Biometrika (1985), **72**, 1, *pp*. 45–58
Printed in Great Britain

Bootstrap confidence intervals for a class of parametric problems

BY BRADLEY EFRON

Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.

SUMMARY

We consider the following class of problems: having observed a multivariate normal data vector y with unknown mean vector η, covariance matrix the identity, find an approximate confidence interval for $\theta = t(\eta)$, a real-valued function of η. A simple geometric construction is given which leads to highly accurate solutions. This construction shows that the standard approximation based on maximum likelihood theory, $\hat{\theta} \pm \hat{\sigma} z^{(\alpha)}$, can be quite misleading when θ is nonlinear in η. We discuss bootstrap-based confidence intervals which remove most of the error in the standard approximation, at the expense of considerably more calculation. The bootstrap intervals are invariant under transformation of both y and η, and so they automatically produce accurate solutions in problems which can be transformed to multivariate normality, without requiring knowledge of the normalizing transformation.

Some key words: Bootstrap; Computer-intensive method; Confidence interval; Higher-order asymptotics; Maximum likelihood; Nonlinear regression.

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 η, we wish to set confidence intervals for a real-valued function of η, say $\theta = t(\eta)$. For example we might have y bivariate normal with mean vector η and covariance matrix the identity, and desire confidence intervals for $\theta = \eta_1 \eta_2$.

Exact confidence intervals for θ 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,

$$\theta \in \hat{\theta} \pm \hat{\sigma} z^{(\alpha)}, \tag{1}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ, $\hat{\sigma}$ an estimate of its standard deviation, usually based on the Fisher information matrix, and $z^{(\alpha)}$ the α 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 §§ 3 and 4 is quite noticeable. Changing the parameter of interest from θ to θ^2 makes the comparison much worse.

The bias-corrected percentile method is another automatic algorithm for producing confidence intervals for θ a real-valued parameter. It is based on the bootstrap (Efron, 1979, 1982a), and can be applied to either parametric or nonparametric situations. In the present paper we discuss only its use in parametric families $f_\eta(y)$. By implication, however, its success here is intended to support its credibility in nonparametric situations, where it is more likely to be actually used. This argument is by no means conclusive; see the cautionary note at the end of § 6.

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.

Table 1. *Central 90% confidence intervals for $\theta = \eta_1 \eta_2$ and for $\phi = \theta^2$, having observed bivariate normal vector $y = (2, 4)$*

	For θ	For ϕ
Standard intervals (1)	[0·64, 15·36]	[−53·7, 181·7]
Almost exact intervals	[1·77, 17·03]	[3·1, 290·0]
Bootstrap intervals	[1·77, 17·12]	[3·1, 293·1]

The almost exact intervals are based on the signed distance theory of §§ 3 and 4. The bootstrap intervals are the bias-corrected percentile intervals of § 5.

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 θ: where $f_\eta(y)$ is the multivariate normal family $N_k(\eta, I)$ and θ is a real-valued function of η. These intervals are useful in their own right, and are closely related to Bartlett's improvements on the likelihood ratio method; see § 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 thrust of the paper.

A final point is made in § 6: because the bootstrap intervals are calculated in a transformation-invariant way, they automatically produce accurate intervals in any family $f_\eta(y)$ which can be transformed to multivariate normality, with constant covariance matrix, by mappings of y and η. No knowledge of the normalizing transformation is required of the statistician.

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 η and covariance matrix the identity,

$$y : N_k(\eta, I), \tag{2}$$

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

$$\mathscr{C}_\theta = \{\eta : t(\eta) = \theta\},$$

are assumed to be smooth, the function $t(\eta)$ having continuous second partial derivatives. In § 6 we will see how the family (2) can be considerably generalized.

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 $\mathscr{C}_\theta$ are straight lines passing through the origin.

Example 2, noncentral $\chi_6 : k = 6, \theta = \| \eta \|$. In this case the level surfaces $\mathscr{C}_\theta$ are spheres of radius θ centred at the origin.

Example 3, product of means: $k = 2, \theta = \eta_1 \eta_2$. This is the example discussed in § 1. The level surfaces are hyperbolae in the plane. Efron (1984) 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 η. The point on any particular $\mathscr{C}_\theta$ nearest to y in Euclidean distance is labelled $\hat{\eta}(\theta)$, it being the restricted maximum likelihood estimator for η assuming $\eta \in \mathscr{C}_\theta$. The line through $\hat{\eta}(\theta)$ orthogonal to $\mathscr{C}_\theta$ is labelled $\mathscr{L}_{\hat{\eta}(\theta)}$. We think of this as a one-dimensional axis, with origin at $\hat{\eta}(\theta)$; x will indicate signed distance along $\mathscr{L}_{\hat{\eta}(\theta)}$. It does not 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 $\mathscr{C}_\theta$, as indicated by the arrowhead in Fig. 1.

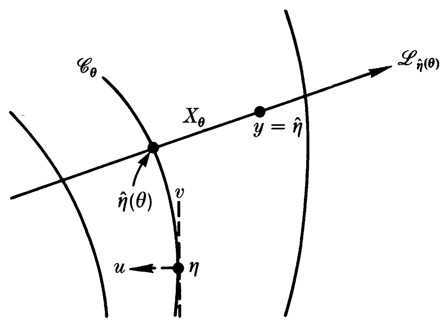

Fig. 1. Schematic illustration shows the level surfaces $\mathscr{C}_\theta$ and the data vector y. Nearest point to y on a particular surface $\mathscr{C}_\theta$, $\hat{\eta}(\theta)$; line through $\hat{\eta}(\theta)$ orthogonal to $\mathscr{C}_\theta$, $\mathscr{L}_{\hat{\eta}(\theta)}$. Signed distance of y from $\hat{\eta}(\theta)$ along $\mathscr{L}_{\hat{\eta}(\theta)}$, X_θ.

