ON PARAMETRIC BOOTSTRAP PROCEDURES FOR SECOND-ORDER ACCURATE CONFIDENCE LIMITS

BY

THOMAS J. DICICCIO AND JOSEPH P. ROMANO

TECHNICAL REPORT NO. 293
APRIL 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON PARAMETRIC BOOTSTRAP PROCEDURES FOR
SECOND-ORDER ACCURATE CONFIDENCE LIMITS

BY

THOMAS J. DICICCIO AND JOSEPH P. ROMANO

TECHNICAL REPORT NO. 293
APRIL 1988

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON PARAMETRIC BOOTSTRAP PROCEDURES FOR
SECOND-ORDER ACCURATE CONFIDENCE LIMITS

by

Thomas J. DiCiccio and Joseph P. Romano
Department of Statistics, Stanford University
Stanford, California 94305-4065, U.S.A.
Summary

The use of simulation procedures to construct approximate confidence limits for scalar parameters is considered in parametric models. Efron's (1987) accelerated bias-corrected percentile bootstrap method, which requires an analytical correction, is shown to be second-order accurate, and a general formula is given for the correction. An alternative second-order accurate percentile method that does not require such theoretical adjustments is proposed. This procedure, called the automatic percentile method, involves the construction of a least favourable subfamily of the model which is related to orthogonal parameters. The use of variance-stabilizing transformations to improve the accuracy of the bootstrap-\(t\) method is also discussed, and the second-order accuracy of a related procedure recently proposed by Tibshirani (1988) is shown.

Some key words: Approximate confidence limit; Bootstrap-\(t\); Least favourable family; Orthogonal parameters; Percentile method; Second-order accuracy; Variance-stabilizing transformation.
1. Introduction

This article concerns the use of bootstrap or simulation procedures to construct approximate confidence limits for scalar parameters in parametric settings. The emphasis is on second-order accurate procedures, where second-order accuracy is defined for convenience in the sense of Hall (1988). Two such procedures, both introduced by Efron (1981, 1987), are the bootstrap–$t$ method and the accelerated bias–corrected ($BC_a$) percentile method.

Since an estimator of the variance of an estimator is typically available in parametric situations, the bootstrap–$t$ method is especially convenient to use because it is automatic from a computational standpoint. However, this method lacks the property of exact invariance under reparameterization. As Efron (1981) suggests, situations can arise, particularly when dealing with small samples, where the bootstrap–$t$ method produces quite inaccurate approximations, even if it is applied in a natural parameterization. One such situation, considered for Table 1, is that of a sample of size $n = 8$ drawn from a bivariate normal distribution, with the correlation coefficient $\rho$ being the parameter of interest. Table 1 shows the bootstrap–$t$ 95% confidence interval for $\rho$, based on the usual sample correlation coefficient $r$ and $\text{var}(r) = n^{-1}(1 - \rho^2)^2 + O(n^{-2})$, in the case that $r = 0.5$ is observed. The exact interval for this problem also shown in Table 1 is discussed further in Section 2.2.

To circumvent difficulties that can arise with the bootstrap–$t$, Efron has developed various percentile methods; see DiCiccio and Romano (1988) for a review. These methods are parameterization invariant, and in increasing order of sophistication they are the simple percentile method, the bias–corrected ($BC$) percentile method, and the $BC_a$ method mentioned earlier. Among these methods, only the $BC_a$ is generally second–order accurate. Approximations obtained by using the percentile methods for the correlation coefficient example are shown in Table 1. Because of special features of this example, the $BC$ and $BC_a$ methods coincide; typically, these two procedures are distinct.

One drawback of the $BC_a$ method is that it involves a quantity known as the accelera-
tion constant, which must be determined by a theoretical calculation in most applications. Both the bootstrap–$t$ and $BC_a$ methods are discussed in detail in Section 2.1, and an explicit formula is given there for the acceleration constant. This formula generalizes the one given by Efron (1987) for the important case of maximum likelihood estimation, and it can be used, for example, to show that zero is an appropriate choice for the acceleration constant in the correlation coefficient problem.

