BOOTSTRAP CONFIDENCE INTERVALS
FOR PARAMETRIC PROBLEMS

by
Bradley Efron
Stanford University

TECHNICAL REPORT NO. 90
MARCH 1984

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT
2 R01 GM21215-09

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
BOOTSTRAP CONFIDENCE INTERVALS FOR PARAMETRIC PROBLEMS

By

BRADLEY EFRON

TECHNICAL REPORT NO. 90
March 1984

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT 2 R01 GM21215-09

Also prepared under National Science Foundation Grant MCS 80-24649, and issued as Technical Report No. 215, Stanford University, Department of Statistics.

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
BOOTSTRAP CONFIDENCE INTERVALS FOR PARAMETRIC PROBLEMS

Bradley Efron

Abstract

The problem of setting approximate confidence intervals for a real-valued function $\theta$ of an unknown parameter vector $\eta$ is important in statistical practice. The standard approximation based on maximum likelihood theory, $\hat{\theta} \pm \hat{\sigma} z^{(a)}$ is asymptotically correct, but can be quite misleading in small samples. We will discuss a bootstrap-based method which removes much of the error in the standard approximation, at the expense of considerably more computation. The discussion centers on a simple set of parametric families, in which it is possible to directly compare the standard and bootstrap methods with the correct answer.
BOOTSTRAP CONFIDENCE INTERVALS FOR PARAMETRIC PROBLEMS

Bradley Efron

1. Approximate Confidence Intervals.

The following problem is important in statistical applications: having observed a data vector \( y \) from a family of densities \( f_\eta(y) \) depending on an unknown parameter vector \( \eta \), we wish to set confidence intervals for a real-valued function of \( \eta \), say \( \theta = t(\eta) \). For example we might have \( y \) bivariate normal with mean vector \( \eta \) and covariance matrix the identity, and desire confidence intervals for \( \theta = \eta_1 \eta_2 \) (see Section 2.)

Exact confidence intervals for \( \theta \) are sometimes available, but in most cases, including the one just mentioned, only approximations are possible. The most famous approximation is that based on the maximum likelihood estimate (MLE),

\[
\theta \in \hat{\theta} \pm \hat{\sigma} z(\alpha),
\]

where \( \hat{\theta} \) is the MLE of \( \theta \), \( \hat{\sigma} \) an estimate of its standard deviation, usually based on the Fisher information matrix, and \( z(\alpha) \) the 100 \( - \alpha \) percentile point of a standard normal variate.

The standard intervals (1) have proved immensely useful in statistical practice. They have the great virtue of being automatic: a computer program can be written which, given the observation \( y \) and the parametric form \( f_\eta(y) \), produces the intervals (1) with no further input required from the statistician. Nevertheless the standard intervals can be quite inaccurate. Table 1 shows them applied to the case \( \theta = \eta_1 \eta_2 \) described above, having observed \( y = (2,4) \). The deviation from the almost exact intervals derived in Sections 3 and 4 is quite
noticeable. Changing the parameter of interest from $\theta$ to $\phi = \theta^2$ makes the comparison much worse.

The bias-corrected percentile method is another automatic algorithm for producing confidence vectors for $\theta$ a real-valued parameter. It is based on the bootstrap, Efron (1979, 1982), and can be applied in either parametric or non-parametric situations. In this paper we will only discuss its use in parametric families $f_\eta(y)$.

<table>
<thead>
<tr>
<th>for $\theta$</th>
<th>for $\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Intervals (1):</td>
<td>[0.64, 15.36]</td>
</tr>
<tr>
<td>Almost Exact Intervals:</td>
<td>[1.77, 17.03]</td>
</tr>
<tr>
<td>Bootstrap Intervals:</td>
<td>[1.77, 17.12]</td>
</tr>
</tbody>
</table>

Table 1. Central 90% confidence intervals for $\theta = \eta_1 \eta_2$ and $\phi = \theta^2$, having observed bivariate normal vector $y = (2,4)$. The almost exact intervals are based on the signed distance theory of Sections 3 and 4. The bootstrap intervals are the bias-corrected percentile intervals of Section 5.

The bootstrap intervals typically require more computational expenditure than the standard intervals (1). The reward is a considerable improvement in accuracy, at least in some situations. This is evident in Table 1.

Sections 2-4 discuss a restricted but still quite flexible class of problems for which it is possible to calculate almost exact confidence intervals for $\theta$. These intervals are useful in their own right, relating closely to Bartlett’s improvements on the likelihood ratio method, see Section 6. Section 5 describes the bias-corrected bootstrap method, and shows that the bootstrap intervals are a good approximation to the almost exact intervals, within our restricted class of problems. This is the main point of the paper. Section 7 briefly discusses a more difficult class of problems for which the bootstrap intervals are not
fully satisfactory, though still an improvement over (1), and where computationally still more ambitious methods are required. The discussion in the body of the paper is at a descriptive level, with all proofs deferred to the Appendix.

2. A Simple Class of Problems.

We introduce a simple class of parametric problems for which it is easy to see how the bootstrap method corrects certain deficiencies of the standard intervals (1).

Suppose that the observed data vector $y$ has a multivariate normal distribution in $k$ dimensions, with mean vector $\eta$ and covariance matrix the identity,

$$ y \sim N_k(\eta, I) \text{,} $$

and that having observed $y$ we want to set confidence intervals for a real-valued function of $\eta$, say $\theta = t(\eta)$. The level surfaces of constant $\theta$ value

$$ C_{\theta} = \{ \eta : t(\eta) = \theta \} \text{,} $$

are assumed to be smooth, the function $t(\eta)$ having continuous second partial derivatives.

Three simple examples will be used for numerical illustration in what follows. Example 1 (ratio estimation): dimension $k = 2$, $\theta = \eta_2 / \eta_1$. In this case the surfaces $C_{\theta}$ are straight lines passing through the origin. Example 2 (non-central $\chi^2$): $k = 6$, $\theta = \|\eta\|$. In this case the level surfaces $C_{\theta}$ are spheres of radius $\theta$ centered at the origin. Example 3 (product of means): $k = 2$, $\theta = \eta_1 \eta_2$. This is the example discussed in Section 1. The level surfaces are hyperbolas in the plane. Efron (1983) showed this problem arising naturally in the comparison of nonnested linear models; the good coverage properties of the bootstrap intervals observed there suggested the results of this paper.
Figure 1 schematically illustrates the curved level surfaces and the data vector $y$. Of course $y$ is also $\hat{\eta}$, the unrestricted maximum likelihood estimate for $\eta$. The point on any particular $C_\theta$ nearest to $y$ in Euclidean distance is labelled $\hat{n}(\theta)$, it being the restricted MLE for $\eta$ assuming $\eta \in C_\theta$. The line through $\hat{n}(\theta)$ orthogonal to $C_\theta$ is labelled $L_{\hat{n}(\theta)}$. We will think of this as a one-dimensional axis, with origin at $\hat{n}(\theta)$; $x$ will indicate signed distance along $L_{\hat{n}(\theta)}$. It doesn't matter which sign convention is used, as long as it is.
defined consistently, but for the illustrations in this paper, $x$ will be taken positive in the direction away from the curvature of $C_\theta$, as indicated by the arrowhead in Figure 1.