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

The shift of X_θ can be quite dramatic. For Example 2, noncentral χ_6, with $\theta = 5$, the theorem of § 3 gives X_θ approximately $N(0{\cdot}5, 0{\cdot}95^2)$. Table 2 shows this approximation to be accurate. In terms of Fig. 1, the surface $\mathscr{C}_\theta$ containing the true η is a 6-dimensional sphere of radius 5; X_θ, taken positive or negative as y is outside or inside of $\mathscr{C}_\theta$, is almost perfectly normal, but with distribution centred half a standard deviation outside $\mathscr{C}_\theta$.

Table 2. *Percentiles of the distribution of X_θ in Example 2, $\theta = 5$; $y \sim N_6(\eta, I)$, with η on a sphere of radius 5; X_θ 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*

	0·025	0·05	0·10	0·25	0·75	0·90	0·95	0·975
Theoretical percentiles $0{\cdot}5+0{\cdot}95\,z^{(\alpha)}$	−1·36	−1·06	−0·72	−0·14	1·14	1·72	2·06	2·36
Actual percentiles	−1·36	−1·08	−0·73	−0·16	1·14	1·73	2·07	2·38

The theorem in §3 gives X_θ approximately $N(0{\cdot}5, 0{\cdot}95^2)$. This agrees well with the actual distribution of X_θ, determined from 100,000 Monte Carlo simulations. The standard errors of these numbers are about 0·01.

3. Distribution of the signed distance

This section presents a theorem on the approximate normality of the signed distance X_θ. First we need a quantitative description of the curvature of the level surfaces $\mathscr{C}_\theta$ at a point η.

Let $\mathscr{T}_\eta$ be the tangent hyperplane to $\mathscr{C}_\theta$ at a point η. We can choose any basis of $k-1$ orthogonal unit vectors in $\mathscr{T}_\eta$, and let $v = (v_1, \ldots, v_{k-1})$ be the coordinates of a point in $\mathscr{T}_\eta$ with respect to this basis, $v = 0$ corresponding to the point η. Also let u indicate distance orthogonal to $\mathscr{T}_\eta$, 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_θ. This convention is illustrated in Fig. 1.

Since $\mathscr{C}_\theta$ is tangent to $\mathscr{T}_\eta$ at η, the Taylor series describing $\mathscr{C}_\theta$ for points near η, that is for v near 0, begins

$$u = v' d_\eta v \tag{3}$$

for some symmetric $(k-1)\times(k-1)$ matrix d_η, not necessarily positive-definite. For $\mathscr{C}_\theta$ the 6-dimensional sphere of radius 5, considered in Table 2, $d_\eta = 0{\cdot}1I$. The magnitude of the elements of d_η measures the curvature of $\mathscr{C}_\theta$ at η, in a way made more precise later. If $\mathscr{C}_\theta$ is flat, as in Example 1, then $d_\eta = 0$. For dimension $k = 2$, d_η equals one-half the usual definition of curvature for $\mathscr{C}_\theta$ at point η. We call d_η simply the curvature matrix of $\mathscr{C}_\theta$ at point η.

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

Rescaling the sufficient vector $\bar{y}$ to $y = \bar{y}\sqrt{n}$ with distribution $N_k\{\eta\sqrt{n}, I\}$ restores its covariance matrix to I. Figure 1 applies again exactly as shown, except that we have to remember that every parameter vector η has been mapped into $\eta(n) = \eta\sqrt{n}$. The level surfaces $\mathscr{C}_\theta$ map into $\mathscr{C}_\theta(n) \equiv \mathscr{C}_\theta\sqrt{n}$, 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 $\mathscr{C}_\theta(n)$ near $\eta(n)$. The theorem which follows can be interpreted as saying that

$$X_\theta \equiv \bar{X}_\theta\sqrt{n} \tag{4}$$

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

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 I. This curvature matrix is called simply d_η, rather than $d_{\eta(n)}(n)$, in what follows. The elements of d_η are assumed to be of magnitude $O(n^{-\frac{1}{2}})$ because of the rescaling argument above.

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

Theorem 1. *If $\eta \in \mathscr{C}_\theta$, then the signed distance X_θ is asymptotically normal with first four cumulants*

$$[\mathrm{tr}\,(d_\eta), \{1-\mathrm{tr}\,(d^2_\eta)\}^2, 0, 0], \tag{5}$$

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

For any $(k-1)\times(k-1)$ orthogonal matrix Γ, the trace satisfies $\mathrm{tr}\,(\Gamma d_\eta \Gamma') = \mathrm{tr}\,(d_\eta)$ and $\mathrm{tr}\,(\Gamma d^2_\eta \Gamma') = \mathrm{tr}\,(d^2_\eta)$, so that the terms of (5) do not depend on the choice of basis in $\mathscr{T}_\eta$. Geometrically, $2\,\mathrm{tr}\,(d_\eta)$ is the sum of the usual curvatures of $\mathscr{C}_\theta$ in the $k-1$ orthogonal directions through η.