Since the theoretical evaluation of the acceleration constant is often inconvenient or intractable, it is of interest to consider procedures related to the percentile methods that are second-order accurate yet do not require any analytical corrections. One such procedure, called the automatic percentile method, is introduced in Section 2.2. Being fully automatic, this method is naturally more computationally demanding than the $BC_a$; however, the automatic percentile method is not a double-bootstrap procedure. Approximations obtained by using this method for the correlation coefficient problem are shown in Table 1. The automatic percentile method involves a least favourable family construction similar to the one used by Efron (1987) to develop his formula for the acceleration constant in maximum likelihood estimation. A connection between this least favourable family and orthogonal nuisance parameters, where orthogonality is defined in the sense of Cox and Reid (1987), is also indicated in Section 2.2.

In some cases, the accuracy of the bootstrap–$t$ method can be improved substantially by appropriately transforming the parameter of interest. Table 1 shows the improvement obtained in the correlation coefficient example by introducing the reparameterization $\phi = g(\rho) = \tanh^{-1} \rho$, for which $\text{var}\{g(\rho)\} = n^{-1} + O(n^{-2})$. The use of such variance-stabilizing transformations in connection with the bootstrap–$t$ method is discussed generally in Section 2.3. The least favourable family construction is again involved in these transformations. Related transformations have been considered by Tibshirani (1988) to devise a variant of the bootstrap–$t$ method. His transformations are also discussed in Section 2.3, and the second-order accuracy of the procedure that he has proposed is established for parametric settings.

Further examples are considered in Section 3. Section 4 contains some technical
justification of the discussion that is given in Section 2.
2. Bootstrap Procedures

2.1. The bootstrap–t and $BC_a$ methods

Consider a family of densities indexed by the vector parameter $\eta = (\eta^1, \ldots, \eta^p)$, and let $\hat{\eta} = (\hat{\eta}^1, \ldots, \hat{\eta}^p)$ be an estimator of $\eta$ based on a sample of size $n$. Suppose that $\theta = \theta(\eta)$ is a real-valued parameter of interest and that the standard deviation of the estimator $\hat{\theta} = \theta(\hat{\eta})$ is $n^{-1/2} \sigma(\eta) + O(n^{-3/2})$. Following Hall (1988), the exact upper $1 - \alpha$ confidence limit for $\theta$ is defined to be

\[
\hat{\theta}_{EX}(1 - \alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} K^{-1}_\eta(\alpha),
\]

where $\hat{\sigma} = \sigma(\hat{\eta})$ and $K_\eta$ is the distribution function of $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$. Since by definition $pr_\eta\{K^{-1}_\eta(\alpha) \leq n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \} = 1 - \alpha$, it follows that $pr_\eta\{\theta \leq \hat{\theta}_{EX}(1 - \alpha)\} = 1 - \alpha$. An approximate upper $1 - \alpha$ confidence limit $\hat{\theta}_A(1 - \alpha)$ is said to be second-order accurate if the difference between $\hat{\theta}_A(1 - \alpha)$ and $\hat{\theta}_{EX}(1 - \alpha)$ is of order $O_p(n^{-3/2})$, in which case the error in coverage level for $\hat{\theta}_A(1 - \alpha)$ is typically of order $O(n^{-1})$.

The parametric bootstrap distribution of $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ is $K_\hat{\eta}$, and the bootstrap–t approximate upper $1 - \alpha$ confidence limit for $\theta$ is

\[
\hat{\theta}_{BT}(1 - \alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} K^{-1}_{\hat{\eta}}(\alpha).
\]

The bootstrap–t method is easily seen to be second-order accurate since

\[
K^{-1}_{\hat{\eta}}(\alpha) = K^{-1}_\eta(\alpha) + O_p(n^{-1}).
\]

Although in many situations, such as location-scale families, the bootstrap–t method does produce very accurate approximate limits, examples can arise where this method performs poorly. As alternatives to the bootstrap–t method, Efron (1981, 1987) has developed various percentile methods for constructing bootstrap confidence limits.