The signed distance of $y$ from $\hat{n}(\theta)$, labelled $X_\theta$, plays a central role in the confidence interval theory for $\theta$. Our first theorem, stated in the next section, is that if $\eta \in C_\theta$, then $X_\theta$ is approximately normal, $X_\theta \sim N(\mu_\eta, \tau^2_\eta)$. The mean $\mu_\eta$ and variance $\tau^2_\eta$ depend in a simple way on the curvature of $C_\theta$ at the point $\eta$. If $C_\theta$ is flat, as in example 1 (ratio estimation), then $X_\theta$ is obviously $N(0,1)$. However any curvature in $C_\theta$ makes $\mu_\eta$ positive and $\tau^2_\eta < 1$; the approximately normal distribution of $X_\theta$ is shifted in the positive direction, away from the curvature of $C_\theta$ in Figure 1.

The shift of $X_\theta$ can be quite dramatic. For example 2 (non-central $X^2_\theta$) with $\theta = 5$, the theorem of Section 3 gives $X_\theta \sim N(.5, .95^2)$. Table 2 shows this approximation to be accurate. In terms of Figure 1, the surface $C_\theta$ containing the true $\eta$ is a 6-dimensional sphere of radius 5; $X_\theta$, taken positive or negative as $y$ is outside or inside of $C_\theta$, is almost perfectly normal, but with distribution centered half a standard deviation outside of $C_\theta$.

<table>
<thead>
<tr>
<th>Theoretical percentiles</th>
<th>.025</th>
<th>.05</th>
<th>.10</th>
<th>.25</th>
<th>.75</th>
<th>.90</th>
<th>.95</th>
<th>.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>$.5 + .95 z^{(\alpha)}$</td>
<td>-1.36</td>
<td>-1.06</td>
<td>-.72</td>
<td>-.14</td>
<td>1.14</td>
<td>1.72</td>
<td>2.06</td>
<td>2.36</td>
</tr>
<tr>
<td>Actual percentiles</td>
<td>-1.36</td>
<td>-1.08</td>
<td>-.73</td>
<td>-.16</td>
<td>1.14</td>
<td>1.73</td>
<td>2.07</td>
<td>2.38</td>
</tr>
</tbody>
</table>

Table 2. Percentiles of the distribution of $X_\theta$ in example 2, $\theta = 5$; $y \sim N_6(\eta, I)$, with $\eta$ on a sphere of radius 5; $X_\theta$ is the signed distance from $y$ to the nearest point on the sphere, taken positive or negative as $y$ is outside or inside the sphere. The theorem in Section 3 gives $X_\theta \sim N(.5, .95^2)$. This agrees well with the actual distribution of $X_\theta$, determined from 100,000 Monte Carlo simulations.
3. Distribution of the Signed Distance.

This section presents a theorem on the approximate normality of the signed distance \( X_\theta \). First we need a quantitative description of the curvature of the level surfaces \( C_\theta \) at a point \( \eta \).

Let \( T_\eta \) be the tangent hyperplane to \( C_\theta \) at a point \( \eta \). We can choose any basis of \( k-1 \) orthogonal unit vectors in \( T_\eta \), and let \( v = (v_1, v_2, \ldots, v_{k-1}) \) be the coordinates of a point in \( T_\eta \) with respect to this basis, \( v = 0 \) corresponding to the point \( \eta \). Also let \( u \) indicate distance orthogonal to \( T_\eta \), so \( (u, v) \) together constitute a rotated \( k \)-dimensional coordinate system for the space containing \( y \). We have two choices for the positive direction of \( u \), but whichever one we choose, the opposite direction will indicate the positive direction of the signed distance \( X_\theta \). This convention is illustrated in Figure 1.

Since \( C_\theta \) is tangent to \( T_\eta \) at \( \eta \), the Taylor series describing \( C_\theta \) for points near \( \eta \), that is for \( v \) near 0, begins

\[
u = v' d_\eta v
\]

for some symmetric \( (k-1) \times (k-1) \) matrix \( d_\eta \), not necessarily positive definite.

For \( C_\theta \) the 6-dimensional sphere of radius 5, considered in Table 1, \( d_\eta = (0.1) I \). The magnitude of the elements of \( d_\eta \) measures the curvature of \( C_\theta \) at \( \eta \), in a way made more precise later. If \( C_\theta \) is flat, as in example 1, then \( d_\eta = 0 \).

For dimension \( k = 2 \), \( d_\eta \) equals one half the usual definition of curvature for \( C_\theta \) at point \( \eta \). We will call \( d_\eta \) simply the curvature matrix of \( C_\theta \) at point \( \eta \).

The confidence interval theory presented here is asymptotic in the sense that it becomes increasingly accurate as \( d_\eta \) approaches 0. We can state this more conventionally by assuming that the observed data actually consists of \( n \) independent and identically distributed (i.i.d.) vectors \( y_1, y_2, \ldots, y_n \sim N_k(\eta, I) \). Then \( \hat{\eta} = \bar{y} \) is sufficient for \( \eta \), \( \bar{y} \sim N_k(\eta, I/n) \), and we can call \( \hat{X}_\theta \) the
signed distance of $\tilde{y}$ from $\hat{\eta}(\theta)$. Figure 1 still describes the situation, except with $y$ and $X_\theta$ replaced by $\tilde{y}$ and $\tilde{X}_\theta$.

Rescaling the sufficient vector $\tilde{y}$ to $y = \sqrt{n} \tilde{y} \sim N_k(\sqrt{n}\eta, \mathbf{I})$ restores its covariance matrix to $\mathbf{I}$. Figure 1 applies again exactly as shown, except that we have to remember that every parameter vector $\eta$ has been mapped into $\eta(n) = \sqrt{n} \eta$.

The level surfaces $C_\theta$ map into $C_{\theta}(n) \equiv \sqrt{n} C_\theta$, with curvature matrix at point $\eta(n)$ divided by $\sqrt{n}$, say $d_{\eta(n)}(n) = d_{\eta}/\sqrt{n}$ [as can be seen from (3), which gives $(u/\sqrt{n}) = (v/\sqrt{n})' \ d_{\eta}(v/\sqrt{n})$, or $u = v'(d_{\eta}/\sqrt{n})v$, as the local equation for $C_{\theta}(n)$ near $\eta(n)$]. The theorem which follows can be interpreted as saying that

$$X_\theta \equiv \sqrt{n} \tilde{X}_\theta \quad (4)$$

is asymptotically normal, with expectation $\text{tr}(d_{\eta(n)}(n)) + O(n^{-3/2})$ and standard deviation $[1 - \text{tr} d_{\eta(n)}^2(n)] + O(n^{-3/2})$.

In any specific situation, for example that of Table 2, there will be a well-defined curvature matrix which applies after the covariance matrix of the sufficient vector has been rescaled to $\mathbf{I}$. This curvature matrix is called simply $d_\eta$, rather than $d_{\eta(n)}(n)$, in what follows. The elements of $d_\eta$ are assumed to be of magnitude $O(n^{-1/2})$ because of the rescaling argument above.

We can now state the main theorem. Its proof appears in the Appendix.

**Theorem 1.** If $\eta \in C_\theta$, then the signed distance $X_\theta$ is asymptotically normal with first four cumulants

$$X_\theta \sim \left[ \text{tr}(d_{\eta}), \{1 - \text{tr} d_{\eta}^2\}^2, 0, 0 \right], \quad (5)$$

through $O(n^{-1})$, the errors in (5) being $O(n^{-3/2})$.