The expectation $\mathrm{tr}\,(d_\eta)$ in (5) is $O(n^{-\frac{1}{2}})$. This is the main term disturbing X_θ from its asymptotic $N(0,1)$ distribution. The standard deviation $1-\mathrm{tr}\,(d^2_\eta)$ differs from 1 by only $O(n^{-1})$. Notation $\sim$ and $\doteq$ will indicate accuracy to $O(n^{-1})$, with error $O(n^{-3/2})$. In the example of Table 2, $\mathrm{tr}\,(d_\eta) = 0{\cdot}50$, $1-\mathrm{tr}\,(d^2_\eta) = 0{\cdot}950$. The fact that the third and fourth cumulants $\doteq 0$ accounts for the impressive accuracy of the normal approximation seen in Table 2. Another example is given in §4.

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

The proof of Theorem 1 and other results in the Appendix, is formal in nature, and does not provide error bounds for the approximation. The accuracy of the Theorem breaks down in regions of high curvature of $\mathscr{C}_\theta$. Numerical experimentation for $\mathscr{C}_\theta$ a circle in the plane showed good accuracy when the radius of the circle was $\geqslant 3$, and reasonable accuracy even for radius as small as 2 except in the lower tail of X_θ.

4. Confidence intervals for θ

The near normality of the signed distance X_θ can be used to construct accurate confidence intervals for θ. Theorem 1 suggests that the normalized signed distance W_θ should be nearly unit normal,

$$W_\theta \equiv \frac{X_\theta - \mathrm{tr}\,(d_{\hat\eta(\theta)})}{1-\mathrm{tr}\,(d^2_{\hat\eta(\theta)})}, \tag{6}$$

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

Corollary. *If $\eta \in \mathscr{C}_\theta$ then the normalized signed distance W_θ is asymptotically normal with first four cumulants $W_\theta \sim \{0,1,0,0\}$, to $O(n^{-1})$.*

Table 3. *Actual distribution of W_θ in Example 3, compared to standard normal distribution*

	$\alpha = 0{\cdot}025$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}90$	$\alpha = 0{\cdot}95$	$\alpha = 0{\cdot}975$
Percentiles of W_θ	$-1{\cdot}94$	$-1{\cdot}63$	$-1{\cdot}28$	$1{\cdot}29$	$1{\cdot}66$	$1{\cdot}98$
Percentiles $z^{(\alpha)}$	$-1{\cdot}96$	$-1{\cdot}64$	$-1{\cdot}28$	$1{\cdot}28$	$1{\cdot}64$	$1{\cdot}96$

The proof appears in the Appendix. For Example 3, $\theta = \eta_1 \eta_2$, with $\eta = (2, 4)$, 100,000 Monte Carlo replications gave the comparison between the actual distribution of W_θ and its limiting $N(0, 1)$ distribution shown in Table 3.

Figure 2 illustrates how W_θ is used to test the hypothesis $\eta \in \mathscr{C}_\theta$. According to the Corollary, and a straightforward application of Edgeworth expansions, the interval

$$z^{(\alpha)} \leqslant W_\theta \leqslant z^{(1-\alpha)} \tag{7}$$

is an acceptance region of size $1 - 2\alpha + O(n^{-3/2})$ for the hypothesis $\eta \in \mathscr{C}_\theta$. If $\alpha = 0{\cdot}05$ for instance, then $-1{\cdot}645 \leqslant W_\theta \leqslant 1{\cdot}645$ tests $\eta \in \mathscr{C}_\theta$ at level $1 - 2\alpha \doteq 0{\cdot}90$, with probability of error $\doteq 0{\cdot}05$ on each side of (7).

Relationship (7) can be expressed as

$$\operatorname{tr}(d_{\hat\eta(\theta)}) + \{1 - \operatorname{tr}(d^2_{\hat\eta(\theta)})\}\, z^{(\alpha)} \leqslant X_\theta \leqslant \operatorname{tr}(d_{\hat\eta(\theta)}) + \{1 - \operatorname{tr}(d^2_{\hat\eta(\theta)})\}\, z^{(1-\alpha)},$$

which shows that the acceptance region is shifted in the positive direction from $\mathscr{C}_\theta$ along each line $\mathscr{L}_{\hat\eta(\theta)}$, by amount $\operatorname{tr}(d_{\hat\eta(\theta)})$. The more curved $\mathscr{C}_\theta$ is at $\hat\eta(\theta)$, the larger the shift.

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

Table 4 applies to Example 2, noncentral χ_6, supposing that we have observed a data vector y with $\| y \| = 5$. Exact confidence limits for θ were obtained in the usual way using the noncentral chi-squared distribution. For instance the $0{\cdot}05$ limit point $2{\cdot}68$ was obtained from $\operatorname{pr}\{\chi_6^2(2{\cdot}68^2) < 5^2\} = 0{\cdot}95$. The signed distance confidence limits obtained from (7) agree well with the exact results.

Let points on $\mathscr{L}_{\hat\eta}$ be denoted by $\hat\eta + x\hat\Delta$, where $\hat\Delta$ is a unit vector pointing in the positive direction along $\mathscr{L}_{\hat\eta}$. In Fig. 2 we have indicated the x values of the points on $\mathscr{L}_{\hat\eta}$

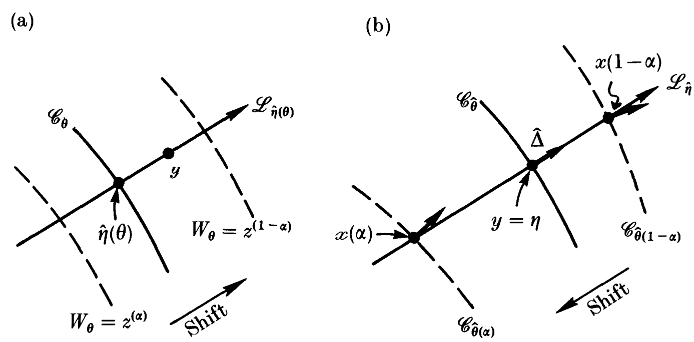

Fig. 2. Test based on the normalized signed distance accepts hypothesis $\eta \in \mathscr{C}_\theta$ for y in band about $\mathscr{C}_\theta$, as indicated in (a). Band shifted in positive direction away from $\mathscr{C}_\theta$ by amount depending upon curvature of $\mathscr{C}_\theta$. Corresponding confidence interval for θ, having observed $y = \hat\eta$, shifted in the negative direction from $\mathscr{C}_{\hat\theta}$, as indicated in (b).