The most accurate of the percentile procedures is the accelerated bias–corrected ($BC_a$) percentile method. If $G_\eta$ is the distribution function of the estimator $\hat{\theta}$, i.e., $G_\eta(x) = pr_\eta(\hat{\theta} \leq x)$, then the parametric bootstrap distribution of $\hat{\theta}$ is $G_{\hat{\eta}}$. The $BC_a$ approximate upper $1 - \alpha$ confidence limit for $\theta$ is given by

\[
\hat{\theta}_{BC_a}(1 - \alpha) = G^{-1}_{\hat{\eta}} \left\{ \Phi \left( \frac{Z + (Z + z_{1-\alpha})}{1 - A(Z + z_{1-\alpha})} \right) \right\},
\]

where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution. The quantities $Z$ and $A$ that appear in (1) are referred to respectively as the bias correction and the acceleration constant, and they are to be determined. The $BC$ limit $\hat{\theta}_{BC}(1 - \alpha)$ is obtained by setting
\( A = 0 \) in (1), and the simple percentile limit \( \hat\theta_S(1 - \alpha) \) is obtained by setting \( A = Z = 0 \). Thus, \( \hat\theta_{BC}(1 - \alpha) = G^{-1}_\hat{\eta} \{ \Phi(z_{1-\alpha} + 2Z) \} \) and \( \hat\theta_S(1 - \alpha) = G^{-1}_\hat{\eta}(1 - \alpha) \). Efron (1981) argues from transformation theory to show that \( Z \) can be suitably approximated by \( \Phi^{-1}_\hat{\eta}( \hat{\theta} ) \).

Although the bias correction can be obtained directly from the bootstrap distribution \( G_\hat{\eta} \), the acceleration constant cannot usually be so easily derived, and a theoretical calculation is typically required for its evaluation. Efron (1987) presents a formula for the computation of \( A \) in the case that \( \hat{\eta} \) is the maximum likelihood estimator.

To more fully elucidate the bias correction and the acceleration constant, some further notation is required. Let \( U^i = n^{1/2}(\hat{\eta}^i - \eta^i) \), \( i = 1, \ldots, p \), and suppose that

\[
E(U^i) = n^{-1/2} \lambda^i + O(n^{-1}), \quad \text{cov}(U^i, U^j) = \lambda^{i,j} + O(n^{-1}), \\
\text{cum}(U^i, U^j, U^k) = n^{-1/2} \lambda^{i,j,k} + O(n^{-1}),
\]

(2)

and the higher-order cumulants of \( U^1, \ldots, U^p \) are \( O(n^{-1}) \) or smaller. It is shown in Section 4.2 that the \( BC_a \) method is second-order accurate if, to error of order \( O(n^{-1}) \), \( Z \) and \( A \) are given by

\[
Z = n^{-1/2} \left\{ \left( \frac{1}{2} \lambda^{i,j,k} - \lambda^i \lambda^{j,k} \right) \theta_i \theta_j \theta_k + \left( \frac{1}{2} \lambda^{i,k} \lambda^{j,\ell} - \frac{1}{2} \lambda^{i,j} \lambda^{k,\ell} \right) \theta_i \theta_j \theta_{k\ell} \right\} (\lambda^{i,j} \theta_i \theta_j)^{-3/2},
\]

\[
A = n^{-1/2} \left\{ \left( \frac{1}{2} \lambda^{i,j} \lambda^{\ell,k} - \frac{1}{3} \lambda^{i,j,k} \right) \theta_i \theta_j \theta_k \right\} (\lambda^{i,j} \theta_i \theta_j)^{-3/2},
\]

(3)

where

\[
\theta_i = \partial \theta(\eta)/\partial \eta^i, \quad \theta_{ij} = \partial^2 \theta(\eta)/\partial \eta^i \partial \eta^j, \\
\lambda^{i,j} = \partial \lambda^{i,j}(\eta)/\partial \eta^k \quad (i, j, k = 1, \ldots, p).
\]

In these expressions, the usual convention is used whereby summation over the range \( 1, \ldots, p \) is assumed for every index that appears both as a subscript and as a superscript. In practice, \( \eta \) is unknown, and it suffices to use estimates of \( Z \) and \( A \) that differ from the expressions in (3) by terms of order \( O_p(n^{-1}) \). Examples of such estimates include Efron’s formula for the bias correction and \( A(\hat{\eta}) \) for the acceleration constant.