For any $(k-1) \times (k-1)$ orthogonal matrix $\Gamma$, the trace satisfies $\text{tr}((d_{\eta}\Gamma') = \text{tr}(d_{\eta})$ and $\text{tr}(d_{\eta}(\Gamma^2) = \text{tr}(d_{\eta})$, so the terms of (5) do not depend on the choice of basis in $T_\eta$. Geometrically, $2\text{tr}(d_{\eta})$ is the sum of the usual curvatures of $C_\theta$ in the $k-1$ orthogonal directions through $\eta$. 

7
The expectation \( \text{tr}(d_{\hat{\eta}}) \) in (5) is \( O(n^{-1/2}) \). This is the main term disturbing \( \chi_\theta \) from its asymptotic \( N(0,1) \) distribution. The standard deviation \( 1 - \text{tr}(d_{\hat{\eta}}^2) \) differs from 1 by only \( O(n^{-1}) \). (Notations \( \hat{\cdot} \) and \( \hat{\cdot} \) indicate accuracy through \( O(n^{-1}) \), with error \( O(n^{-3/2}) \).) In the example of Table 2, \( \text{tr}(d_{\eta}) = .50 \), \( 1 - \text{tr}(d_{\eta}^2) = .950 \). The fact that the third and fourth cumulants \( \hat{\cdot} \) 0 accounts for the impressive accuracy of the normal approximation seen in Table 2. Another example is given in Section 4.

As a point of comparison, the student's t correction, which would be necessary if we had estimated the scale of \( \gamma \) instead of assuming it known, is \( O(n^{-1}) \) in the coordinates we are using. The t effect is smaller by order \( O(n^{-1/2}) \) than the shift in \( \chi_\theta \)'s mean due to the curvature of the level surfaces. (See the last paragraph of Section 5.)

The proof of Theorem 1 (and the other results in the Appendix, is formal in nature, and doesn't provide error bounds for the approximation. The accuracy of the Theorem breaks down in regions of high curvature of \( C_{\theta} \). Numerical experimentation for \( C_{\theta} \), a circle in the plane showed good accuracy when the circle's radius was \( \geq 3 \), and reasonable accuracy even for radius as small as 2 except in the lower tail of \( \chi_\theta \).

4. Confidence Intervals for \( \theta \).

The near normality of the signed distance \( \chi_\theta \) can be used to construct accurate confidence intervals for \( \theta \). Theorem 1 suggests that the normalized signed distance \( W_{\theta} \) should be nearly unit normal,

\[
W_{\theta} \equiv \frac{\chi_{\theta} - \text{tr} d_{\hat{\eta}(\theta)}}{1 - \text{tr} d_{\hat{\eta}(\theta)}^2},
\]

where \( d_{\hat{\eta}(\theta)} \) in the matrix (3) describing the curvature of \( C_{\theta} \) at \( \hat{\eta}(\theta) \), the point on \( C_{\theta} \) nearest to \( y \).

Corollary. If \( \eta \in C_{\theta} \) then the normalized signed distance \( W_{\theta} \) is asymptotically normal with first four cumulants \( W_{\theta} \approx [0,1,0,0] \), through \( O(n^{-1}) \).
The proof appears in the Appendix. For example \( \theta = \eta_1 \eta_2 \), with \( \eta = (2,4) \). 40,000 Monte Carlo replications gave the following comparison between the actual distribution of \( W_\theta \) and its limiting \( N(0,1) \) distribution:

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>.025</th>
<th>.05</th>
<th>.10</th>
<th>.90</th>
<th>.95</th>
<th>.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentiles of ( W_\theta ):</td>
<td>-1.94</td>
<td>-1.64</td>
<td>-1.28</td>
<td>1.28</td>
<td>1.65</td>
<td>1.97</td>
</tr>
<tr>
<td>( z(\alpha) ):</td>
<td>(-1.96)</td>
<td>(-1.64)</td>
<td>(-1.28)</td>
<td>(1.28)</td>
<td>(1.64)</td>
<td>(1.96)</td>
</tr>
</tbody>
</table>

Figure 2 illustrates how \( W_\theta \) is used to test the hypothesis \( \eta \in C_\theta \). According to the Corollary, and a straightforward application of Edgeworth expansions, the interval

\[
z(\alpha) \leq W_\theta \leq z(1-\alpha)
\]

(7)

is an acceptance region of size \( 1-2\alpha + o(n^{-3/2}) \) for the hypothesis \( \eta \in C_\theta \). If \( \alpha = .05 \) for instance, then \(-1.645 \leq W_\theta \leq 1.645 \) tests \( \eta \in C_\theta \) at level \( 1-2\alpha = .90 \), with probability of error \( \approx .05 \) on each side of (7).

Figure 2. The test based on the normalized signed distance accepts the hypothesis \( \bar{\eta} \in C_{\bar{\theta}} \) for \( y \) in a band about \( C_{\bar{\theta}} \), as indicated at left. The band is shifted in the positive direction away from \( C_{\bar{\theta}} \) by an amount depending upon the curvature of \( C_{\bar{\theta}} \). The corresponding confidence interval for \( \theta \), having observed \( y = \hat{\eta} \), is shifted in the negative direction from \( C_{\bar{\theta}} \), as indicated at right.
Relationship (7) can be expressed as

\[ \text{tr } d_{\hat{\theta}}(\theta) + (1 - \text{tr } d_{\hat{\theta}}^2(\theta))z^{(\alpha)} \leq X_\theta \leq \text{tr } d_{\hat{\theta}}(\theta) + (1 - \text{tr } d_{\hat{\theta}}^2(\theta))z^{(1-\alpha)} \]

which shows that the acceptance region is shifted in the positive direction from \( C_{\theta} \) along each line \( L_{\hat{\theta}}(\theta) \), by amount \( \text{tr } d_{\hat{\theta}}(\theta) \). The more curved \( C_{\theta} \) is at \( \hat{\theta}(\theta) \), the larger the shift.

The test regions (7) can be inverted to give an approximate 1-2\( \alpha \) central confidence interval for \( \theta \), say \( \theta \in [\hat{\theta}(\alpha), \hat{\theta}(1-\alpha)] \). Having observed \( y \), the interval consists of those values of \( \theta \) such that \( W_{\hat{\theta}} \) satisfies (7). As usual, inverting the test regions reverses the asymmetry of the intervals: the confidence interval for \( \theta \) is shifted along \( L_{\hat{\theta}} \) in the negative direction from \( y = \hat{\theta} \), as illustrated in Figure 2.

Table 3 applies to example 2 (non-central \( \chi^2_0 \)), supposing that we have observed a data vector \( y \) with \( ||y|| = 5 \). Exact confidence limits for \( \theta \) were obtained in the usual way using the non-central \( \chi^2_0 \) distribution. For instance the .05 limit point 2.68 was obtained from \( P\{\chi^2_0(2.68^2) < s^2\} = .95 \). The signed distance confidence limits obtained from (7) are seen to agree well with the exact results.