Table 4. *Confidence intervals for Example* 2, *noncentral* χ_6, *having observed* $\| y \| = 5$

	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}25$	$\alpha = 0{\cdot}75$	$\alpha = 0{\cdot}90$	$\alpha = 0{\cdot}95$
Exact limits	2·68	3·08	3·75	5·18	5·82	6·19
Signed distance limits (7)	2·71	3·08	3·71	5·16	5·80	6·19
Approximate limits (8)	2·77	3·15	3·79	5·21	5·85	6·23
Approximate bootstrap limits (13)	2·94	3·28	3·86	5·14	5·75	6·06
Standard limits (1)	3·36	3·72	4·33	5·67	6·28	6·64

delimiting the confidence interval for θ 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 §5:

$$x(\alpha) \doteq \frac{-\operatorname{tr}(\hat{d}) + \{1 - \operatorname{tr}(\hat{d}^2)\}\, z^{(\alpha)}}{1 - \frac{1}{2}\, \| \hat{c} \|^2 z^{(\alpha)2}} - \operatorname{tr}(\hat{e})\, z^{(\alpha)}. \tag{8}$$

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

Denote by $\mathscr{C}(x)$ the level surface $\mathscr{C}_\theta$ intersecting $\mathscr{L}_{\hat{\eta}}$ at $\hat{\eta} + x\hat{\Delta}$. At $x = 0$, $\mathscr{C}(0) = \mathscr{C}_{\hat{\theta}}$ is described locally by $u = v'\hat{d}v$ as in (3), with $\hat{d} \equiv d_{\hat{\eta}}$. The $(k-1) \times (k-1)$ matrix $\hat{e}$ is the first derivative of the curvature matrix of $\mathscr{C}(x)$ as x moves away from 0,

$$\hat{e} = \left[\frac{\partial}{\partial x} d_{\hat{\eta} + x\hat{\Delta}}\right]_{x=0}. \tag{8A}$$

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

$$\operatorname{tr}(d_{\hat{\eta} + x\hat{\Delta}}) \doteq \operatorname{tr}(\hat{d}) + x \operatorname{tr}(\hat{e}).$$

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

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

$$\operatorname{co}^2(x) \doteq 1 - \| \hat{c} \|^2 x^2. \tag{8B}$$

In other words, $\| \hat{c} \|$ measures the rate of rotation of the level surfaces near the point $y = \hat{\eta}$. The quantity $\| \hat{c} \|^2$, like $\operatorname{tr}(\hat{e})$, is $O(n^{-1})$.

For the noncentral χ_6 example it is easy to calculate $\hat{d} = (2 \| y \|)^{-1} I$, $\hat{e} = -2\hat{d}^2$, and $\| \hat{c} \|^2 = 0$. The third line of Table 4 shows approximation (8) performing well.

In Example 1, ratio estimation, $d_\eta = 0$ since the $\mathscr{C}_\theta$ are straight lines. In this case $W_\theta = X_\theta$ has the distribution $N(0, 1)$ exactly, and the W_θ intervals based on inverting (7) agree exactly with those of Fieller (1954),

$$\theta \in \frac{\hat{\theta} \pm \{\hat{\delta}^2(1 - \hat{\delta}^2 + \hat{\theta}^2)\}^{\frac{1}{2}}}{1 - \hat{\delta}^2}, \quad \hat{\delta}^2 \equiv (z^{(\alpha)}/y_1)^2. \tag{9}$$

Letting $r = z^{(1-\alpha)}/\| y \|$, approximation (8) turns out to be

$$\theta \in \left[\tan\left\{\tan^{-1}\hat{\theta} - \tan^{-1}\left(\frac{r}{1 - \frac{1}{2}r^2}\right)\right\},\ \tan\left\{\tan^{-1}\hat{\theta} + \tan^{-1}\left(\frac{r}{1 - \frac{1}{2}r^2}\right)\right\}\right], \tag{10}$$

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 that $\hat{\theta} = y_2/y_1 = 1$, then the exact central 90%, $\alpha = 0{\cdot}05$, interval for θ based on (9) is $[0{\cdot}408, 2{\cdot}452]$, compared to $[0{\cdot}409, 2{\cdot}443]$ from (10).

To summarize this section, the normalized signed distance W_θ is an approximate normal pivotal for θ. Because of this, 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. Section 6 discusses their close relationship to standard likelihood ratio tests. Formula (8), which is based on the local geometry of the surfaces $\mathscr{C}_\theta$ near the point $y = \hat{\eta}$, 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, 1982a), applied to the simple class of parametric problems introduced in § 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 § 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 $\mathscr{C}_\theta$; it does not depend on the name θ attached to the level surfaces $\mathscr{C}_\theta$, and so cannot be misled by transformations like the change from θ to $\phi = \theta^2$ in § 1.

