributions, we should take $\alpha = .994$. This kind of extreme correction is worrisome, but it produces an interesting result in Table 3: it moves the upper endpoint of the nonparametric interval much closer to the normal theory value 3.25.

Calibrating the ABC intervals improves their accuracy from second to third order, with coverage errors, as in (2.10), reduced to $O(1/n^{3/2})$. We are talking about a lot of computation here, on the order of 1000 times as much as for the ABC intervals themselves. The computational efficiency of ABC compared to $BC_{\alpha}$ becomes crucial in the calibration context. Calibrating the $BC_{\alpha}$ intervals would require on the order of 1,000,000 recomputations of the original statistic $\hat{\theta}$.
Table 6. Percentiles of the distributions of \( \hat{\sigma}^2 \) shown in Figure 6. The .05 and .95 values were used for the calibrated ABC endpoints in Tables 2 and 3.

8. Second Order Accuracy and Correctness. In order to validate the second-order accuracy and correctness of bootstrap confidence intervals we need asymptotic expansions for the cumulative distribution functions of \( \tilde{\theta} \) and \( T = (\hat{\theta} - \theta)/\hat{\sigma} \). Later these expressions will be used to connect bootstrap theory to several other second-order confidence interval methods. In many situations, including those considered in the preceding sections, the asymptotic distribution of
\[
U = (\hat{\theta} - \theta)/\sigma
\]
is standard normal, and the first three cumulants of \( U \) are given by
\[
E(U) = k_1/\sqrt{n}, \quad \text{var}(U) = 1, \quad \text{skew}(U) = k_3/\sqrt{n},
\]
where \( k_1 \) and \( k_3 \) are of order \( O(1) \); the fourth- and higher-order cumulants are of order \( O(n^{-1}) \) or smaller. It follows that the first three cumulants of
\[
T = \frac{(\hat{\theta} - \theta)}{\hat{\sigma}} = U \left\{ 1 - \frac{1}{2} \frac{(\hat{\sigma}^2 - \sigma^2)}{\sigma^3} \right\} + O_p(n^{-1})
\]
are given by
\[
E(T) = (k_1 - \frac{1}{2} k_2)/\sqrt{n} + O(n^{-1}), \quad \text{var}(T) = 1 + O(n^{-1}),
\]
\[
\text{skew}(T) = -(3k_2 - k_3)/\sqrt{n} + O(n^{-1})
\]
where
\[
\frac{k_2}{\sqrt{n}} = E\left\{ \frac{(\hat{\sigma}^2 - \sigma^2)(\hat{\theta} - \theta)}{\sigma^3} \right\}.
\]
Observe that \( k_2 \) is of order \( O(1) \), since \( \sigma^2 \) is of order \( O(n^{-1}) \) and \( \hat{\sigma}^2 \) generally differs from \( \sigma^2 \) by order \( O_p(n^{-3/2}) \). The fourth and higher-order cumulants of \( T \) are of order \( O(n^{-1}) \) or smaller. Thus, when \( \theta \) is continuous, the cumulative distribution functions \( H(u) \) and \( K(t) \) of \( U \) and \( T \) typically have Cornish-Fisher expansions
\[
H(u) = \text{pr}\{(\hat{\theta} - \theta)/\sigma \leq u\} = \Phi\left[u - n^{-1/2}\{(k_1 - \frac{1}{2} k_3) + \frac{1}{6} k_3 u^2\}\right] + O(n^{-1}),
\]
\[
K(t) = \text{pr}\{(\hat{\theta} - \theta)/\hat{\sigma} \leq t\} = \Phi\left[t - n^{-1/2}\{(k_1 - \frac{1}{6} k_3) - (\frac{1}{2} k_2 - \frac{1}{6} k_3)t^2\}\right] + O(n^{-1}).
\]
Furthermore, the inverse cumulative distribution functions $H^{-1}(\alpha)$ and $K^{-1}(\alpha)$ have expansions

\begin{align*}
H^{-1}(\alpha) &= z^{(\alpha)} + n^{-1/2} \left[ (k_1 - \frac{1}{8} k_3) + \frac{1}{8} k_3 \{z^{(\alpha)}\}^2 \right] + O(n^{-1}), \quad (8.3) \\
K^{-1}(\alpha) &= z^{(\alpha)} + n^{-1/2} \left[ (k_1 - \frac{1}{8} k_3) - (\frac{1}{2} k_2 - \frac{1}{8} k_3) \{z^{(\alpha)}\}^2 \right] + O(n^{-1}). \quad (8.4)
\end{align*}

To compare approximate confidence limits, Hall (1988) defined an "exact" upper $\alpha$ confidence limit for $\theta$ as $\hat{\theta}_{\text{exact}}(\alpha) = \hat{\theta} - \hat{\sigma} K^{-1}(1 - \alpha)$. This limit is exact in the sense of coverage; note that \( \text{pr} \{ K^{-1}(1 - \alpha) \leq (\hat{\theta} - \theta) / \hat{\sigma} \} = \alpha \) implies \( \text{pr} \{ \theta \leq \hat{\theta}_{\text{exact}}(\alpha) \} = 1 - \alpha \). It requires the cumulative distribution function $K$, which is rarely known in practice; however, although usually unavailable, $\hat{\theta}_{\text{exact}}(\alpha)$ does provide a useful benchmark for making comparisons. By using (8.4), the exact limit is seen to satisfy

\begin{equation}
\hat{\theta}_{\text{exact}}(\alpha) = \hat{\theta} + \hat{\sigma} z^{(\alpha)} - n^{-1/2} \hat{\sigma} \left[ (k_1 - \frac{1}{8} k_3) - (\frac{1}{2} k_2 - \frac{1}{8} k_3) \{z^{(\alpha)}\}^2 \right] + O_p(n^{-3/2}). \quad (8.5)
\end{equation}

An approximate $\alpha$ confidence limit $\hat{\theta}(\alpha)$ is said to be second-order correct if it differs from $\hat{\theta}_{\text{exact}}(\alpha)$ by order $O_p(n^{-3/2})$, that is, if it equals the right-hand side of (8.5). It is easily seen from (8.2) that a second-order correct limit $\hat{\theta}(\alpha)$ is also second-order accurate, that is, \( \text{pr} \{ \theta \leq \hat{\theta}(\alpha) \} = \alpha + O(n^{-1}) \).

Let $\hat{K}(t)$ be the bootstrap cumulative distribution function of $T$, so that $\hat{K}(t)$ is the cumulative distribution function of $T^* = (\hat{\theta}^* - \hat{\theta}) / \hat{\sigma}^*$. The first three cumulants of $T^*$ typically differ from those of $T$ by order $O_p(n^{-1})$, and $\hat{K}(t)$ has the expansion

\begin{equation}
\hat{K}(t) = \Phi \left[ t - n^{-1/2} \left\{ (\hat{k}_1 - \frac{1}{8} \hat{k}_3) - (\frac{1}{2} \hat{k}_2 - \frac{1}{8} \hat{k}_3) t^2 \right\} \right] + O_p(n^{-1}),
\end{equation}

where $\hat{k}_j = k_j + O_p(n^{-1/2})$. Hence, $\hat{K}(t) = K(t) + O_p(n^{-1})$ and $\hat{K}^{-1}(\alpha) = K^{-1}(\alpha) + O_p(n^{-1})$, and since $\hat{\sigma}$ is of order $O_p(n^{-1/2})$, the bootstrap-$t$ confidence limit $\hat{\theta}_T(\alpha)$ satisfies

\begin{equation}
\hat{\theta}_T(\alpha) = \hat{\theta} - \hat{\sigma} K^{-1}(1 - \alpha) = \hat{\theta} - \hat{\sigma} K^{-1}(1 - \alpha) + O_p(n^{-3/2}) = \hat{\theta}_{\text{exact}}(\alpha) + O_p(n^{-3/2}). \quad (8.6)
\end{equation}

Expression (8.6) shows that the bootstrap-$t$ method is second-order correct.