<table>
<thead>
<tr>
<th>.</th>
<th>( \alpha : )</th>
<th>.05</th>
<th>.10</th>
<th>.25</th>
<th>.75</th>
<th>.90</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact limits:</td>
<td>2.68</td>
<td>3.08</td>
<td>3.75</td>
<td>5.18</td>
<td>5.82</td>
<td>6.19</td>
<td></td>
</tr>
<tr>
<td>signed distance limits (7):</td>
<td>2.71</td>
<td>3.08</td>
<td>3.71</td>
<td>5.16</td>
<td>5.80</td>
<td>6.19</td>
<td></td>
</tr>
<tr>
<td>approximate limits (8):</td>
<td>2.77</td>
<td>3.15</td>
<td>3.79</td>
<td>5.21</td>
<td>5.85</td>
<td>6.23</td>
<td></td>
</tr>
<tr>
<td>approximate bootstrap limits (13):</td>
<td>2.94</td>
<td>3.28</td>
<td>3.86</td>
<td>5.14</td>
<td>5.75</td>
<td>6.06</td>
<td></td>
</tr>
<tr>
<td>standard limits (1):</td>
<td>3.36</td>
<td>3.72</td>
<td>4.33</td>
<td>5.67</td>
<td>6.28</td>
<td>6.64</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Confidence intervals for example 2 (non-central \( \chi^2_0 \)), having observed \( ||y|| = 5 \). The exact confidence limits for \( \theta = ||\hat{\theta}|| \), obtained from the non-central \( \chi^2_0 \) distribution, agree well with the signed distance confidence limits obtained from (7).
Let points on $L_{\hat{n}}$ be denoted by $\hat{n} + x\hat{\Delta}$, where $\hat{\Delta}$ is a unit vector pointing in the positive direction along $L_{\hat{n}}$. In Figure 2 we have indicated the $x$ values of the points on $L_{\hat{n}}$ delimiting the confidence interval for $\theta$ by $x(\alpha)$ and $x(1-\alpha)$. The following approximation for $x(\alpha)$ (and correspondingly for $x(1-\alpha)$) is useful for comparison with the bootstrap intervals of Section 5:

$$x(\alpha) \approx -\text{tr} \hat{d} + \frac{(1 - \text{tr} \hat{d}^2)z(\alpha)}{1 - \frac{\|\hat{c}\|^2}{z(\alpha)^2}} - (\text{tr} \hat{e})z(\alpha).$$  \hspace{1cm} (8)

This formula, which is accurate through $O(n^{-1})$, is motivated in the Appendix. Here we will only describe the meaning of the various terms appearing in (8).

Denote by $C(x)$ the level surface $C_\theta$ intersecting $L_{\hat{n}}$ at $\hat{n} + x\hat{\Delta}$. At $x = 0$, $C(0) = C_{\hat{n}}$ is described locally by $u = v'\hat{d}v$ as in (3), with $\hat{d} \equiv d_{\hat{n}}$.

The $(k-1) \times (k-1)$ matrix $\hat{e}$ is the first derivative of the curvature matrix of $C(x)$ as $x$ moves away from $0$,

$$\hat{e} = \frac{\partial}{\partial x} d_{\hat{n}+x\hat{\Delta}} \bigg|_{x=0}. \hspace{1cm} (8A)$$

The trace of $\hat{e}$ determines how quickly the curvature of the level surfaces is changing near $y = \hat{n}$,

$$\text{tr} \ d_{\hat{n}+x\hat{\Delta}} = \text{tr} \hat{d} + x \text{ tr} \hat{e}.$$

The elements of $\hat{e}$ are $O(n^{-1})$, compared to $O(n^{-1/2})$ for $\hat{d}$, as shown in the Appendix.

Finally, let $\hat{\Delta}(x)$ be the unit orthogonal vector to $C(x)$ at the point $\hat{n} + x\hat{\Delta}$, as illustrated in Figure 2. The cosine of the angle between $\hat{\Delta}(x)$ and $\hat{\Delta} = \hat{\Delta}(0)$, say $\text{Co}(x) \equiv \hat{\Delta}(x) \cdot \hat{\Delta}$, is locally quadratic near $x = 0$, and the term $\|\hat{c}\|^2$ appearing in (8) is just the quadratic coefficient,

$$\text{Co}^2(x) \equiv 1 - \|\hat{c}\|^2 x^2. \hspace{1cm} (8B)$$
In other words, \( \| \hat{c} \| \) measures the rate of rotation of the level surfaces near the point \( \hat{y} = \hat{n} \). The quantity \( \| \hat{c} \|^2 \), like \( \text{tr} \hat{e} \), is \( O(n^{-1}) \).

For the noncentral \( \chi_0^2 \) example it is easy to calculate \( \hat{d} = (2 \| \hat{y} \|)^{-1} I \),
\( \hat{e} = -2 \hat{d} \), and \( \| \hat{c} \|^2 = 0 \). The third line of Table 2 shows approximation (8) performing well.

In example 1 (ratio estimation), \( d_1 = 0 \) since the \( C_\theta \) are straight lines. In this case \( W_\theta = \chi_0 \sim N(0,1) \) exactly, and the \( W_\theta \) intervals based on inverting (7) agree exactly with the Fieller situation, Fieller (1954),

\[
\theta \in \hat{\theta} \pm \frac{\hat{\delta}^2 (1 - \hat{\delta}^2 + \hat{\theta}^2)}{1 - \hat{\delta}^2} \cdot (\hat{\delta}^2 \equiv \left( \frac{z(\alpha)}{y_1} \right)^2).
\]  

(9)

Approximation (8) turns out to be

\[
\theta \in \left[ \tan^{-1} \hat{\theta} + \tan^{-1} \left( \frac{z(\alpha)}{\| y \|} \right), \tan^{-1} \hat{\theta} + \tan^{-1} \left( \frac{z(1-\alpha)}{\| y \|} \right) \right],
\]

(10)

where we have used \( \hat{d} = \hat{e} = 0 \), \( \| \hat{c} \|^2 = 1/\| y \|^2 \). Numerical calculation shows that (10) closely approximates (9). If the observed vector \( y \) equals \( (3,3) \), so \( \hat{\theta} = y_2/y_1 = 1 \), then the exact central 90% \( (\alpha=.05) \) interval for \( \theta \) based on (9) is \( [.408,2.452] \), compared to \( [.409,2.443] \) from (10).

To summarize this section, the normalized signed distance \( W_\theta \) is an approximate normal pivotal for \( \theta \). The signed distance confidence intervals based on inverting (7) are both accurate and appropriate: they tend to give nearly the claimed coverage probabilities, and also to be inferentially correct since they are based on an appropriate test statistic. Formula (8), which is based on the local geometry of the surfaces \( C_\theta \) near the point \( y = \hat{n} \), gives an accurate approximation to the intervals obtained from (7).
5. Bootstrap Confidence Intervals.

This section discusses a bootstrap method for setting confidence intervals, Efron (1981, 1982), applied to the simple class of parametric problems introduced in Section 2. Asymptotically the bootstrap intervals coincide with the standard intervals (1). However we will see that the bootstrap method agrees more closely with the almost exact intervals of Section 4, and so has better small-sample properties. In particular the bootstrap method has two advantages over the standard intervals: it takes into account the geometry of the level surfaces $C_\theta$; it does not depend on the name "$\theta$" attached to the level surfaces $C_\theta$, and so cannot be mislead by transformations like the change from $\theta$ to $\phi = \theta^2$ in Section 1.

The bias-corrected percentile method, applied to the maximum likelihood estimate $\hat{\theta} = t(\hat{\eta}) = t(y)$, constructs approximate confidence intervals for $\theta$ in the following way: (1) the probability mechanism generating the data is estimated by maximum likelihood, which in this case means estimating $\eta$ by $\hat{\eta} = y$. (2) Bootstrap data vectors $y^*(1), y^*(2), \ldots, y^*(B)$ are obtained by i.i.d. sampling from $f_\hat{\eta}(\cdot)$. (3) The corresponding bootstrap MLE's $\hat{\theta}^*(b) = t(y^*(b))$ are calculated, $b = 1, 2, \ldots, B$, giving cumulative distribution function (c.d.f.)

$$\hat{G}(s) = \#\{\hat{\theta}^*(b) < s\}/B.$$  (4) The quantity

$$z_0 = \Phi^{-1}(\hat{G}(\hat{\theta}))$$  (11)

is calculated, $\Phi$ being the standard normal c.d.f. (5) Finally, the central 1-2$\alpha$ interval for $\theta$ is taken to be

$$\theta \in [\hat{G}^{-1}\Phi(2z_0 + z(\alpha)), \hat{G}^{-1}\Phi(2z_0 + z(1-\alpha))] \tag{12}$$

In this paper (12) will be referred to simply as a bootstrap confidence interval for $\theta$. 