The bias-corrected percentile method, applied to the maximum likelihood estimate $\hat{\theta} = t(\hat{\eta}) = t(y)$, constructs approximate confidence intervals for θ in the following way: (i) the probability mechanism generating the data is estimated by maximum likelihood, which in this case means estimating η by $\hat{\eta} = y$; (ii) bootstrap data vectors $y^*(1), \ldots, y^*(B)$ are obtained by independently identically distributed sampling from $f_{\hat{\eta}}(.)$; (iii) the corresponding bootstrap maximum likelihood estimates $\hat{\theta}^*(b) = t\{y^*(b)\}$ are calculated, $b = 1, \ldots, B$, giving cumulative distribution function

$$\hat{G}(s) \equiv \text{card}\,\{\hat{\theta}^*(b) < s\}/B;$$

(iv) the quantity

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

is calculated, Φ being the standard normal cumulative; (v) finally, the central $1-2\alpha$ interval for θ 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 θ. Efron (1981, 1982a) and Buckland (1983) discuss other ways of forming confidence intervals from $\hat{G}$.

Notice that if $\hat{G}(\hat{\theta}) = \frac{1}{2}$ then $z_0 = 0$ and the interval (12) is $[\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1-\alpha)]$, the α and $1-\alpha$ percentiles of the bootstrap distribution. If $\hat{G}(\hat{\theta}) \neq \frac{1}{2}$ then the z_0 term compensates for the bias of $\hat{\theta}$ as an estimator of θ, as motivated in § 10·7 of Efron (1982a).

Step (ii) of the bootstrap algorithm can be, but does not have to be, carried out by Monte Carlo. In this paper we are considering only parametric applications of the bootstrap, so that the bootstrap cumulative distribution $\hat{G}$ can be approximated by a variety of familiar parametric techniques. Efron (1984), for example, used Edgeworth expansions to approximate $\hat{G}$. 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 $\mathscr{L}_{\hat{\eta}}$, using the notation in Fig. 2b.

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

$$x_B(\alpha) \doteq \frac{-\operatorname{tr}(\hat{d}) + \{1 - \operatorname{tr}(\hat{d}^2)\} z^{(\alpha)}}{1 - \frac{1}{2} \| \hat{c} \|^2 z^{(\alpha)2}} + \{2 \operatorname{tr}(\hat{d}^2) + \operatorname{tr}(\hat{e})\} z^{(\alpha)} \tag{13}$$

to $O(n^{-1})$, *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 $O(n^{-\frac{1}{2}})$ term, $\operatorname{tr}(\hat{d})$, and a portion of the $O(n^{-1})$ term, the denominator $1 - \frac{1}{2} \| \hat{c} \|^2 z^{(\alpha)2}$, which comes from the rotation of the level surfaces near $\hat{\eta} = y$. They err by $O(n^{-1})$ in the term $\operatorname{tr}(\hat{e}) z^{(\alpha)}$, which relates to the change in curvature of the level surfaces near $\hat{\eta} = y$. The fourth line of Table 4 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 θ attached to the level surfaces $\mathscr{C}_\theta$, and not only upon their geometric shapes. The comparison is easiest if the function $t(\hat{\eta} + x\hat{\Delta})$ is linear in x, that is if the name θ is linear in distance along $\mathscr{L}_{\hat{\eta}}$. Then the standard intervals (1) are delimited on $\mathscr{L}_{\hat{\eta}}$ 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 $O(n^{-\frac{1}{2}})$. The standard method is correct to first order but not to second order. The bootstrap intervals are correct to second order but not to third order.

In Example 2, $t(\hat{\eta} + x\hat{\Delta})$ is linear in x. The large effect of the $O(n^{-\frac{1}{2}})$ error is apparent in the fifth line of Table 4. Changing the name of the level surfaces, for example considering $\phi = \theta^2$ rather than θ, adds another error of $O(n^{-\frac{1}{2}})$ to (1) as compared to (8). As we saw in § 1, the naming error can be enormous.

Table 5. *Confidence intervals for Example* 3 ($\theta = \eta_1 \eta_2$), *having observed* $y = (2, 4)$

	$\alpha = 0{\cdot}95$	$\alpha = 0{\cdot}90$	$\alpha = 0{\cdot}10$	$\alpha = 0{\cdot}05$
Signed distance limits	1·77	3·19	14·87	17·03
Bootstrap limits	1·77	3·10	14·97	17·12
Standard limits	0·64	2·27	13·73	15·36

Table 5 compares, for Example 3, the almost exact signed distance intervals of § 4 obtained from either (7) or (8) with the bootstrap intervals and also with the standard intervals. The bootstrap intervals match better the almost exact answers here than in Table 4, because the $\operatorname{tr}(\hat{e})$ 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 Creasy's (1954) fiducial solution rather than Fieller's solution for $\theta = \eta_2/\eta_1$.

Robison (1964) discusses a large class of problems where the level surface $\mathscr{C}_\theta$ are linear;

in such problems we expect the bootstrap intervals to closely match the exact answers, since they are accurate to $O(n^{-1})$.