To demonstrate the second-order correctness of the $BC_a$ method, let $\hat{H}(u)$ be the cumulative bootstrap distribution function of $U$, so that $\hat{H}(u)$ is the cumulative distribution function of $U^* = (\hat{\theta}^* - \hat{\theta}) / \hat{\sigma}$. It is assumed that the estimator $\hat{\sigma}^2$ is such that the bootstrap distribution of $\hat{\theta}$ has variance that differs from $\sigma^2$ by order $O_p(n^{-2})$, that is, \( \text{var}(\hat{\sigma}) = \hat{\sigma}^2 + O_p(n^{-2}) \). The first three cumulants of $U^*$ typically differ from those of $U$ by order $O_p(n^{-1})$, so $\hat{H}(u) = H(u) + O_p(n^{-1})$ and $\hat{H}^{-1}(\alpha) = H^{-1}(\alpha) + O_p(n^{-1})$. The bootstrap cumulative distribution function $\hat{G}(c)$ of $\hat{\theta}$ satisfies $\hat{G}(c) = \hat{H} \{ (c - \hat{\theta}) / \hat{\sigma} \}$, and $\hat{G}^{-1}(\alpha) = \hat{\theta} + \hat{\sigma} \hat{H}^{-1}(\alpha)$. Thus, (8.3) gives

\begin{equation}
\hat{G}^{-1}(\alpha) = \hat{\theta} + \hat{\sigma} z^{(\alpha)} + n^{-1/2} \hat{\sigma} \left[ (k_1 - \frac{1}{8} k_3) + \frac{1}{8} k_3 \{z^{(\alpha)}\}^2 \right] + O_p(n^{-3/2}),
\end{equation}

28
and by definition (2.3),
\[
\hat{\theta}_{BC_a}[\alpha] = \hat{G}^{-1} \left[ \Phi \left\{ z_0 + \frac{z_0 + z^{(a)}}{1 - a(z_0 + z^{(a)})} \right\} \right] \\
= \hat{G}^{-1} \left\{ \Phi \left( z^{(a)} + 2z_0 + a\{z^{(a)}\}^2 \right) \right\} + \mathcal{O}_p(n^{-3/2}) \\
= \hat{\theta} + \hat{\sigma}z^{(a)} + n^{-1/2} \hat{\sigma} \left[ 2\sqrt{n}z_0 + (k_1 - \frac{1}{6}k_3) + (\sqrt{n}a + \frac{1}{6}k_3)\{z^{(a)}\}^2 \right] + \mathcal{O}_p(n^{-3/2}). \tag{8.7}
\]

Comparison of (8.5) and (8.7) shows that \( \hat{\theta}_{BC_a}[\alpha] \) is second-order correct when \( a \) and \( z_0 \) are defined by
\[
a = \left( \frac{1}{2}k_2 - \frac{1}{3}k_3 \right)/\sqrt{n}, \tag{8.8}
\]
\[
z_0 = -(k_1 - \frac{1}{6}k_3)/\sqrt{n}. \tag{8.9}
\]
The quantities \( a \) and \( z_0 \) are of order \( O(n^{-1/2}) \). The quantity \( a \) satisfies
\[
a = -\frac{1}{6}\{\text{skew}(U) + \text{skew}(T)\} + O(n^{-1}),
\]
and interpretation (2.7) for \( z_0 \) is easily seen from (8.1), for
\[
\Phi(z_0) = \Phi\{-(k_1 - \frac{1}{6}k_3)/\sqrt{n}\} = H(0) + O(n^{-1}) = \text{pr}\{\hat{\theta} \leq \theta\} + O(n^{-1}).
\]
In practice, \( \hat{\theta}_{BC_a}[\alpha] \) is calculated using estimates \( \hat{a} \) and \( \hat{z}_0 \) that differ from \( a \) and \( z_0 \) by order \( O_p(n^{-1}) \); expression (8.7) shows that this change does not affect the second-order correctness of \( \hat{\theta}_{BC_a}[\alpha] \). The estimate \( \hat{z}_0 \) given in expression (2.8) has this property, since
\[
\hat{z}_0 = \Phi^{-1}\{\hat{G}(\hat{\theta})\} = \Phi^{-1}\{\hat{H}(0)\} = \Phi^{-1}\{H(0)\} + \mathcal{O}_p(n^{-1}) \\
= \Phi^{-1}\{\Phi\{-(k_1 - \frac{1}{6}k_3)/\sqrt{n}\}\} + \mathcal{O}_p(n^{-1}) = z_0 + \mathcal{O}_p(n^{-1}).
\]
The second-order correctness of the bootstrap-\( t \) and the \( BC_a \) methods has been discussed by Efron (1987), Bickel (1987, 1988), Hall (1988), and DiCiccio and Romano (1995).
Definitions (8.8) and (8.9) for \( a \) and \( z_0 \) can be used to cast expansion (8.5) for \( \hat{\theta}_{exact}[\alpha] \) into the form of (4.20). In particular,
\[
\hat{\theta}_{exact}[\alpha] = \hat{\theta} + \hat{\sigma}z^{(a)} + \hat{\sigma}z_0 + (2a + c_q)\{z^{(a)}\}^2 + \mathcal{O}_p(n^{-3/2}), \tag{8.10}
\]
where
\[
c_q = -(\frac{1}{2}k_2 - \frac{1}{3}k_3)/\sqrt{n}. \tag{8.11}
\]
The bias of \( \hat{\theta} \) is
\[
b = \sigma k_1/\sqrt{n}, \tag{8.12}
\]
29
and \( z_0 \) can be expressed in terms of \( a, c_q, \) and \( b \) by

\[
z_0 = a + c_q - b/\sigma = \Phi^{-1}(2\Phi(a)\Phi(c_q - b/\sigma)) + O(n^{-1}).
\]  

(8.13)

If \( \hat{c}_q \) and \( \hat{b} \) are estimates that differ from \( c_q \) and \( b \) by order \( O_p(n^{-1}) \), then estimate (4.12),

\[
\hat{z}_0 = \Phi^{-1}(2\Phi(\hat{a})\Phi(\hat{c}_q - \hat{b}/\hat{\sigma})),
\]

(8.14)

differs from \( z_0 \) by the same order.

Once estimates \((\hat{\theta}, \hat{\sigma}, \hat{\alpha}, \hat{z}_0, \hat{c}_q)\) are obtained, the quadratic version of the ABC confidence limit, \( \hat{\theta}_{ABC}[\alpha] = \hat{\theta} + \hat{\sigma}\hat{\xi} \), can be constructed according to definition (4.13). This limit is second-order correct. Since

\[
w = \hat{z}_0 + z^{(\alpha)} = z_0 + z^{(\alpha)} + O_p(n^{-1}),
\]

\[
\lambda = w(1 - \hat{a}w)^{-2} = z^{(\alpha)} + z_0 + 2a\{z^{(\alpha)}\}^2 + O_p(n^{-1}),
\]

\[
\xi = \lambda + \hat{c}_q \lambda^2 = z^{(\alpha)} + z_0 + (2a + c_q)\{z^{(\alpha)}\}^2 + O_p(n^{-1}),
\]

(8.15)

\( \hat{\theta}_{ABC}[\alpha] \) agrees with (8.10) to error of order \( O_p(n^{-3/2}) \).