13
Notice that if \( \hat{G}(\hat{\theta}) = .50 \) then \( z_0 = 0 \) and the interval (12) is \([\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)]\), the \( \alpha \) and \( 1-\alpha \) percentiles of the bootstrap distribution. If \( \hat{G}(\hat{\theta}) \neq .50 \) then the \( z_0 \) term compensates for the bias of \( \hat{\theta} \) as an estimator of \( \theta \), as motivated in Section 10.7 of Efron (1982).

Step (2) of the bootstrap algorithm can be, but doesn't have to be, carried out by Monte Carlo. In this paper we are considering only parametric applications of the bootstrap, so that the bootstrap c.d.f. \( \hat{G} \) can be approximated by a variety of familiar parametric techniques. In Efron (1983), for example, Edgeworth expansions are used to approximate \( \hat{G} \) (giving, effectively, \( B = \infty \) at step (2)). Our next theorem gives a direct approximation for the intervals (12).

The bootstrap interval (12), call it \( \theta \in [\hat{\theta}_B(\alpha), \hat{\theta}_B(1-\alpha)] \), corresponds to an interval \([x_B(\alpha), x_B(1-\alpha)]\) of \( \hat{\eta} \), using the notation pictured on the right side of Figure 2.

Theorem 2. The limits of the bootstrap interval (12) are

\[
x_B(\alpha) = -\text{tr} \hat{d} + (1-\text{tr} \hat{d}^2)z(\alpha) + (2\text{tr} \hat{d}^2 + \text{tr} \hat{e})z(\alpha)
\]

through \( O(n^{-1}) \), and similarly for \( x_B(1-\alpha) \). (Proof appears in the Appendix.)

Comparing (13) with (8), we see that the bootstrap intervals correctly match the \( O(1) \) term \([z(\alpha)]\), the \( O(n^{-1/2}) \) term \([\text{tr} \hat{d}]\), and a portion of the \( O(n^{-1}) \) term [the denominator \( 1 - \frac{\|\hat{e}\|^2}{2} z(\alpha)^2 \), which comes from the rotation of the level surfaces near \( \hat{\eta} = y \)]. They err by \( O(n^{-1}) \) in the term \([\text{tr} \hat{e}]z(\alpha)\), which relates to the change in curvature of the level surfaces near \( \hat{\eta} = y \). The fourth line of Table 2 shows the bootstrap intervals performing moderately well in example 2.

At this point it is interesting to return to the standard intervals (1). Comparison with the other intervals is complicated by the fact that (1), unlike
(7), (8), or (13), depends upon the name "\( \theta \)" attached to the level surfaces \( C_\theta \), and not only upon their geometric shapes. The comparison is easiest if the function \( t(\hat{n}+x\hat{A}) \) is linear in \( x \), i.e. if the name \( \theta \) is linear in distance along \( L_\hat{n} \). Then the standard intervals (1) are delimited on \( L_\hat{n} \) by say, \( x_S(\alpha) \) and \( x_S(1-\alpha) \), where

\[
x_S(\alpha) = z(\alpha).
\]

This agrees with (8) to \( O(1) \), but not to the next term \( O(n^{-1/2}) \). The standard method is first order but not second order correct. The bootstrap intervals are second order but not third order correct.

Example 2 happens to have \( t(\hat{n}+x\hat{A}) \) linear in \( x \). The large effect of the \( O(n^{-1/2}) \) error is apparent in the fifth line of Table 2. Changing the name of the level surfaces, for example considering \( \phi = \theta^2 \) rather than \( \theta \), adds another error of \( O(n^{-1/2}) \) to (1) as compared to (8). As we saw in Section 1, the naming error can be enormous.

Table 4 compares, for example 3, the almost exact signed distance intervals of Section 4 obtained from either (7) or (8) with the bootstrap intervals and also with the standard intervals. The bootstrap intervals better match the almost exact answers here than in Table 3, because the \( \text{tr} \hat{e} \) curvature term is smaller in this case. (In fact, example 2 was chosen to exhibit especially large curvature effects.)

In example 1 (ratio estimation), (13) and (8) agree exactly, since \( \hat{d} = \hat{e} = 0 \). Direct Monte Carlo simulation confirmed that the bootstrap intervals are virtually identical to the signed distance intervals (10) in this case. We have already seen that the latter agrees closely with the exact Fieller solutions. (Interestingly, the bootstrap interval \( [\hat{G}^{-1}(\alpha),\hat{G}^{-1}(1-\alpha)] \) which ignores the bias adjustment term \( z_0 \) in (11), (12) gives the Creasy (1954) fiducial solution rather than the Fieller solution for \( \theta = \eta_2/\eta_1 \).)
Table 4. Confidence intervals for example 3 \((\theta = \eta_1 \eta_2)\), having observed \(y = (2,4)\). The bootstrap intervals agree closely with the signed distance intervals based on (7). The Bayes limits are obtained from the improper prior \((\eta_1^2 \eta_2)^{1/2} d\eta_1 d\eta_2\), suggested as being appropriate for this problem, in a frequentist sense, in an unpublished Stanford technical report by Charles Stein.

Robison (1964) discusses a large class of problems where the level surfaces \(C_{\theta}\) are linear; in such problems we expect the bootstrap intervals to closely match the exact answers, since they are accurate through \(O(n^{-1})\).

Suppose that instead of (2) we observe \(y \sim N_k(n, \sigma^2 I)\), with \(\sigma^2\) unknown but estimated by \(\hat{\sigma}^2 = \sigma^2 \chi^2_m / m\) independent of \(y\). If actually \(y = \sqrt{n} \bar{y}\) as in the repeated sampling argument leading up to (4), then \(m = k(n-1)\). The signed distance theory of Section 4 is modified in the obvious way for this situation, for example replacing \(z^{(\alpha)}\) by \(\tilde{\sigma}_m t^{(\alpha)}_m\) in (8), where \(t^{(\alpha)}_m\) is the 100\( \cdot \alpha\) percentile point of a student's \(t_m\) distribution. The bootstrap algorithm can still be carried out, by bootstrap sampling from \(N_k(\bar{y}, \hat{\sigma}^2 I)\) at step (2), but the resulting interval (12) does not incorporate a \(t\) correction, so another error of \(O(n^{-1})\) has crept in. A quick remedy is to replace \(\alpha\) in (12) by \(\alpha_m\), defined by \(z^{(\alpha_m)} = t^{(\alpha)}_m\) (e.g. if \(\alpha = .05\), \(m = 20\), then \(\alpha_m = .042\)), which according to (13) corrects the bootstrap intervals. However in the more complicated situation of Section 6 the proper choice of \(m\) is not likely to be obvious, so that this correction cannot be made.

The reason for pursuing general methods like the bootstrap is the hope that they can be applied in an automatic way to general problems, with some promise of good performance. The standard intervals (1), despite their limitations, have served applied statisticians well in this respect. So far we have shown that the bootstrap intervals improve on (1) within the rather specialized class of problems introduced in Section 2. This section extends that class considerably. These results are closely related to Bartlett's improvements on the likelihood ratio test, Barndorff-Nielsen and Cox (1983), as described later.