Suppose that instead of (2) we observe y distributed as $N_k(\eta, \sigma^2 I)$, with σ^2 unknown but estimated by $\bar{\sigma}^2$ distributed as $\sigma^2 \chi_m^2/m$ independent of y. If actually $y = \bar{y}\sqrt{n}$ as in the repeated sampling argument leading up to (4), then $m = k(n-1)$. For convenient discussion here, suppose that $\bar{\sigma} = 1$, which can always be achieved by rescaling the original coordinates. The signed distance theory of §4 is modified in the obvious way for this situation, for example replacing $z^{(\alpha)}$ by $t_m^{(\alpha)}$ in (8), where $t_m^{(\alpha)}$ is the α point of Student's t_m distribution. The bootstrap algorithm can still be carried out, by bootstrap sampling from $N_k(\bar{y}, \bar{\sigma}^2 I)$ at step (ii), 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 α in (12) by α_m, defined by $z^{(\alpha_m)} = t_m^{(\alpha)}$, for example if $\alpha = 0{\cdot}05$, $m = 20$, then $\alpha_m = 0{\cdot}042$. This, according to (13), corrects the bootstrap intervals. However in the more complicated situation of §6 the proper choice of m is not likely to be obvious, so that this correction cannot be made.

6. Invariance under transformations

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 §2. This section extends that class considerably. These results are closely related to Bartlett's improvements on the likelihood ratio test (Barndorff-Nielsen & Cox, 1984) as described later.

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

$$\tilde{y} = h_1(y), \quad \tilde{\eta} = h_2(\eta), \tag{14}$$

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 $\tilde{f}_{\tilde{\eta}}(\tilde{y})$, 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 θ. Unfortunately, it may be practically impossible to discover h_1 and h_2.

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

We need to say what the 'bootstrap method' is for a general parametric family $\tilde{f}_{\tilde{\eta}}(.)$. Let $\tilde{\eta}^{\max}$ indicate the maximum likelihood estimator of $\tilde{\eta}$, and generate bootstrap data vectors $\tilde{y}^*(1), \ldots, \tilde{y}^*(B)$ by independent and identical sampling from $\tilde{f}_{\tilde{\eta}^{\max}}$. This gives corresponding bootstrap maximum likelihood estimators for θ, say $\hat{\theta}^*(b)$ based on $\tilde{y}^*(b)$. Now proceed as in steps (iii), (iv), (v) of the algorithm described in §5, leading to the interval (12), which is what we mean here by the bootstrap interval for θ.

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

The signed distance intervals of §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_θ equals the likelihood ratio statistic for testing the hypothesis $\eta \in \mathscr{C}_\theta$ in model (2),

$$L(y) \equiv 2\log\{f_{\hat{\eta}}(y)/f_{\hat{\eta}(\theta)}(y)\} = X_\theta^2.$$

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

$$\tilde{L}(\tilde{y}) \equiv 2\log\{\sup_{\tilde{\eta}} \tilde{f}_{\tilde{\eta}}(\tilde{y}) / \sup_{\tilde{\eta}:\tilde{t}(\tilde{\eta})=\theta} \tilde{f}_{\tilde{\eta}}(\tilde{y})\}$$

still equals $L(y) = X_\theta^2$. Therefore we can automatically recover X_θ^2 from the likelihood ratio statistic, without knowledge of the transformations (14). This fact relates to Bartlett's improvements on Wilks's likelihood ratio criterion, as we now briefly discuss.

Theorem 1 can be used to show that

$$B_\eta X_\theta^2 \sim \chi_1^2, \quad B_\eta^{-1} \equiv \{1+\mathrm{tr}^2(d_\eta)\}\{1-2\,\mathrm{tr}\,(d_\eta^2)\}, \tag{15}$$

the moments of $B_\eta X_\theta^2$ equalling those of χ_1^2 to $O(n^{-1})$. The factor B_η, which is of the form $1+b_\eta/n+O(n^{-3/2})$, is called a Bartlett factor by Barndorff-Nielsen & Cox (1984); (15) is a special case of the general theory presented there.

We can use (15), as Bartlett intended, to set confidence intervals for θ in terms of the likelihood ratio statistic $\tilde{L}(\tilde{y}) = X_\theta^2$: θ exists in the confidence interval if $X_\theta^2 \leqslant B_{\hat{\eta}}^{-1}\chi_1^{2(1-2\alpha)}$, or equivalently if $|X_\theta| \leqslant z^{(1-\alpha)}/B_{\hat{\eta}}^{1/2}$. For the example in Table 4, $B_\eta^{-1} = \{1+6{\cdot}25/\theta^2\}\{1-2{\cdot}5/\theta^2\}$. The central 90% Bartlett interval for θ is [3·12, 6·71], compared with the exact answer [2·68, 6·19]. The Bartlett interval has nearly the correct length, but is shifted about 0·50 units right.

The trouble here is that Bartlett's theory works with X_θ^2, ignoring the sign of X_θ. The good properties of the signed distance intervals depend on keeping track of the side of $\mathscr{C}_\theta$ 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_θ. McCullagh (1984) 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 $\mathrm{tr}\,(d_{\hat{\eta}(\theta)})$ and $\mathrm{tr}\,(d_{\hat{\eta}(\theta)}^2)$ occurring in (6). The problem of putting the signed distance theory of §4 into transformation-invariant form will not be pursued here. The calculations seem to require bootstrap-like results: for example the Appendix shows that $-\mathrm{tr}\,(d_{\hat{\eta}}) \doteqdot z_0$, 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.

To summarize this paper, almost exact confidence intervals, useful in their own right, can be found for the simple class of problems introduced in §2 and extended in §6. Within this class, the bootstrap method of §5 improves on the standard intervals (1), eliminating the error term of $O(n^{-\frac{1}{2}})$. This improvement gives some theoretical support to the use of the bootstrap method for more general confidence interval problems.