In many contexts, there exists a vector of parameters \( \zeta = (\zeta_1, \ldots, \zeta_p)' \) and an estimator \( \hat{\zeta} = (\hat{\zeta}_1, \ldots, \hat{\zeta}_p)' \) such that the parameter of interest is \( \theta = t(\zeta) \), and the variance of the estimator \( \hat{\theta} = t(\hat{\zeta}) \) is of the form \( \sigma^2 = v(\zeta) + O(n^{-2}) \), so the variance is estimated by \( \hat{\sigma}^2 = v(\hat{\zeta}) \). This situation arises in parametric models and in the smooth function of means model. For the smooth function model, inference is based on independent and identically distributed vectors \( x_1, \ldots, x_n \), each having mean \( \mu \); the parameter of interest is \( \theta = t(\mu) \), which is estimated by \( \hat{\theta} = t(\hat{\mu}) \). In fact the smooth function model is closely related to exponential families, as shown in Section 4 of DiCiccio and Efron (1992).

Assume that \( \sqrt{n}(\hat{\zeta} - \zeta) \) is normally distributed asymptotically. Typically, the first three joint cumulants of \( \hat{\zeta}_1, \ldots, \hat{\zeta}_p \) are

\[
E(\hat{\zeta}_i) = \zeta_i + \kappa_i, \quad \text{cov}(\hat{\zeta}_i, \hat{\zeta}_j) = \kappa_{i,j}, \quad \text{cum}(\hat{\zeta}_i, \hat{\zeta}_j, \hat{\zeta}_k) = \kappa_{i,j,k}, \quad (i, j, k = 1, \ldots, p),
\]

where \( \kappa_i \) and \( \kappa_{i,j} \) are of order \( O(n^{-1}) \) and \( \kappa_{i,j,k} \) is of order \( O(n^{-2}) \), and the fourth- and higher-order joint cumulants are of order \( O(n^{-3}) \) or smaller. Straightforward calculations show that

\[
\sigma^2 = \kappa_{i,j}t_it_j + O(n^{-2}), \quad \text{where} \quad t_i = \partial t(\zeta)/\partial \zeta_i \quad (i = 1, \ldots, p).
\]

In this expression and subsequently, the usual convention is used whereby summation over repeated indices is understood, with the range of summation being \( 1, \ldots, p \). Now, suppose \( \zeta \) is sufficiently rich so that \( \kappa_{i,j} \) depends on the underlying distribution only through \( \zeta \) for indices \( i \) and \( j \) such that \( t_i \) and \( t_j \) are nonvanishing.

Then, it is possible to write

\[
v(\zeta) = \kappa_{i,j}(\zeta)t_i(\zeta)t_j(\zeta) + O(n^{-2}),
\]

30
and
\[ \hat{\sigma}^2 = n(\hat{\zeta}) = \kappa_{i,j}(\hat{\zeta})t_i(\hat{\zeta})t_j(\hat{\zeta}) + O_p(n^{-2}). \]

In this case, the quantities \(k_1, k_2, k_3\) are given by
\[
\begin{align*}
k_1 &= \sqrt{n}(\kappa_{i,i} + \frac{1}{2}\kappa_{i,j}t_{ij})/(\kappa_{i,j}t_i t_j)^{1/2}, \\
k_2 &= \sqrt{n}\kappa_{i,j}u_i t_j / (\kappa_{i,j}t_i t_j)^{3/2} \\
&= \sqrt{n}(\kappa_{i,j}k_{i,j}t_i t_j t_k + 2\kappa_{i,j}\kappa_{k,l}t_i t_k t_j j)/(\kappa_{i,j}t_i t_j)^{3/2}, \\
k_3 &= \sqrt{n}(\kappa_{i,j}k_{i,j}k_{k,l}t_i t_j t_k + 3\kappa_{i,j}\kappa_{k,l}t_i t_k t_j j)/(\kappa_{i,j}t_i t_j)^{3/2},
\end{align*}
\]

(8.16)

to error of order \(O(n^{-1/2})\), where \(t_{ij} = \partial^2 t(\zeta)/\partial \zeta_i \partial \zeta_j, v_i = \partial v(\zeta)/\partial \zeta_i, \kappa_{i,j/k} = \partial \kappa_{i,j}(\zeta)/\partial \zeta_k,\)

\((i, j, k = 1, \ldots, p). \) It follows from (8.8), (8.11), and (8.12) that
\[
\begin{align*}
a &= \left(\frac{1}{2}\kappa_{i,j/k} - \frac{2}{3}\kappa_{i,j,k}\right)t_i t_j t_k / (\kappa_{i,j}t_i t_j)^{3/2}, \\
b &= \kappa_{i,i} + \kappa_{i,j}t_{ij}, \\
c_q &= -\left(\frac{1}{2}\kappa_{i,j/k} - \frac{2}{3}\kappa_{i,j,k}\right)t_i t_j t_k / (\kappa_{i,j}t_i t_j)^{3/2} + \frac{1}{2}\kappa_{i,j}k_{i,j}t_i t_k t_j / (\kappa_{i,j}t_i t_j)^{3/2},
\end{align*}
\]

(8.17)
to error of order \(O(n^{-1})\). An expression for \(z_0\) having error of order \(O(n^{-1})\) can be deduced from

(8.17) by using (8.13).

The ABC method applies to both exponential families and the smooth function of means
to error of order \(O(n^{-1})\). An expression for \(z_0\) having error of order \(O(n^{-1})\) can be deduced from

(8.17) by using (8.13).

The ABC method applies to both exponential families and the smooth function of means
model. For these cases, \(\hat{\zeta}\) is an unbiased estimate of \(\zeta\), and the cumulant generating function of \(\hat{\zeta}\),
\(\Psi(\xi) = \log E\{\exp(\xi \hat{\zeta})\},\) has an approximation \(\hat{\Psi}(\xi)\) such that
\[
\frac{\partial \hat{\Psi}(\xi)}{\partial \xi_i} \bigg|_{\xi = 0} = \hat{\zeta}_i, \\
\frac{\partial^2 \hat{\Psi}(\xi)}{\partial \xi_i \partial \xi_j} \bigg|_{\xi = 0} = \kappa_{i,j}(\hat{\zeta}) + O_p(n^{-2}), \\
\frac{\partial^3 \hat{\Psi}(\xi)}{\partial \xi_i \partial \xi_j \partial \xi_k} \bigg|_{\xi = 0} = \kappa_{i,j,k} + O_p(n^{-s/2}).
\]

In particular, it is reasonable to take \(\hat{\sigma}^2 = \hat{\Psi}_{ij} \hat{t}_i \hat{t}_j\), where \(\hat{t}_i = t_i(\hat{\zeta})\) \((i = 1, \ldots, p)\). The ABC
algorithm uses numerical differentiation of \(t(\zeta)\) and \(\hat{\Psi}(\xi)\) to facilitate calculation of estimates
\(\hat{\sigma}, \hat{a}, \hat{z}_0, \hat{c}_q\).