Suppose that instead of having \( y \sim N_k(\eta, I) \) as in (2), the statistician sees a transformed version of the same problem, say

\[
\tilde{y} = h_1(y), \quad \tilde{\eta} = h_2(\eta),
\]

(14)

where \( h_1 \) and \( h_2 \) are continuously differentiable one-to-one mappings of \( k \)-dimensional space into itself. That is, he observes \( \tilde{y} \) from the mapped density \( \tilde{f}_{\tilde{\eta}}(\cdot) \), and desires a confidence interval for a function \( \theta = \tilde{t}(\tilde{\eta}) \).

If \( h_1 \) and \( h_2 \) are known, then the inverse mappings \( y = h_1^{-1}(\tilde{y}) \) and \( \eta = h_2^{-1}(\tilde{\eta}) \) convert the situation back to (2). The confidence interval desired is for the function \( \theta = t(\eta) \equiv \tilde{t}(h_2(\eta)) \). The theory of the previous sections applies, giving nearly exact confidence intervals for \( \theta \). Unfortunately, it may be practically impossible to discover \( h_1 \) and \( h_2 \).

A useful feature of the bootstrap intervals is that the statistician need not know \( h_1 \) and \( h_2 \); the bias-corrected percentile method, applied to the MLE, automatically produces the same confidence intervals for \( \theta \) whatever the mappings \( h_1 \) and \( h_2 \) may be. The method is invariant under transformations (14). In other words, if there exists transformations \( y = h_1^{-1}(\tilde{y}), \eta = h_2^{-1}(\tilde{\eta}) \) such that
\( y \sim N_k(\eta, I) \) (for all values of \( \tilde{\eta} \)), then the bootstrap method produces nearly correct confidence intervals for \( \theta \), with properties as stated in Section 5. A proof is given in the Appendix.

We need to say what "the bootstrap method" is for a general parametric family \( \tilde{f}_\tilde{\eta}(\cdot) \). Let \( \tilde{\eta}^{\text{MAX}} \) indicate the MLE of \( \tilde{\eta} \), and generate bootstrap data vectors \( \tilde{y}^*(1), \tilde{y}^*(2), \ldots, \tilde{y}^*(B) \) iid \( \tilde{f}_{\tilde{\eta}^{\text{MAX}}} \). This gives corresponding bootstrap MLE's for \( \theta \), say \( \hat{\theta}^*(b) \) based on \( \tilde{y}^*(b) \). Now proceed as in steps (3), (4), (5) of the algorithm described in Section 5, leading to the interval (12), which is what we mean here by the bootstrap interval for \( \theta \).

The standard intervals (1) also are invariant under transformations (14) since both the MLE \( \hat{\theta} \) and the Fisher information standard deviation estimate \( \hat{\sigma} \) are invariant. (The name \( \theta \) of the function of interest remains unchanged in (14), so that the unpleasant properties of (1), vis-a-vis name changes has no effect here.)

The signed distance intervals of Section 4 can also be described in a transformation invariant manner, though not in a computationally simple way. Notice that the square of the signed distance \( X^2_\theta \) equals the likelihood ratio statistic for testing the hypothesis \( \eta \in C_{\theta} \) in model (2),

\[
L(y) = 2 \log \frac{f_\tilde{\eta}(y)}{f_{\tilde{\eta}(\tilde{\theta})}(y)} = X^2_{\tilde{\theta}}.
\]

After transformations (14), a standard calculation shows that the likelihood ratio statistic

\[
\tilde{L}(y) = 2 \log \left( \frac{\sup_{\tilde{\eta}} \tilde{f}_{\tilde{\eta}}(\tilde{y})}{\sup_{\tilde{\eta}:f(\tilde{\eta})=\theta} \tilde{f}_{\tilde{\eta}}(\tilde{y})} \right)
\]

still equals \( L(y) = X^2_{\tilde{\theta}} \). Therefore we can automatically recover \( X^2_{\tilde{\theta}} \) from the likelihood ratio statistic, without knowledge of the transformations (14). This fact relates to Bartlett's improvements on Wilk's likelihood ratio criterion, as we now briefly discuss.
Theorem 1 can be used to show that

\[ B \cdot X_\theta^2 \sim \chi^2_1 \quad (B^{-1} \equiv [1 + \text{tr}^2 \, d_\eta][1-2\text{tr} \, d_\eta^2]), \]  \hspace{1cm} (15)

the moments of \( B \cdot X_\theta^2 \) equaling those of \( \chi^2_1 \) through \( O(n^{-1}) \). The factor \( B \),
which is of the form \( 1 + b/n + O(n^{-3/2}) \), is called a Bartlett factor in
Barndorff-Nielsen and Cox (1983); (15) is a special case of the general theory
presented there.

We can use (15), as Bartlett intended, to set confidence intervals for \( \theta \)
in terms of the likelihood ratio statistic \( \tilde{L}(\bar{y}) = X_\theta^2 : \theta \) exists in the confidence interval if \( X_\theta^2 \geq B^{-1} \cdot X^2(1-2\alpha) \), or equivalently if \( |X_\theta| \leq z^{(1-\alpha)/B^{1/2}} \). For the example in Table 3, \( B^{-1} = [1 + 6.25/\theta^2][1-2.5/\theta^2] \). The central 90% Bartlett interval for \( \theta \) is \([3.12,6.71]\), compared to the exact answer \([2.68,6.19]\). The Bartlett interval has nearly the correct length, but is shifted about .50 units right.

The trouble here is that the Bartlett theory works with \( X_\theta^2 \), ignoring the sign of \( X_\theta \). The good properties of the signed distance intervals depend on keeping track of the side of \( C_\theta \) in which the data vector \( y \) lies, i.e. on the sign of the signed distance.

In fact it is not difficult to construct a "signed likelihood", which equals the signed distance \( X_\theta \). (Peter McCullagh, in an unpublished report, gives a confidence interval theory based on signed likelihoods, for arbitrary one parameter families \( f_\theta(y) \).) However in order to construct the intervals (7), we also need to know the terms \( \text{tr} \, d^2_\eta(\theta) \) and \( \text{tr} \, d^2_\eta(\theta) \) occurring in (6). The problem of putting the signed distance theory of Section 4 into transformation-invariant form will not be pursued here. The calculations seem to require bootstrap-like results: for example the Appendix shows that \( -\text{tr} \, d^2_\eta \neq z_0 \), definition (11). In any case we have seen that the bootstrap intervals, which capture the main aspects of the signed distance intervals, are naturally calculated in a transformation-invariant manner.

Table 5 gives central 90% confidence intervals for $\theta$, having observed $\bar{Y} \sim \theta X^2_{19}$. The exact interval, based on the $X^2_{19}$ distribution, extends 2.38 times as far to the right of the MLE $\hat{\theta}$ as to the left. The bootstrap interval (12) is not nearly skewed enough toward the right, though it is an improvement over the standard interval (1).