However it is easy, using the theory developed by Efron (1982b) to describe families $f_\eta(y)$ which cannot be transformed even approximately to the form of § 2, and for which the bootstrap method is only a partial improvement over (1). Current research involves a modification of definition (12) which makes the bootstrap intervals correct to $O(n^{-\frac{1}{2}})$ for general parametric families.

APPENDIX

Proofs of the main results

Suppose that y has the distribution $N_k(0, I)$, and that $A \equiv A_n(y_{(2)}), B \equiv B_n(y_{(2)})$ are functions of $y_{(2)} = (y_2, \ldots, y_k)'$ depending on n, of stochastic order $O_p(n^{-\frac{1}{2}})$. Let $Q(y) \equiv (1+A)\, y_1 + B$. Then it is easy to show that Q has first four cumulants

$$Q \sim \{E(B), E(1+A)^2 + \operatorname{var}(B), 6 \operatorname{cov}(A, B), 12 \operatorname{var}(A)\}, \tag{A1}$$

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

It is convenient to first verify our results making some special assumptions: that the true η equals 0, so $y : N_k(0, I)$; that $\theta = t(0) = 0$, so that $\eta \in \mathscr{C}_0$; and that the level surfaces $\mathscr{C}_\theta$ are described by the equations

$$\mathscr{C}_\theta : \eta_1 = \theta\{1 - C(\eta_{(2)})\} - \{D(\eta_{(2)}) + \theta E(\eta_{(2)})\}. \tag{A2}$$

Here $\eta_{(2)} = (\eta_2, \ldots, \eta_k)'$ and

$$C(\eta) = c'\eta_{(2)}, \quad D(\eta_{(2)}) = \eta'_{(2)} d\eta_{(2)}, \quad E(\eta_{(2)}) = \eta'_{(2)} e\eta_{(2)}, \tag{A3}$$

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 these assumptions do not affect the final conclusions.

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

$$\mathscr{C}_{\theta_n}(n) : \eta_1 = \theta_n(1 - c'_n \eta_{(2)}) - \eta'_{(2)}(d_n + \theta_n e_n)\, \eta_{(2)}, \tag{A4}$$

where $c_n = c/\sqrt{n}$, $d_n = d/\sqrt{n}$, $e_n = e/n$ and $\theta_n = \sqrt{n}\theta$. Renaming the parameter θ in this way makes $\mathscr{C}_{\theta_n}(n)$ cross the η_1 axis at the point $(\theta_n, 0, \ldots, 0)$, so that θ_n has the same geometric meaning for all n. In what follows we denote the level surfaces as in (A2), (A3), dropping the explicit notation for n used in (A4), but remembering that in fact

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

Theorem 1 refers only to the level surface $\mathscr{C}_0$, which is described by

$$\eta_1 = -\eta'_{(2)} d\eta_{(2)} = -\sum_{j=2}^{k} d_{jj} \eta_j^2, \tag{A6}$$

according to (A2), (A3). For convenient notation let $z = \hat{\eta}(0)$ indicate the nearest point to y on $\mathscr{C}_0$, as in Fig. 1. Call $\Delta(z)$ the unit vector in the positive direction along $\mathscr{L}_z$, evaluated to be

$$\Delta(z) \doteq \left(1 - 2\sum_{j=2}^{k} d_{jj}^2 z_j^2, 2d_{22}\, z_2, \ldots, 2d_{kk}\, z_k\right)'. \tag{A7}$$

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

$$z_j \doteq y_j/(1 + 2d_{jj} X_0) \quad (j = 2, \ldots, k). \tag{A8}$$

Then the relations

$$y_1 = z_1 + X_0\, \Delta(z), \quad z_1 = -\sum_{j=2}^{k} d_{jj} z_j^2,$$

(A6), give

$$X_0 \doteq (1-2\sum_{j=2}^{k} d_{jj}^2 y_j^2)\, y_1 + \sum_{j=2}^{k} d_{jj} y_j^2. \tag{A9}$$

Application of (A1), with $A = -2\Sigma d_j^2 y_j^2$ and $B = \Sigma\, d_{jj} y_j^2$, gives Theorem 1.

The Corollary at the beginning of §4 follows in a similar way. Direct evaluation from (A6) shows that the curvature matrix d_z of $\mathscr{C}_0$ at $\hat{\eta}(0) = z$ equals $d + O_p(n^{-3/2})$. Then (6) and (A9) show that W_0 is of the form $(1+A)\,y_1 + B$,

$$1+A \doteq \frac{1-2\Sigma\, d_{jj}^2 y_j^2}{1-\operatorname{tr}(d^2)}, \quad B \doteq \frac{\Sigma\, d_{jj}(y_j^2-1)}{1-\operatorname{tr}(d^2)}, \tag{A10}$$

where the sums are over $j = 2, \ldots, k$, leading via (A1) to the Corollary.

In addition to having chosen $\eta = 0$, we have specified in (A2), (A3) that the unit orthogonal $\Delta(\eta)$ to $\mathscr{C}_0$ at $\eta = 0$ equals e_1, the first coordinate vector. The curvature matrix d_η 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 η, $\Delta(\eta)$ and d_η. The choice $t(0) = 0$ is also innocuous, since none of the results depends on the name θ assigned to the level surfaces.