In exponential families, the distribution of an observed random vector \(y = (y_1, \ldots, y_p)'\) is
indexed by an unknown parameter \(\eta = (\eta_1, \ldots, \eta_p)'\), and the log likelihood function for \(\eta\) based on
\(\Psi(\eta) = n\{\eta_1 \psi_1(\eta) - \psi(\eta)\}, \) where \(E(\eta) = \psi(\eta), \) and both \(\Psi(\eta)\) are
of order \(O(1)\). In this case, \(y\) plays the role of \(\hat{\zeta}\), and \(\zeta\) corresponds to the expectation parameter
\(\mu = E(y) = \partial \psi(\eta)/\partial \eta. \) Upon defining \(\eta = \psi(\eta)\) and \(\psi(\eta) = n \psi(\eta) = n \psi(\eta)/n\), the log
likelihood function for \(\eta\) based on \(y\) is \(l(\eta; y) = \eta' y - \psi(\eta)\), which agrees with (3.1). The cumulant

31
generating function for $y$ is $\Psi(\xi) = \psi(\eta + \xi) - \psi(\eta)$, and the approximate cumulant generating function is

$$\hat{\Psi}(\xi) = \psi(\hat{\eta} + \xi) - \psi(\hat{\eta}),$$

where $\hat{\eta}$ is the maximum likelihood estimator obtained from the equations $\psi_i(\hat{\eta}) = y_i\; (i = 1, \ldots, p)$. The usual estimate of variance is $\hat{\sigma}^2 = \psi_{ij}(\hat{\eta})\hat{t}_i\hat{t}_j = \hat{\Psi}_{ij}\hat{t}_i\hat{t}_j$.

In the smooth function model, the cumulant generating function is approximated by

$$\hat{\Psi}(\xi) = n \log \left\{ \frac{1}{n} \sum_{j=1}^{n} \exp \left( \xi_i x_{ij}/n \right) \right\},$$

which is the true cumulant generating function for the model that puts probability mass $1/n$ on each of the observed random vectors $x_j = (x_{1j}, \ldots, x_{pj})'$ ($j = 1, \ldots, n$). The usual estimate of variance obtained from the delta-method is

$$\hat{\sigma}^2 = \frac{1}{n^2} \left\{ \sum_{j=1}^{n} (x_{ik} - \overline{x}_i)(x_{jk} - \overline{x}_j) \right\} \hat{t}_i\hat{t}_j = \hat{\Psi}_{ij}\hat{t}_i\hat{t}_j,$$

where $\overline{x}_i = \sum x_{ij}/n$.

Key features of exponential families and the smooth function model are that $\kappa_i = 0$ and $\kappa_{ij/l}\kappa_{k,l} = \kappa_{i,j,k} (i, j, k = 1, \ldots, p)$, so the expressions for $a$, $b$, and $c$ given in (5.17) undergo considerable simplification; in particular,

$$a = \frac{1}{2}\kappa_{i,j,k}\kappa_{k,l}t_{ij}\kappa_{i,j}t_{ij}/(\kappa_{i,j}\kappa_{i,j}t_{ij})^{3/2}, \quad b = \frac{1}{2}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}\kappa_{i,j}t_{ij}/(\kappa_{i,j}\kappa_{i,j})^{3/2},$$

to error of order $O(n^{-1})$.

The ABC method requires only that $t(\xi)$ and $\hat{\Psi}_i(\xi)$ be specified; the estimates $\hat{\sigma}$, $\hat{a}$, $\hat{\sigma}_0$, and $\hat{c}_q$ are obtained by numerical differentiation. The details are as follows. By definition,

$$\hat{t}_i = \left. \frac{d}{d\epsilon} t(\xi + \epsilon e_i) \right|_{\epsilon=0} \quad (i = 1, \ldots, p),$$

$$\hat{\Psi}_{ij} = \left. \frac{d}{d\epsilon} \hat{\Psi}_i(\epsilon e_j) \right|_{\epsilon=0} \quad (i, j = 1, \ldots, p),$$

where $e_i$ is the $p$-dimensional unit vector whose $i$th entry is 1. Let $\hat{t} = (\hat{t}_1, \ldots, \hat{t}_p)'$, $\hat{\Psi} = (\hat{\Psi}_{ij})$, $\hat{\sigma}^2 = \hat{\Psi}_{ij}\hat{t}_i\hat{t}_j = t'\hat{\Psi} t$. Then

$$\hat{a} = \left. \frac{\hat{\Psi}_{ijk}\hat{t}_i\hat{t}_j\hat{t}_k}{6\hat{\sigma}^3} \right|_{\epsilon=0} = \left. \frac{1}{6\hat{\sigma}^3} \frac{d^2}{d\epsilon^2} \hat{\Psi}_i(\epsilon\hat{t}) \right|_{\epsilon=0},$$

32
\[ \hat{c}_q = \frac{\hat{\Psi}_{ij} \hat{\Psi}_{kl} \hat{t}_{ij} \hat{t}_{kl}}{2\hat{q}^3} = \left. \frac{1}{2\hat{q}} \frac{d^2}{d\hat{q}^2} t \left( \hat{c} + \epsilon \hat{\hat{q}} \right) \right|_{\epsilon=0}. \]

Now \( \hat{\hat{q}} = \Gamma D \Gamma' \), where \( D \) is a diagonal matrix of eigenvalues of \( \hat{\hat{q}} \) and \( \Gamma \) is an orthogonal matrix whose columns are corresponding eigenvectors. Denote the \( i \)th diagonal element of \( D \) by \( d_i \) and the \( i \)th column of \( \Gamma \) by \( \gamma_i = (\gamma_{1i}, \ldots, \gamma_{pi})' \), so that \( \hat{\Psi}_{ij} = \sum_k d_k \gamma_{ik} \gamma_{jk} \). The quantity \( \hat{b} \) can be estimated by

\[ \hat{b} = \frac{\hat{\Psi}_{ij} \hat{t}_{ij}}{2} = \left. \frac{1}{2} \sum_{i=1}^p \frac{d^2}{d\hat{q}^2} t \left( \hat{c} + \epsilon d_i^{1/2} \gamma_i \right) \right|_{\epsilon=0}. \]

If calculating the eigenvalues and eigenvectors is too cumbersome, then \( \hat{b} \) can be obtained from

\[ \hat{b} = \left. \frac{1}{2} \sum_{i=1}^p \frac{\partial^2}{\partial \epsilon_1 \partial \epsilon_2} t \left( \hat{c} + \epsilon_1 e_i + \epsilon_2 \hat{\hat{q}} e_i \right) \right|_{(\epsilon_1, \epsilon_2)=(0,0)}. \]

Once \( \hat{\hat{q}}^2, \hat{a}, \hat{b}, \) and \( \hat{c} \) are calculated, then \( \hat{z}_0 \) can be obtained using (8.14).