In the case where \( \hat{\eta} \) is the maximum likelihood estimator, formulae (3) for \( Z \) and \( A \) can be expressed in terms of moments of the derivatives of the log-likelihood function \( \ell(\eta) \).
based on the entire sample of size $n$. Let

$$
\ell_i = \partial \ell(\eta)/\partial \eta^i, \quad \ell_{ij} = \partial^2 \ell(\eta)/\partial \eta^i \partial \eta^j,
$$

$$
\kappa_{i,j} = E(\ell_i \ell_j), \quad \kappa_{i,j,k} = E(\ell_i \ell_j \ell_k), \quad \kappa_{i,j,k,l} = E(\ell_i \ell_j \ell_k \ell_l).
$$

Thus $(\kappa_{i,j})$ is the expected information matrix, and let $(\kappa^{i,j})$ be its inverse. Then (3) becomes

$$
Z = A + \frac{1}{2} \left\{ \{ (\kappa_{i,j,k} + \kappa_{i,j,k}) \mu^i - \theta_{jk} \} \{ \kappa^{i,k}(\kappa_{\ell,m} \mu^\ell \mu^m) - \mu^i \mu^k \} \right\}(\kappa_{i,j} \mu^i \mu^j)^{-3/2},
$$

$$
A = \frac{1}{6} (\kappa_{i,j,k} \mu^i \mu^j \mu^k)(\kappa_{i,j} \mu^i \mu^j)^{-3/2},
$$

(4)

where $\mu^i = \kappa^{i,j} \theta_j$. Efron (1987) gives essentially this formula for the acceleration constant.

Rather than recommending the use of $A(\hat{\eta})$ to estimate $A$, Efron suggests using the quantity obtained by replacing $\kappa_{i,j}$ in (4) with its observed counterpart $-\ell_{ij}(\hat{\eta})$.

Although the $BC_a$ method is second–order accurate, the necessary acceleration constant can be difficult to estimate. The automatic percentile method, introduced in the following subsection, is also second–order accurate, yet it does not require such analytical adjustments.

### 2.2. The automatic percentile method

To describe this method first in the simplest context, suppose that $\theta = \eta^1$ and that $\eta^2, \ldots, \eta^p$ are orthogonal to $\theta$, i.e., $\lambda^{1,i} = 0$ ($i = 2, \ldots, p$). For convenience, let $\psi = (\eta^2, \ldots, \eta^p)$ be the vector of nuisance parameters, so that $\eta = (\theta, \psi)$. The automatic percentile approximate upper $1 - \alpha$ confidence limit for $\theta$ is

$$
\hat{\theta}_{AP}(1 - \alpha) = G_{\eta}^{-1}\{G_{(\theta_0, \psi)}(\theta_0)\},
$$

(5)

where $\theta_0$ is an initial estimate of $\hat{\theta}_{EX}(1 - \alpha)$ that differs from $\theta$ by order $O_p(n^{-1/2})$, and $\theta' = G_{(\theta_0, \psi)}^{-1}(\alpha)$. Possible choices for $\theta_0$ include the simple percentile limit $\hat{\theta}_S(1 - \alpha)$, the $BC$ percentile limit $\hat{\theta}_{BC}(1 - \alpha)$, and even $\hat{\theta}$ itself. To illustrate the automatic percentile method in Table 1, the simple percentile limit is used for $\theta_0$.

The automatic percentile procedure requires that the distribution function $G_\eta$ of the estimator $\hat{\theta}$ be considered at three values of $\eta : \hat{\eta}, (\theta_0, \hat{\psi})$, and $(\theta'_0, \hat{\psi})$. By choosing $\theta_0 = \hat{\theta}$,
the first two of these values coincide, and only two distribution functions are required. Experience with numerical examples suggests, however, that the accuracy of the method is improved by taking $\theta_0$ to be a better approximation of the exact upper limit. It is thus advisable to use $\hat{\theta}_S(1 - \alpha)$ or $\hat{\theta}_{BC}(1 - \alpha)$ for $\theta_0$. Indeed, for the correlation coefficient example, using $\theta_0 = \hat{\theta}$ produces the interval $[-0.4157(1.18), 0.8990(1.12)]$, while using the $BC$ limits for $\theta_0$ yields $[-0.2791(2.72), 0.8651(2.56)]$. In interpreting the latter interval, it should be recalled that the $BC$ method is already second-order accurate for this particular problem.