To complete the argument for Theorem 1 and its Corollary, we need to extend expression (A6) for $\mathscr{C}_0$ to higher terms, say to

$$\eta_1 = -\{\eta_{(2)}' d\eta_{(2)} + d^{(3)}(\eta_{(2)}^3) + d^{(4)}(\eta_{(2)}^4 + \ldots\}, \tag{A11}$$

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

The calculation leading to (A9) now gives $X_0 \doteq (1+A)\,y_1 + B$,

$$A = -2\sum_{j=2}^{k} d_{jj}^2 y_j^2, \quad B = \sum_{j=2}^{k} d_{jj} y_j^2 + d^{(3)}(y_{(2)}^3).$$

Formula (A1) again gives Theorem 1, and the Corollary. Notice that the cubic term in B contributes nothing to $E(B)$ because $E\{d^{(3)}(y_{(2)}^3)\} = 0$, and that it adds terms of $o(n^{-1})$ everywhere else in (A1).

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

$$\Delta(\theta) = \frac{1}{(1+\theta^2 \,\|\, c \,\|^2)^{\frac{1}{2}}}\begin{pmatrix} 1 \\ \theta c \end{pmatrix}, \tag{A12}$$

so that the squared cosine of the angle between $\Delta(\theta)$ and $\Delta(0) = e_1$ is $\mathrm{co}^2(\theta) = (1+\theta^2 \,\|\, c\,\|^2)^{-1} \doteq (1-\theta^2 \,\|\, c\,\|^2)$. The curvature matrix of $\mathscr{C}_\theta$ at $(\theta, 0, \ldots, 0)$ is evaluated to be $(d+\theta e)\,(1+O(\theta^2 \,\|\, c\,\|^2)) \doteq (d+\theta e)$.

Now we can verify Theorem 2. Suppose that the level surfaces $\mathscr{C}_\theta$ 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); θ in (A2) measures distance along $e_1 = \Delta(0)$, the normal vector to $\mathscr{C}_\theta$ at 0, so θ plays the role of x in (8A), (8B). The arguments following (A12) confirm relations (8A), (8B).

Notice that for situation (A2), (A3), the maximum likelihood estimate $\hat{\theta}$ is

$$\hat{\theta} = \frac{y_1 + D(y_{(2)})}{1 - C(y_{(2)}) - E(y_{(2)})} \doteq \{1 + C(y_{(2)}) + E(y_{(2)}) + C^2(y_{(2)})\}\,\{y_1 + D(y_{(2)})\}, \tag{A13}$$

and likewise under bootstrap sampling

$$\hat{\theta}^* \doteq \{1+C(y^*_{(2)})+E(y^*_{(2)})+C^2(y^*_{(2)})\}\{y^*_1+D(y^*_{(2)})\}.$$

Using (A1) again, applied to the bootstrap distribution $y^* \sim N_k(0, I)$, gives the first four cumulants of $\hat{\theta}^*$ to be to $O(n^{-1})$

$$\{\mathrm{tr}\,(d), 1+2\,\mathrm{tr}\,(d^2)+2\,\mathrm{tr}\,(e)+3\,\|\,c\,\|^2, 0, 12\,\|\,c\,\|^2\}. \tag{A14}$$

Theorem 2 follows from standard Edgeworth, Cornish–Fisher expansions (Kendall & Stuart, 1977, §6.18) used to evaluate the terms in (11), (12) from the approximate cumulants (A14). For example $\mathrm{pr}_*\{\hat{\theta}^* < \hat{\theta} = 0\} \doteq \Phi\{1-\mathrm{tr}\,(d)\}$, so that

$$z_0 \doteq -\mathrm{tr}\,(d). \tag{A15}$$

To understand approximation (8), suppose first that $d = e = 0$ in (A3), so that the $\mathscr{C}_\theta$ are straight lines. Having observed $y = 0$, we have

$$W_\theta = -\theta\,\mathrm{co}\,(\theta) \doteq -\theta(1-\tfrac{1}{2}\,\|\,c\,\|^2\,\theta^2). \tag{A16}$$

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

Finally, it is not difficult to see why the bootstrap intervals are invariant under transformations (14). Following through the definitions shows that $\hat{\theta}^*(b) = \tilde{t}\{\hat{\eta}^{\max *}(b)\}$, where $\tilde{\eta}^{\max *}(b) = h_2\{\hat{\eta}^*(b)\} = h_2\{y^*(b)\}$, and also that $\tilde{t}(\tilde{\eta}) = t\{h_2^{-1}(\eta)\}$. 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

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator (with discussion). *J. R. Statist. Soc.* B **46**. To appear.

BUCKLAND, S. (1983). Monte Carlo methods for confidence interval estimation using the bootstrap method. *Bias* **10**, 194–212.

CREASY, M. A. (1954). Limits for the ratio of the means (with discussion). *J. R. Statist. Soc.* B **16**, 186–94.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7**, 1–25.

EFRON, B. (1981). Nonparametric standard errors and confidence intervals. *Can. J. Statist.* **9**, 139–72.

EFRON, B. (1982a). *The Jackknife, the Bootstrap, and Other Resampling Plans.* Philadelphia: S.I.A.M.

EFRON, B. (1982b). Transformation theory: How normal is a one parameter family of distributions? *Ann. Statist.* **10**, 323–39.

EFRON, B. (1984). Comparing non-nested linear models. *J. Am. Statist. Assoc.* **79**. To appear.

FIELLER, E. C. (1954). Some problems in interval estimation (with discussion). *J. R. Statist. Soc.* B **16**, 175–85.

KENDALL, M. G. & STUART, A. (1977). *The Advanced Theory of Statistics* **1**, 4th ed. London: Griffin.

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71**, 233–44.

ROBISON, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. *J. Am. Statist. Assoc.* **59**, 214–24.

[*Received March* 1984. *Revised July* 1984]