The ABC confidence limit \( \hat{\theta}_{ABC}[\alpha] \) is defined in (8.14) as

\[ \hat{\theta}_{ABC}[\alpha] = t \left( \hat{c} + \frac{\lambda \hat{\hat{q}}}{\hat{\hat{q}}} \right). \]

This confidence limit is second-order correct; by (5.10) and (5.15),

\[ \hat{\theta}_{ABC}[\alpha] = \hat{\theta} + \lambda \frac{\hat{t}_{ij} \hat{\Psi}_{ij}}{\hat{\hat{q}}} + \frac{\lambda^2 \hat{t}_{ij} \hat{\Psi}_{kl} \hat{\Psi}_{ij} \hat{t}_{kl}}{2\hat{\hat{q}}^2} + O_p(n^{-3/2}) \]

\[ = \hat{\theta} + \hat{\sigma} \lambda + \hat{\sigma} \hat{c}_q \lambda^2 + O_p(n^{-3/2}) \]

\[ = \hat{\theta} + \hat{\sigma} \left[ \hat{z}(\alpha) + z_0 + 2a \{z(\alpha)\}^2 \right] + \hat{\sigma} \hat{c}_q \{z(\alpha)\}^2 + O_p(n^{-3/2}) \]

\[ = \hat{\theta}_{exact}[\alpha] + O_p(n^{-3/2}). \]

The second-order correctness of the ABC method for exponential families was shown by DiCicco and Efron (1992).

9. Parametric Models and Conditional Confidence Intervals. An impressive likelihood-based theory of higher-order accurate confidence intervals has been developed during the past decade. This effort has involved many authors, including Barndorff-Nielsen (1986), Cox and Reid (1987), Pierce and Peters (1992), and McCullagh and Tibshirani (1990). This section concerns the connection of bootstrap confidence intervals with the likelihood-based theory. We will see that in exponential families, including nonparametric situations, the bootstrap can be thought of as an easy, automatic way of constructing the likelihood intervals. However in parametric families that are not exponential, the two theories diverge. There the likelihood intervals are second-order accurate in a conditional sense, while the bootstrap intervals' accuracy is only unconditional. In
order to get good conditional properties the bootstrap resampling would have to be done according to the appropriate conditional distribution, which would usually be difficult to implement.

Consider an observed random vector \( y = (y_1, \ldots, y_n)' \) whose distribution depends on an unknown parameter \( \zeta = (\zeta_1, \ldots, \zeta_p)' \), and let \( l(\zeta) = l(\zeta; y) \) be the log likelihood function for \( \zeta \) based on \( y \). Suppose the parameter \( \theta = t(\zeta) \) is estimated by \( \hat{\theta} = t(\hat{\zeta}) \), where \( \hat{\zeta} = (\hat{\zeta}_1, \ldots, \hat{\zeta}_p)' \) is the maximum likelihood estimator. Parametric bootstrap distributions are generally constructed using samples \( y^* \) drawn from the fitted distribution for \( y \), that is, from the distribution having \( \zeta = \hat{\zeta} \).

Asymptotic formulae for the first three cumulants of \( \hat{\zeta} \) are given by McCullagh (1987, chapter 7), and using these formulae in conjunction with (8.16) shows that \( \sigma^2 = \lambda_{ii,tt} + O(n^{-2}) \) and

\[
\begin{align*}
  k_1 &= -\sqrt{n} \left[ (\frac{1}{2} \lambda_{ii,j,k} + \frac{1}{2} \lambda_{ij,k}) \lambda_{ii,j} \lambda_{ik,t} t_i \lambda_{ij,t} t_j / (\lambda_{ii,j} t_i t_j)^{1/2}, \\
  k_2 &= -\sqrt{n} \left[ (\lambda_{ii,j,k} + 2 \lambda_{ij,k}) \lambda_{ii,j} \lambda_{ii,m} \lambda_{i,k} \lambda_{i,m} t_i t_m t_k t_m / (\lambda_{ii,j} t_i t_j)^{3/2}, \\
  k_3 &= -\sqrt{n} \left[ (2 \lambda_{ii,j,k} + 3 \lambda_{ij,k}) \lambda_{ii,j} \lambda_{ii,m} \lambda_{i,k} \lambda_{i,m} t_i t_m t_k t_m / (\lambda_{ii,j} t_i t_j)^{3/2},
\end{align*}
\]  

(9.1)

to error of order \( O(n^{-1/2}) \), where \( \lambda_{ii,j} = E(l_i l_j), \lambda_{ii,k} = E(l_i l_j l_k), \lambda_{i,j,k} = E(l_i l_j l_k), \) with \( l_i = \partial l(\zeta)/\partial \xi_i, l_{ij} = \partial^2 l(\zeta)/\partial \xi_i \partial \xi_j, \) and \( (\lambda_{ii,j}) \) is the \( p \times p \) matrix inverse of \( (\lambda_{ii,j}) \). The quantities \( \lambda_{ii,j}, \lambda_{ii,k}, \) and \( \lambda_{i,j,k} \) are assumed to be of order \( O(n) \). The expected information estimate of variance is \( \sigma^2 = \hat{\lambda}_{ii,j} t_i t_j \), where \( \hat{\lambda}_{ii,j} = \lambda_{ii,j}(\hat{\zeta}) \), and the variance of the bootstrap distribution of \( \hat{\theta} \) satisfies \( \text{var}(\hat{\theta}^*) = \hat{\sigma}^2 + O_p(n^{-2}) \). Thus, if the studentized statistic is defined using the expected information estimate of variance, say \( T_E = (\hat{\theta} - \theta)/\hat{\sigma} \), then the results of Section 5 show that the \( BCA \) method is second-order correct with respect to \( T_E \). Using (8.8) in conjunction with (9.1) to calculate \( a \) yields

\[
a = \frac{1}{\hat{\sigma}} \lambda_{ii,j,k} \lambda_{ii,m} \lambda_{i,k} \lambda_{i,m} t_i t_m t_k t_m / (\lambda_{ii,j} t_i t_j)^{3/2},
\]  

(9.2)

to error of order \( O(n^{-1}) \). This formula for \( a \) was given by Efron (1987).

If nuisance parameters are absent \( (p = 1) \) and \( \theta = \zeta \), then (8.9), (9.1), and (9.2) show that

\[
a = z_0 = \frac{1}{\hat{\sigma}} \lambda_{1,1,1} \lambda_{1,1}^{-3/2} = \frac{1}{\hat{\sigma}} \text{skew}(\partial l(\theta)/\partial \theta),
\]  

(9.3)

to error of order \( O(n^{-1}) \). The equality of \( z_0 \) and \( a \) in this context was demonstrated by Efron (1987).

In addition to being invariant under monotonically increasing transformations of the parameter of interest as described in section 3, the quantities \( a \) and \( z_0 \) are also invariant under reparameterizations \( \eta = \eta(\zeta) \) of the model. Expression (9.2) for \( a \) is invariant under reparameterizations of the model, as is the formula for \( z_0 \) obtained by substituting (9.1) into (8.9). There is no restriction then in assuming the model is parameterized so that \( \theta = \zeta^1 \) and the nuisance parameters \( \zeta^2, \ldots, \zeta^p \) are orthogonal to \( \theta \). Here, orthogonality means \( \lambda_{1,a} = \lambda_{1,a}^1 = 0 \) \( (a = 2, \ldots, p) \); see Cox and Reid (1987). In this case, (6.2) becomes

\[
a = \frac{1}{\hat{\sigma}} \lambda_{1,1,1} \lambda_{1,1}^{-3/2} = \frac{1}{\hat{\sigma}} \text{skew}(\partial l(\zeta)/\partial \zeta^1).
\]  

(9.4)
Comparison of (9.4) with (9.3) indicates that, to error of order $O(n^{-1})$, $a$ coincides with its version that would apply if the orthogonal nuisance parameters were known. In this sense, $a$ can be regarded as unaffected by the presence of nuisance parameters. In contrast, for the orthogonal case,

$$z_0 = (\frac{1}{2}\lambda_{a, b, 1} + \frac{1}{2}\lambda_{a, a, 1})\lambda_{a, b, 1, 1}^{-1/2} + \frac{1}{6}\lambda_{a, b, 1, 1, 1}\lambda_{a, b, 1, 1}^{-3/2}, \quad (9.5)$$

to error of order $O(n^{-1})$, where, for purpose of the summation convention, the indices $a$ and $b$ range over $2, \ldots, p$. Expression (9.5) shows that $z_0$ reflects the presence of unknown nuisance parameters.