Although always possible in principle, orthogonal parameterizations can in practice be difficult to obtain. The automatic percentile method can be extended to nonorthogonal situations by means of the least favourable family construction that was introduced by Stein (1956) and later used by Efron (1987) to develop his expression for the acceleration constant. The least favourable direction $\mu = (\mu_1, \ldots, \mu_p)$ at the point $\eta$ in the parameter space has components defined by $\mu^i = \lambda_{i,j} \theta_j (\lambda_{i,j} \theta_i \theta_j)^{-1}$. Based on the sample of size $n$, the least favourable family is the line $\eta(\tau) = \hat{\eta} + \tau \hat{\mu}$ in the parameter space, indexed by $\tau$ and running through $\hat{\eta}$ in the direction $\hat{\mu} = \mu(\hat{\eta})$.

The automatic percentile approximate upper $1 - \alpha$ confidence limit is generally given by

$$\hat{\theta}_{AP}(1 - \alpha) = G_{\hat{\eta}}^{-1}\{G_{\eta(\tau'_0)}(\theta_0)\};$$

(6)

where $\theta_0$ is an initial approximation to the exact limit $\hat{\theta}_{EX}(1 - \alpha)$, $\tau_0$ satisfies $\theta_0 = \theta(\eta(\tau_0))$, $\theta'_0 = G_{\eta(\tau_0)}^{-1}(\alpha)$, and $\tau'_0$ satisfies $\theta'_0 = \theta(\eta(\tau'_0))$. As in the orthogonal case, $\theta_0$ should be chosen so as to differ from $\theta$ by $O_p(n^{-1/2})$, and reasonable choices for $\theta_0$ are $\hat{\theta}_S(1 - \alpha)$ and $\hat{\theta}_{BC}(1 - \alpha)$. If orthogonality holds, then the least favourable family is $\eta(\tau) = (\hat{\theta} + \tau, \hat{\psi})$, and (6) is easily seen to reduce to (5). For the case of maximum likelihood estimation, either observed or expected information can be used in determining the least favourable family. Quite generally, the covariance matrix necessary for computing the least favourable direction can in principle be derived by simulation.

The second-order accuracy of the automatic percentile method is established in Section 4.3.
In the case of a scalar parameter model, for which no nuisance parameters are present, it is possible to define the exact upper \( 1 - \alpha \) confidence limit as that value of \( \theta \) satisfying

\[
G_\theta(\hat{\theta}) = \alpha
\]  

(7)

where \( G_\theta(x) = \text{pr}_\theta(\hat{\theta} \leq x) \) is the distribution function for the estimator \( \hat{\theta} \). This definition of the exact limit is used by Efron (1987). The automatic percentile method has been studied in this context by DiCiccio and Romano (1987). Necessary and sufficient conditions are given there for the exactness of the method, and furthermore, it is shown generally that if the method is iterated, each successive round of iteration improves the accuracy of the resulting approximate limit by an order of \( n^{-1/2} \). Unfortunately, a similar improvement cannot be achieved by iteration in the multiparameter context because of the presence of nuisance parameters.

To gain further insight into the automatic percentile method for multiparameter situations, consider the approximate limit \( \hat{\theta}_A(1 - \alpha) \) defined by \( \hat{\theta}_A(1 - \alpha) = \theta\{\eta(\tau_A)\} \), where \( \tau_A \) is that value of \( \tau \) satisfying

\[
G_{\eta(\tau)}(\hat{\theta}) = \alpha.
\]  

(8)

It is shown in Section 4.3 that \( \hat{\theta}_A(1 - \alpha) \) is also second–order accurate. In an orthogonal situation, \( \hat{\theta}_A(1 - \alpha) \) is simply that value of \( \theta \) which satisfies

\[
G_{(\theta,\hat{\theta})}(\hat{\theta}) = \alpha.
\]  

(9)

The automatic percentile method can be viewed as an algorithm to solve (8) for \( \tau \), and hence for \( \theta \), to a sufficiently high order of accuracy. The results for the scalar–parameter case suggest that by iterating the automatic percentile method, the approximate limits thus obtained converge to \( \hat{\theta}_A(1 - \alpha) \); however, \( \hat{\theta}_A(1 - \alpha) \) is itself only second–order accurate with coverage error typically of order \( O(n^{-1}) \). Hence, at least asymptotically, no improvement in accuracy is to be gained by iteration.

It is worthwhile to remark that the approximate limits \( \hat{\theta}_{AP}(1 - \alpha) \) and \( \hat{\theta}_A(1 - \alpha) \) given at (6) and (8), whose definitions depend upon the least favourable family, do not generally
have the property of parameterization invariance, although their versions defined in terms of orthogonal parameters at (5) and (9) do have this property.