<table>
<thead>
<tr>
<th>Method</th>
<th>Interval</th>
<th>R/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact:</td>
<td>$[0.631, 1.88]$</td>
<td>2.38</td>
</tr>
<tr>
<td>Bootstrap:</td>
<td>$[0.580, 1.69]$</td>
<td>1.64</td>
</tr>
<tr>
<td>Standard (1):</td>
<td>$[0.466, 1.53]$</td>
<td>1.00</td>
</tr>
<tr>
<td>Bootstrap$_{E}$:</td>
<td>$[0.631, 1.88]$</td>
<td>2.38</td>
</tr>
</tbody>
</table>

Table 5. Central 90% confidence intervals for $\theta$, having observed $\bar{Y} \sim \theta X^2_{19}$, so $\hat{\theta} = \bar{Y}/19$; R/L = ratio of right side of interval (measured for $\hat{\theta}$) to left side. Nathaniel Schenker of the University of Chicago has pointed out, in an unpublished technical report, that the bootstrap intervals do not perform well here.

Using the results of Efron (1982B), one can show the existence of monotone transformations $y = h^{-1}_1(\bar{Y})$, $\eta = h^{-1}_2(\theta)$, such that to a close approximation

$$y \sim N(\eta, \sigma^2_{\eta})$$

$$\sigma_{\eta} = 1 + \varepsilon \eta$$

with $\varepsilon = .109$. Notice that if $\varepsilon = 0$ the results of Section 6 would apply and the bootstrap method would perform well. With $\varepsilon \neq 0$, it can be shown, using the results of Efron (1982B), that the term $2z_0 + z^{(\alpha)}$ in (12) should be replaced by

$$\frac{[2 - \varepsilon(z_0 + z^{(\alpha)})]z_0 + z^{(\alpha)}}{1 - \varepsilon(z_0 + z^{(\alpha)})}$$

(17)
The "Bootstrap_ε" line of Table 5 uses (17) with ε = .109, and recovers almost the exact interval.

The difficulty here lies in knowing the correct value of ε. A version of (2) including (16), say \( y \sim N_k(\eta, \sigma^2_\eta I) \), where we still want confidence intervals for a function \( \theta = t(\eta) \), can be approximately solved by analogues of Bootstrap_ε if we know the rate of change of \( \sigma_\eta \) along \( L_\hat{\eta} \) in Figure 2, \( \hat{\varepsilon} = \left. \frac{\partial}{\partial x} \sigma_{\eta+x} \right|_{x=0} \). Not making this correction results in a major error, of \( O(n^{-1/2}) \), in the bootstrap intervals (12). The bootstrap results, as illustrated in Table 5, are likely to be only a partial improvement over the standard intervals (1) (which also don't make such a correction).

If the problem just stated has been disguised by transformations \( \tilde{y} = h_1(y) \), \( \tilde{\eta} = h_2(\eta) \) as in Section 6, it seems to be quite difficult to estimate the parameter ε. In particular, it is difficult to recognize the case \( \varepsilon = 0 \), for which we know the bootstrap method gives good results. The author is currently investigating this problem. Efron (1981, 1982) gives other examples where the bootstrap intervals of Section 5 are not much of an improvement over the standard intervals (1).

To summarize this paper, (I) almost exact confidence intervals can be found for the simple class of problems introduced in Section 2, extended in Section 6; (II) within this class of problems, the bootstrap method of Section 5 improves on the standard intervals (1); (III) for the more difficult class of parametric problems discussed in this section, the bootstrap method of Section 5 is only a partial improvement over (1).
Appendix

Suppose that \( y \sim N_k(0, I) \), and that \( A = A_n(y_2) \), \( B = B_n(y_2) \) are functions of \( y_2 = (y_2, y_3, \ldots, y_k) \) depending on \( n \), of order \( O_p(n^{-3/2}) \). Let \( Q(y) = (1 + A)y_1 + B \). Then it is easy to show that \( Q \) has first four cummulants

\[
Q \sim [E(B), E(1 + A)² + Var(B), 6Cov(A, B), 12Var(A)],
\]

through \( O(n^{-1}) \). By this we mean that the computation has been carried out simply ignoring all terms \( O_p(n^{-3/2}) \) or smaller, and assuming that all relevant moments exist. Formula (A1) is useful in proving the results of Sections 3-5.

It is convenient to first verify our results making some special assumptions: that the true \( \eta \) equals 0, so \( y \sim N_k(0, I) \); that \( \theta = \tau(0) = 0 \), so \( \eta \in C_0 \); and that the level surfaces \( C_0 \) are described by the equations

\[
C_0 : \eta_1 = \theta[1 - C(\eta_2)] - [D(\eta_2) + \theta E(\eta_2)].
\]

Here \( \eta_2 = (\eta_2, \eta_3, \ldots, \eta_k) \) and

\[
C(\eta) = c'\eta_2, \quad D(\eta_2) = \eta_2', \quad d\eta_2', \quad E(\eta_2) = \eta_2', \quad e\eta_2',
\]

for \( c \) a \((k-1)\)-dimensional vector, \( d \) and \( e \) symmetric \((k-1) \times (k-1)\) matrices, with \( d \) diagonal. Later we will argue that those assumptions do not affect the final conclusions.

All of our results are asymptotic: the probability mechanism \( y \sim N_k(0, I) \) stays fixed, but the level surfaces flatten out as \( n \) goes to infinity, according to the rescaling relationships \( C_n(\eta) = \sqrt{n} C_0 \). For the specific case (A2), the level surfaces at stage \( n \) are described by

\[
C_{\eta_n} : \eta_1 = \theta_n[1 - c_n' \eta_2] - \eta_2'[d_n + \theta_n e_n] \eta_2,
\]

where \( c_n = c/\sqrt{n}, \quad d_n = d/\sqrt{n}, \quad e_n = e/\sqrt{n} \), and \( \theta_n = \sqrt{n} \theta \). (Renaming the parameter \( \theta \)
in this way makes \( C_n \) (n) cross the \( \eta \) axis at the point \((\eta, 0, 0, \ldots, 0)\), so that \( \eta \) has the same geometric meaning for all \( n \). In what follows we will denote the level surfaces as in (A2), (A3), dropping the explicit notation for \( n \) used in (A4), but remembering that in fact

\[
c = O(n^{-2}), \quad d = O(n^{-2}), \quad e = O(n^{-1}). \tag{A5}
\]

Theorem 1 refers only to the level surface \( C_0 \), which is described by

\[
\eta = -\eta_0 \quad \text{d} \eta = -\sum_{j=2}^k d_{jj} \eta_j^2,
\]

(A6)

according to (A2), (A3). For convenient notation let \( z = \hat{\eta}(0) \) indicate the nearest point to \( y \) on \( C_0 \), as in Figure 1. Call \( \Delta(z) \) the unit vector in the positive direction along \( L_z \), evaluated to be

\[
\Delta(z) \equiv (1 - 2\sum_{j=2}^k d_{jj} z_j^2, 2d_{22} z_2, \ldots, 2d_{kk} z_k)'.
\]

(A7)

Since by definition \( y = z + X_0 \cdot \Delta(z) \), we get

\[
z_j = y_j / (1 + 2d_{jj} X_0) \quad j = 2, \ldots, k.
\]

(A8)

Then the relations \( y_1 = z_1 + X_0 \Delta(z) \), and \( z_1 = -\sum_{j=2}^k d_{jj} z_j^2 \) (A6) give

\[
X_0 \equiv (1 - 2\sum_{j=2}^k d_{jj} z_j^2) y_1 + \sum_{j=2}^k d_{jj} y_j^2.
\]

(A9)

Applying (A1), with \( A = -2\sum_{j=2}^k d_{jj} y_j^2 \) and \( B = \sum_{j=2}^k d_{jj} y_j^2 \), gives Theorem 1.