Another possibility for studentizing is to use the observed information estimate of variance, $\bar{\sigma}^2 = \bar{L}_{ij} i_j$, where $\bar{L}_{ij}$ is the $p \times p$ matrix inverse of $(\bar{L}_{ij})$ and $\bar{i}_j = i_j(\bar{z})$. Let $T_0 = (\hat{\theta} - \theta)/\bar{\sigma}$. Using the bootstrap-$t$ method with $T_0$ and $T_E$ produces approximate confidence limits $\hat{T}_{T_0}[\alpha]$ and $\hat{T}_{T_E}[\alpha]$ which both have coverage error of order $O(n^{-1})$. However, $\bar{\sigma} = \hat{\sigma} + O_p(n^{-1})$, so $T_0 = T_E + O_p(n^{-1/2})$, and $\hat{T}_{T_0}[\alpha]$ and $\hat{T}_{T_E}[\alpha]$ typically differ by order $O_p(n^{-1})$. The studentized quantities $T_E$ and $T_O$ produce different definitions of second-order correctness. In particular, $\hat{T}_{BC_a}[\alpha]$ differs from $\hat{T}_{T_0}[\alpha]$ by order $O_p(n^{-1})$, and the $BC_a$ method, which is second-order correct with respect to $T_E$, fails to be second-order correct with respect to $T_O$. For exponential families, $\bar{\sigma}^2 = i^2$ since $\lambda_{i,j} = -i^2$, and no distinction arises between $T_E$ and $T_O$ in the definition of second-order correctness.

Although $T_E$ and $T_O$ generally differ by order $O_p(n^{-1/2})$, their first three cumulants agree to error of order $O(n^{-1})$. It follows then from (5.5) that $\hat{T}_{T_0}[\alpha]$ and $\hat{T}_{T_E}[\alpha]$ have expansions

$$\hat{T}_{T_0}[\alpha] = \hat{\theta} + \bar{\sigma} z^{(\alpha)} - n^{-1/2} \bar{\sigma} [[[k_1 - \frac{1}{6} k_3] - (\frac{1}{2} k_2 - \frac{1}{6} k_3)\{z^{(\alpha)}\}^2] + O_p(n^{-3/2}),$$

$$\hat{T}_{T_E}[\alpha] = \hat{\theta} + \bar{\sigma} z^{(\alpha)} - n^{-1/2} \bar{\sigma} [[[k_1 - \frac{1}{6} k_3] - (\frac{1}{2} k_2 - \frac{1}{6} k_3)\{z^{(\alpha)}\}^2] + O_p(n^{-3/2}), \quad (9.6)$$

where $k_1, k_2, k_3$ are given by (9.1). Expression (9.6) shows that if $\hat{T}_{E}[\alpha]$ is a second-order correct confidence limit with respect to $T_E$, such as $\hat{T}_{BC_a}[\alpha]$, then

$$\hat{T}_O[\alpha] = \hat{\theta} + \frac{\bar{\sigma}}{\bar{\sigma}}(\hat{T}_{E}[\alpha] - \hat{\theta})$$

is second-order correct with respect to $T_O$.

Confidence limits that are second-order correct with respect to $T_O$ agree closely with second-order accurate confidence limits obtained from likelihood ratio statistics. The profile log likelihood function for $\theta$ is $l_p(\theta) = l(\hat{\zeta})$, where $\hat{\zeta}$ is the constrained maximum likelihood estimator of $\zeta$ given $\theta$; that is, $\hat{\zeta}$ maximizes $l(\zeta)$ subject to the constraint $t(\zeta) = \theta$. Since $\hat{\zeta}$ is the global maximum likelihood estimator $\hat{\zeta}$, $l_p(\theta)$ is maximized at $\hat{\theta}$. The likelihood ratio statistic for $\theta$ is

$$W_p(\theta) = 2\{l(\hat{\zeta}) - l(\hat{\zeta})\} = 2\{l_p(\hat{\theta}) - l_p(\theta)\},$$

and the signed root of the likelihood ratio statistic is

$$R_p(\theta) = \text{sgn}(\hat{\theta} - \theta)\sqrt{W_p(\theta)}.$$
In wide generality, \( W_p(\theta) \) and \( R_p(\theta) \) are asymptotically distributed as \( \chi^2_1 \) and \( N(0,1) \), respectively. Straightforward calculations show that the derivatives of \( l_p(\theta) \) satisfy
\[
\frac{L_p^{(3)}(\hat{\theta})}{\sigma^6} = \frac{-\tilde{l}_{ijk} l_{ij}^m l_{ik}^n}{\sigma^6}
\]
\[
= (\lambda_{ijk} \lambda_{iij}^m \lambda_{ik}^n + 3\lambda_{ij}^m \lambda_{iik}^n + 3\lambda_{ij}^m \lambda_{iik}^n) / \sigma^6 + O_p(n^{1/2})
\]
\[
= n^{-1/2} (3k_2 - k_3) / \sigma^3 + O_p(n^{1/2})
\]
\[
= (2a + c_q) / \sigma^3 + O_p(n^{1/2});
\]

these calculations make use of the Bartlett identities \( \lambda_{ij} = E(l_{ij}) = -\lambda_{ij} \) and
\[
\lambda_{ijk} = E(l_{ijk}) = -\lambda_{i,j,k} - \lambda_{i,j,k} - \lambda_{i,j,k} - \lambda_{i,j,k}.
\]

Consequently, \( W_p(\theta) \) and \( R_p(\theta) \) have expansions
\[
W_p(\theta) = T_\theta^2 + n^{-1/2} (k_2 - \frac{1}{2} k_3) T_\theta^2 + O_p(n^{-1}),
\]
\[
R_p(\theta) = T_\theta + n^{-1/2} (\frac{1}{2} k_2 - \frac{1}{6} k_3) T_\theta + O_p(n^{-1}).
\] (9.7)

Expansion (9.7) shows that
\[
E(R_p) = n^{-1/2} (k_1 - \frac{1}{6} k_3) + O(n^{-1}) = -z_0 + O(n^{-1}),
\]
\[
\text{var}(R_p) = 1 + O(n^{-1}), \quad \text{skew}(R_p) = O(n^{-1}).
\] (9.8)