For the example of the correlation coefficient, orthogonality holds, and $\hat{\theta}_A(1 - \alpha)$ has zero coverage error since the distribution of $r$ depends only on $\rho$. The exact limit for this problem is usually taken to be $\hat{\theta}_A(1 - \alpha)$, as, for example, in David (1954). This convention is followed in Table 1.

As the foregoing discussion suggests, there is a connection between orthogonal parameters and the least favourable family construction. Suppose that the model is initially specified in terms of the parameter $\eta = (\eta^1, \ldots, \eta^p)$, and let $(\theta, \psi)$ be an orthogonal parameterization. Then $\eta$ can be regarded as a function of $(\theta, \psi)$, and a straightforward calculation shows that $\partial \eta(\theta, \psi)/\partial \theta = \mu^i + O(n^{-1})$. Moreover, for values of $\theta$ which are $O_p(n^{-1/2})$ distant from $\hat{\theta}$,

$$
\eta^i(\theta, \hat{\psi}) = \hat{\eta}^i + (\theta - \hat{\theta})\hat{\mu}^i + O_p(n^{-1}).
$$

In the case where $\hat{\eta}$ is the maximum likelihood estimator, a version of the automatic percentile method can be implemented without direct use of the least favourable family. For a description of how this simplification can be achieved, let $\hat{\eta}(\theta)$ be the restricted maximum likelihood estimator of $\eta$ for a fixed value $\theta$ of the parameter of interest. Thus $\hat{\eta} = \hat{\eta}(\hat{\theta})$. The approximate upper $1 - \alpha$ confidence limit for $\theta$ is given by

$$
G_{\hat{\eta}}^{-1}\{G_{\hat{\eta}(\theta)}(\theta_0)\},
$$

where, as in (6), $\theta_0$ is an initial estimate of $\theta_0 \chi(1 - \alpha)$ that differs from $\theta$ by order $O_p(n^{-1/2})$, and $\theta_0' = G_{\hat{\eta}(\theta)}(\alpha)$. The approximate limit (10) is second-order accurate, and it has the property of parameterization invariance. The curve $\hat{\eta}(\theta)$ is locally close to the least favourable family; indeed,

$$
\frac{\partial \hat{\eta}^i(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} = \frac{I^{ij}}{I_{ij}} = \hat{\mu}^i + O_p(n^{-1/2}),
$$

where $(I^{ij})$ is the inverse of the observed Fisher information matrix. Thus, for values of $\theta$ that are $O_p(n^{-1/2})$ distant from $\hat{\theta}$,

$$
\hat{\eta}^i(\theta) = \hat{\eta}^i + (\theta - \hat{\theta})\mu^i + O_p(n^{-1}).
$$
Another second-order accurate approximate upper $1 - \alpha$ confidence limit is given by that value of $\theta$ which satisfies

$$G_{\tilde{\eta}(\theta)}(\hat{\theta}) = \alpha. \quad (11)$$

Again, the results for the scalar parameter case suggest that if the procedure described at (10) is iterated, then the resulting approximate limits will converge to the approximate limit given by (11).

2.3. Variance-stabilizing transformations

Consider an arbitrary but fixed point $\eta_0$ in the parameter space, having associated least favourable direction $\mu_0$, and let $\theta_0 = \theta(\eta_0)$. The least favourable family indexed by $\tau$ running through $\eta_0$ is $\eta(\tau) = \eta_0 + \tau \mu_0$. It is now natural to regard $\tau$ as a function of $\theta$, so that $\theta(\eta_0 + \tau(\cdot)\mu_0)$ is the identity map. Then the variance of $n^{1/2}(\hat{\theta} - \theta)$ at the point $\eta_0 + \tau(\theta)\mu_0$ is $\text{var}_{\eta_0 + \tau(\theta)\mu_0}(n^{1/2}(\hat{\theta} - \theta)) = \left[\sigma(\eta_0 + \tau(\theta)\mu_0)\right]^2 + O(n^{-1})$. Consider the reparameterization $\phi = g_{\eta_0}(\theta)$ given by

$$g_{\eta_0}(\theta) = \int_0^\theta \frac{du}{\sigma(\eta_0 + \tau(u)\mu_0)}. \quad (12)$$