The Corollary at the beginning of Section 4 follows in a similar way. Direct evaluation from (A6) shows that the curvature matrix \( d_z \) of \( C_0 \) at \( \hat{\eta}(0) = z \) equals \( d + O_p(n^{-3/2}) \). Then (6) and (A9) show that \( W_0 \) is of the form

\[
(1 + A) y_1 + B,
\]

\[
1 + A \equiv 1 - 2\sum_{j=2}^k d_{jj} y_j^2 / \left(1 - \text{tr} \ d^2\right) \quad B \equiv \frac{\sum_{j=2}^k d_{jj} (y_j^2 - 1)}{1 - \text{tr} \ d^2}.
\]

(A10)

Applying (A1) gives the Corollary.
In addition to having chosen \( \eta = 0 \), we have specified in (A2), (A3) that the unit orthogonal \( \Delta(\eta) \) to \( C_0 \) at \( \eta = 0 \) equals \( e_1 \), the first coordinate vector. The curvature matrix \( d_\eta \) at \( \eta = 0 \) has been specified to be diagonal. However none of these choices restricts the generality of our results since they can always be achieved by translations and rotations of any other \( \eta \), \( \Delta(\eta) \), and \( d_\eta \). The choice \( t(0) = 0 \) is also innocuous, since none of the results depends on the name \( \theta \) assigned to the level surfaces.

To complete the argument for Theorem 1 and its Corollary, we need to extend expression (A6) for \( C_0 \) to higher terms, say to

\[
\eta_1 = -[\eta_1^{(2)} d\eta_2^{(2)} + d^{(3)}(\eta_1^{(2)}) + d^{(4)}(\eta_1^{(2)})\ldots],
\]

(A11)

the term \( d^{(3)}(\eta_1^{(2)}) \) for instance indicating a sum of cubic monomials in \( \eta_2, \eta_3, \ldots, \eta_k \). The rescaling argument preceding (4) shows that \( d^{(3)} \) must be \( O(n^{-1}) \); in the notation there, \( d^{(3)}(\eta)(n) = d^{(3)}(\eta)/n \), and likewise \( d^{(4)}(\eta)(n) = d^{(4)}(\eta)/n^{3/2} = O(n^{-3/2}) \), etc.

The calculation leading to (A9) now gives

\[
X_0 = (1+A)y_1 + B, \quad A = -2\sum_j^k d_{jj} y_j^3,
B = \sum_j^k d_{jj} y_j^2 + d^{(3)}(y_2^{(3)}).
\]

Formula (A1) again gives Theorem 1, and the Corollary. Notice that the cubic term in \( B \) contributes nothing to \( \text{E}(B) \) because \( \text{E}(d^{(3)}(y_2^{(3)})) = 0 \), and that it adds terms of order less than \( O(n^{-1}) \) everywhere else in (A1).

The quantities \( c, d, \) and \( e \) in (A3) have simple geometric interpretations in terms of the level surfaces \( C_\theta \) defined by (A2). The unit normal vector to \( C_\theta \) at the point \((-\theta,0,0,\ldots,0)\) is

\[
\Delta(\theta) = \frac{1}{(1+\theta^2 \|c\|^2)^{1/2}} \frac{1}{\|c\|} 
\]

(A12)

so that the cosine \( \cos^2 \) of the angle between \( \Delta(\theta) \) and \( \Delta(\theta) = e_1 \) is

\[
\cos^2(\theta) = (1+\theta^2 \|c\|^2)^{-1} \approx (1-\theta^2 \|c\|^2). \quad \text{The curvature matrix of } C_\theta \text{ at } (\theta,0,\ldots,0) \text{ is evaluated to be } (d+\Theta e)(1+O(\theta^2 \|c\|^2)) = (d+\Theta e).
\]

24
Now we can verify Theorem 2. Suppose that the level surfaces \( C_\theta \) are given by (A2), (A3), and that we have observed \( y = 0 \). As in the previous arguments, these special choices do not restrict the generality of the result. The quantities \( \hat{c}, \hat{d}, \hat{e} \) appearing in (8) equal respectively \( c, d, e \) of (A3); \( \theta \) in (A2) measures distance along \( e_1 = \Delta(0) \), the normal vector to \( C_\theta \) at 0, so \( \theta \) plays the role of \( x \) in (8A), (8B). The arguments following (A12) confirm relations (8A), (8B).

Notice that for situation (A2), (A3), the MLE \( \hat{\theta} \) is

\[
\hat{\theta} = \frac{y_1 + d(y_2)}{1 - c(y_2) - e(y_2)} \pm [1 + c(y_2) + e(y_2) + c^2(y_2)][y_1 + d(y_2)], \quad (A13)
\]

and likewise under bootstrap sampling \( \hat{\theta}^* \pm [1 + c(y^*_2) + e(y^*_2) + c^2(y^*_2)][y_1 + d(y^*_2)] \).

Using (A1) again, applied to the bootstrap distribution \( y^* \sim N_k(0, I) \), gives the first four cumulants of \( \hat{\theta}^* \) to be

\[
\hat{\theta}^* \sim [\text{tr } d, 1 + 2\text{tr } d^2 + 2\text{tr } e + 3\|c\|^2, 0, 12\|c\|^2]. \quad (A14)
\]

Theorem 2 follows from standard Edgeworth, Cornish-Fisher expansions, Kendall and Stuart (1958), used to evaluate the terms in (11), (12) from the approximate cumulants (A14). For example \( \text{Prob}_x(\hat{\theta}^* < \hat{\theta} = 0) \equiv \Phi(1 - \text{tr } d) \), so

\[
z_0 = -\text{tr } d. \quad (A15)
\]

To understand approximation (8), suppose first that \( d = e = 0 \) in (A3), so that the \( C_\theta \) are straight lines. Having observed \( y = 0 \),

\[
W_\theta = -\theta \text{ Co}(\theta) \pm -\theta \left( 1 - \frac{\|c\|^2}{2} \theta^2 \right). \quad (A16)
\]

The interval limits \( W_\theta = \pm z(\alpha) \) are given by \( \theta \pm z(\alpha)(1 - \frac{\|c\|^2}{2} z(\alpha)^2 - 1) \) as in (8). Conversely suppose that \( c = 0 \) in (A3), but \( d \) and \( e \neq 0 \). Then \( X_\theta = -\theta \),
$W_0 = [-\theta - \text{tr}(d+\partial \theta)]/[1 - \text{tr} d^2]$. The interval limits $W_0 = \pm z(\alpha)$ are given by 
$\theta = -\text{tr} d \pm (1 - \text{tr} d^2 - \text{tr} e)z(\alpha)$, again as in (8). Approximation (8) is the result of adding the effects of $e$ (changing curvature of $C_\theta$) and $c$ (rotation of $C_\theta$).

Finally, it is not difficult to see why the bootstrap intervals are invariant under transformations (14). Following through the definitions shows that 
$\hat{\theta}^*(b) = \hat{\epsilon}(\tilde{\eta}^{\text{MAX}^*}(b)) \text{ where } \tilde{\eta}^{\text{MAX}^*}(b) = h_2(\eta^*(b)) = h_2(y^*(b))$, and also that 
$\hat{\epsilon}(\tilde{\eta}) = t(h_2^{-1}(\eta))$. Therefore $\hat{\theta}^*(b) = t(y^*(b))$ so that the bootstrap replications of $\hat{\theta}$ will be the same whether or not transformations (14) have been made. Then $\hat{G}(s)$ and hence the bootstrap interval itself will be the same.

REFERENCES


Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. SIAM-CBMS 38.