Thus, the distribution of \( R_p(\theta) + \tilde{z}_0 \) is standard normal to error of order \( O(n^{-1}) \), and the approximate limit \( \hat{\theta}_p[\alpha] \) that satisfies
\[
R_p(\hat{\theta}_p[\alpha]) + \tilde{z}_0 = -z^{(\alpha)}
\] (9.9)
is second-order accurate. Moreover, comparing (9.7) with the Cornish-Fisher expansion in (8.2) shows that this limit is second-order correct with respect to \( T_\theta \). Approximate confidence limits obtained using (9.9) have been discussed by several authors, including Lawley (1956), Sprott (1980), McCullagh (1984), and Barndorff-Nielsen (1986). McCullagh (1984) and Barndorff-Nielsen (1986) have shown that these limits are second-order accurate conditionally, that is, they have conditional coverage error of order \( O(n^{-1}) \) given exact or approximate ancillary statistics. It follows that second-order conditional coverage accuracy is a property of all approximate confidence limits that are second-order correct with respect to \( T_\theta \). In contrast, limits that are second-order correct with respect to \( T_E \) typically have conditional coverage error of order \( O(n^{-1/2}) \). Conditional validity provides a good reason for preferring \( T_\theta \) over \( T_E \) to define "exact" confidence limits.
The profile log likelihood function \( l_p(\theta) \) is not a genuine likelihood. In particular, the expectation of the profile score, \( l_p^{(1)}(\theta) \), is not identically 0 and is generally of order \( O(1) \). It can be shown that

\[
E\{l_p^{(1)}(\theta)\} = (a - z_0)/\sigma + O(n^{-1}),
\]

and hence, the estimating equation \( l_p^{(1)}(\theta) = 0 \), which yields the estimate \( \hat{\theta} \), is not unbiased. To eliminate this bias, several authors, including Barndorff-Nielsen (1983,1994), Cox and Reid (1987, 1993), and McCullagh and Tibshirani (1990), have recommended that the profile log likelihood function \( l_p(\theta) \) be replaced by an adjusted version

\[
l_{ap}(\theta) = l_p(\theta) + d(\theta),
\]

where the adjustment function \( d(\theta) \) satisfies

\[
d(\theta) = (\hat{a} - z_0)T_O + O_p(n^{-1}), \tag{9.10}
\]

so that

\[
d^{(1)}(\theta) = -E\{l_p^{(1)}(\theta)\} + O_p(n^{-1}).
\]

Hence, \( E\{l_{ap}^{(1)}(\theta)\} = O(n^{-1}) \), and \( l_{ap}(\theta) \) behaves more like a genuine likelihood than does \( l_p(\theta) \). For instance, McCullagh and Tibshirani (1990) suggested the adjustment

\[
m(\theta) = -\int_\theta^\alpha \{a(\zeta_u) - z_0(\zeta_u)\}/\sigma(\zeta_u)\ du. \tag{9.11}
\]

The estimator \( \hat{\theta}_{ap} \) that maximizes \( l_{ap}(\theta) \) satisfies

\[
\hat{\theta}_{ap} = \hat{\theta} + (z_0 - a)\sigma + O_p(n^{-3/2}).
\]

The adjusted likelihood ratio statistic arising from \( l_{ap}(\theta) \) is

\[
W_{ap}(\theta) = 2\{l_{ap}(\hat{\theta}_{ap}) - l_{ap}(\theta)\},
\]

and its signed root is \( R_{ap}(\theta) = \text{sgn}(\hat{\theta}_{ap} - \theta)\sqrt{W_{ap}(\theta)} \). It can be shown that

\[
W_{ap}(\theta) = W_p(\theta) + (z_0 - a)T_O + O_p(n^{-1})
\]

\[
R_{ap}(\theta) = R_p(\theta) + (z_0 - a) + O_p(n^{-1}), \tag{9.12}
\]

so it follows from (6.8) that

\[
E(R_{ap}) = -a + O(n^{-1}), \quad \text{var}(R_{ap}) = 1 + O(n^{-1}), \quad \text{skew}(R_{ap}) = O(n^{-1}).
\]
Consequently, the approximate confidence limit \( \hat{\theta}_{ap}[\alpha] \) that satisfies

\[
R_{ap}(\hat{\theta}_{ap}[\alpha]) + \hat{a} = -z^{(\alpha)}
\]  

(9.13)

is a second-order accurate confidence limit. Expansion (9.12) shows that \( \hat{\theta}_{ap}[\alpha] = \hat{\theta}_p[\alpha] + O_p(n^{-3/2}) \), so \( \hat{\theta}_{ap}[\alpha] \) is also second-order correct with respect to \( T_O \). Confidence limits obtained by (9.13) have been discussed by DiCiccio and Efron (1992), DiCiccio and Martin (1993), Efron (1993) and Barndorff-Nielsen and Chamberlin (1994).

Numerical examples, especially in cases where the number of nuisance parameters is large, indicate that the standard normal approximation for \( R_{ap}(\theta) + \hat{a} \) can be much more accurate than for \( R_p(\theta) + z_0 \), and hence the limits obtained from (9.13) have better coverage accuracy than limits obtained from (9.12). Now, (9.8) suggests that the distribution of \( R_p(\theta) \) is affected by the presence of nuisance parameters at the \( O(n^{-1/2}) \) level through the quantity \( z_0 \). However, the distribution of \( R_{ap}(\theta) \) is insensitive to the presence of nuisance parameters at that level, because of the remarks made about \( a \) at (9.4). Consider again the orthogonal case with \( \theta = \zeta^t \). Let \( R(\theta) \) be the signed root of the likelihood ratio statistic that would apply if the nuisance parameters \( \zeta^2, \ldots, \zeta^p \) were known. It follows from the comparison of (9.3) and (9.4) that the distributions of \( R(\theta) \) and \( R_{ap}(\theta) \) agree to order \( O(n^{-1}) \), while the distributions of \( R(\theta) \) and \( R_p(\theta) \) agree only to order \( O(n^{-1/2}) \). Since \( R(\theta) \) does not require estimation of nuisance parameters, its distribution is likely to be fairly close to standard normal. On the other hand, because of presence of nuisance parameters, the distribution of \( R_p(\theta) \) can be far from standard normal, and asymptotic corrections can fail to remedy adequately the standard normal approximation.

These remarks can be illustrated by taking \( \theta \) to be the variance in a normal linear regression model with \( q \) regression coefficients. In this case, \( \theta \) is orthogonal to the regression coefficients, and

\[
\sigma^2 = \frac{2\theta^2}{n}, \quad a = \frac{2}{3\sqrt{2n}} + O(n^{-1}), \quad z_0 = \frac{q}{\sqrt{2n}} + \frac{2}{3\sqrt{2n}} + O(n^{-1}),
\]

by (9.4) and (9.5). Note that \( a \) does not involve the nuisance parameters, while \( z_0 \) reflects the nuisance parameters through its dependence on \( q \). In this case, \( (a - z_0)/\sigma = -q/(2\theta) \), and (9.11) produces the adjustment function \( d(\theta) = (q/2)\log\theta \). The effect making this adjustment to the profile log likelihood is to account for the degrees of freedom; in particular, \( \hat{\theta}_{ap} = n\hat{\theta}/(n - q) \). Table 4 shows, in the case \( n = 8 \) and \( q = 3 \), the true left-hand tail probabilities of approximate quantiles for \( R_p, R_{ap}, R \) and their mean-adjusted versions obtained using the standard normal approximation. Note the accuracy and the closeness of the approximation for \( R_{ap} \) and \( R \); in constrast, the approximation for \( R_p \) is very poor.
<table>
<thead>
<tr>
<th>nominal</th>
<th>(R_p)</th>
<th>(R_p + \hat{z}_0)</th>
<th>(R_{ap})</th>
<th>(R_{ap} + \hat{a})</th>
<th>(R)</th>
<th>(R + \hat{a})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.67</td>
<td>2.85</td>
<td>1.81</td>
<td>1.19</td>
<td>1.60</td>
<td>1.04</td>
</tr>
<tr>
<td>2.5</td>
<td>22.10</td>
<td>5.69</td>
<td>4.18</td>
<td>2.90</td>
<td>3.75</td>
<td>2.58</td>
</tr>
<tr>
<td>5</td>
<td>31.34</td>
<td>9.62</td>
<td>7.83</td>
<td>5.68</td>
<td>7.12</td>
<td>5.13</td>
</tr>
<tr>
<td>10</td>
<td>43.68</td>
<td>16.32</td>
<td>14.54</td>
<td>11.09</td>
<td>13.44</td>
<td>10.18</td>
</tr>
<tr>
<td>50</td>
<td>84.38</td>
<td>56.81</td>
<td>58.41</td>
<td>51.90</td>
<td>56.65</td>
<td>50.08</td>
</tr>
<tr>
<td>90</td>
<td>98.60</td>
<td>91.09</td>
<td>93.06</td>
<td>90.58</td>
<td>92.51</td>
<td>89.88</td>
</tr>
<tr>
<td>95</td>
<td>99.44</td>
<td>95.38</td>
<td>96.71</td>
<td>95.30</td>
<td>96.42</td>
<td>94.90</td>
</tr>
<tr>
<td>97.5</td>
<td>99.77</td>
<td>97.59</td>
<td>98.43</td>
<td>97.65</td>
<td>98.28</td>
<td>97.43</td>
</tr>
<tr>
<td>99</td>
<td>99.93</td>
<td>98.98</td>
<td>99.40</td>
<td>99.06</td>
<td>99.34</td>
<td>98.96</td>
</tr>
</tbody>
</table>

Table 4. True left-hand tail probabilities of approximate percentage points obtained from the standard normal approximation. Table entries are percentages.

Approximate confidence limits that are second-order correct with respect to \(T_O\) can be used to recover the profile and adjusted profile log likelihoods, at least to error of order \(O_p(n^{-1})\). Suppose that \(\hat{\theta}_O[\alpha]\) is second-order correct; then, by (6.9),

\[
R_p(\hat{\theta}_O[\alpha]) + \hat{z}_0 = -z^{(\alpha)} + O_p(n^{-1}).
\]

It follows that

\[
l_p(\hat{\theta}_O[\alpha]) = \text{constant} - \frac{1}{2}(z^{(\alpha)} + z_0)^2 + O_p(n^{-1}), \tag{9.14}
\]

and by (6.10),

\[
l_{ap}(\hat{\theta}_O[\alpha]) = \text{constant} - \frac{1}{2}(z^{(\alpha)} + z_0)^2 - \{(\hat{a} - \hat{z}_0)/\hat{\sigma}\} \hat{\theta}_O[\alpha] + O_p(n^{-1}). \tag{9.15}
\]

Approximations (9.14) and (9.15) to \(l_p(\theta)\) and \(l_{ap}(\theta)\) are especially useful in complex situations. Efron (1993) discussed the use of second-order correct confidence limits, particularly the ABC limits, to automatically construct implied likelihoods in both parametric and nonparametric situations.

Second-order accurate confidence limits can also be constructed by using Bayesian methods with noninformative prior distributions. Assume \(\theta = \zeta^1\), with the nuisance parameters \(\zeta^2, \ldots, \zeta^p\) not necessarily orthogonal to \(\theta\), and consider Bayesian inference based on a prior density \(\pi(\zeta)\). DiCiccio and Martin (1993) showed that the posterior distribution of

\[
R_p + \frac{1}{R_p} \log \left( \frac{S}{R_p} \right), \tag{9.16}
\]
is standard normal to error of order $O(n^{-3/2})$, where

$$S = l_1(\hat{\theta}_0) \left\{ -l^{11}(\hat{\theta}_0) \right\}^{1/2} \frac{|-l_{ij}(\hat{\theta}_0)|^{1/2}}{|-l_{ij}(\hat{\theta})|^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_0)},$$

and $|-l_{ij}(\hat{\theta}_0)|$ denotes the determinant of the $p \times p$ matrix $(-l_{ij}(\hat{\theta}_0))$. Thus, the quantity $\hat{\theta}_0[\alpha]$ that satisfies

$$R_p(\hat{\theta}_0[\alpha]) + \frac{1}{R_p(\hat{\theta}_0[\alpha])} \log \left( \frac{S(\hat{\theta}_0[\alpha])}{R_p(\hat{\theta}_0[\alpha])} \right) = -z_0 \quad (9.17)$$

agrees with the posterior $\alpha$ quantile of $\theta$ to error of order $O(n^{-2})$.

From a frequentist perspective,

$$S = T_0 + O_p(n^{-1/2}) = R_p + O_p(n^{-1/2}),$$

so the adjustment term $R_p^{-1} \log(S/R_p)$ in (6.16) is of order $O_p(n^{-1/2})$ under repeated sampling. Indeed, standard Taylor expansions show that

$$\frac{1}{R_p} \log \left( \frac{S}{R_p} \right) = z_0 + \sum_{i=1}^{p} \frac{\partial}{\partial \zeta_i} \left\{ \lambda^{i,1}(\lambda^{1,1})^{-1/2} \right\} + \frac{\pi_i(\zeta)}{\pi(\zeta)} \lambda^{i,1}(\lambda^{1,1})^{-1/2} + O_p(n^{-1}) \quad (9.18)$$

where $\pi_i(\zeta) = \partial \pi(\zeta)/\partial \zeta_i$. It is apparent from (6.18) that if the prior density $\pi(\zeta)$ is chosen to satisfy

$$\frac{\pi_i(\zeta)}{\pi(\zeta)} \lambda^{i,1}(\lambda^{1,1})^{-1/2} = -\sum_{i=1}^{p} \frac{\partial}{\partial \zeta_i} \{ \lambda^{i,1}(\lambda^{1,1})^{-1/2} \}, \quad (9.19)$$

then $R_p^{-1} \log(S/R_p) = z_0 + O_p(n^{-1})$. In this case, $\hat{\theta}_0[\alpha]$, the solution to (9.17), agrees to error of order $O_p(n^{-3/2})$ with $\hat{\theta}_0[\alpha]$, the solution to (4.9). Consequently, when the prior $\pi(\zeta)$ satisfies (4.19), $\hat{\theta}_0[\alpha]$ is second-order correct with respect to $T_0$, as is the posterior $\alpha$ quantile of $\theta$. These approximate confidence limits also have conditional coverage error of order $O_p(n^{-1})$ given exact or approximate ancillary statistics. Prior distributions for which the posterior quantiles are second-order accurate approximate confidence limits under repeated sampling are usually called noninformative.

Equation (9.19) was given by Peers (1965). When the nuisance parameters $\zeta^2, \ldots, \zeta^p$ are orthogonal to $\theta = \zeta^1$, this equation reduces to

$$\frac{\pi_1(\zeta)}{\pi(\zeta)} (\lambda_{1,1})^{-1/2} = -\frac{\partial}{\partial \zeta_1} (\lambda_{1,1})^{-1/2}.$$

Tibshirani (1989) showed that this equation has solutions of the form

$$\pi(\zeta) \propto (\lambda_{1,1})^{1/2} g,$$

where $g$ is arbitrary and depends only on the nuisance parameters.
References