Let $\hat{\phi} = g_{\eta_0}(\hat{\theta})$ and $\phi_0 = g_{\eta_0}(\theta_0)$. This transformation is the variance-stabilizing one along the least favourable family in the sense that for all $\theta$,

$$\text{var}_{\eta_0 + \tau(\theta)\mu_0}(n^{1/2}(\hat{\phi} - \phi)) = 1 + O(n^{-1}).$$

In particular, $\text{var}_{\eta_0}(n^{1/2}(\hat{\phi} - \phi_0)) = 1 + O(n^{-1})$. Adapting the notation previously introduced, let $\bar{K}_{\eta_0}$ be the distribution function of $n^{1/2}(\hat{\phi} - \phi_0)$, i.e., $\bar{K}_{\eta_0}(x) = \text{pr}_{\eta_0}(n^{1/2}(\hat{\phi} - \phi) \leq x)$. Thus the so-called exact upper $1 - \alpha$ confidence limit for $\phi_0$ is $\hat{\phi}_{EX}(1 - \alpha) = \hat{\phi} - n^{-1/2} \bar{K}_{\eta_0}^{-1}(\alpha)$, and $\text{pr}_{\eta_0}(\theta_0 \leq g_{\eta_0}^{-1}\{\hat{\phi}_{EX}(1 - \alpha)\}) = 1 - \alpha$. Although the coverage level of $g_{\eta_0}^{-1}\{\hat{\phi}_{EX}(1 - \alpha)\}$ is exact, it is easily shown that $g_{\eta_0}^{-1}\{\hat{\phi}_{EX}(1 - \alpha)\} = \hat{\theta}_{EX}(1 - \alpha) + O_p(n^{-3/2})$, and this discrepancy between $g_{\eta_0}^{-1}\{\hat{\phi}_{EX}(1 - \alpha)\}$ and $\hat{\theta}_{EX}(1 - \alpha)$ highlights the lack of parameterization invariance in the definition of the exact upper limit.

A bootstrap version of this limit is conceivable. Using the estimator $\hat{\eta}$ based on a sample of size $n$, the transformation $g_{\eta_0}$ and the distribution $\bar{K}_{\eta_0}$ can be determined for
\( \eta_0 = \hat{\eta} \). The variance–stabilized bootstrap–t approximate upper \( 1 - \alpha \) confidence limit is given by

\[
\hat{\theta}_{VS}(1 - \alpha) = g_{\hat{\eta}}^{-1}\{\hat{\phi} - n^{-1/2}\bar{K}_{\hat{\eta}}^{-1}(\alpha)\}.
\] (13)

It is shown in Section 4.4 that \( \hat{\theta}_{VS}(1 - \alpha) \) is second–order accurate.

In the orthogonal case \( \eta = (\theta, \psi) \), the least favourable family through \( \eta_0 \) parameterized by \( \theta \) is \( \eta_0 + \tau(\theta)\mu_0 = (\theta, \psi_0) \), and \( \text{var}(\theta, \psi_0)\{n^{1/2}(\hat{\theta} - \theta)\} = \{\sigma(\theta, \psi_0)\}^2 + O(n^{-1}) \). The transformation \( g_{\eta_0} \) is then given by

\[
g_{\eta_0}(\theta) = \int^{\theta}_0 \frac{du}{\sigma(u, \psi_0)}.
\]

In the correlation coefficient example, \( \sigma(\rho, \psi_0) = 1 - \rho^2 \), so that \( g_{\eta_0}(\rho) = \tanh^{-1}\rho \) for all \( \eta_0 \). The results obtained by using (13) in this example are quite accurate, as illustrated in Table 1.

Recently, Tibshirani (1988) has suggested a variant of (13) obtained by using the transformation

\[
g_{\eta_0}(\theta) = \int^{\theta}_0 \frac{du}{\{\text{E}_{\eta_0}(\hat{\theta}^2|\hat{\theta} = u)\}^{1/2}}.
\] (14)

This method has the advantage of not requiring direct use of the least favourable family; however, to estimate the integrand by simulation, as Tibshirani suggests, is very demanding computationally. In the case of the correlation coefficient, (14) also gives the \( \tanh^{-1}\theta \) transformation. It is also shown in Section 4.4 that the variance–stabilized bootstrap–t method based on transformation (14) is second–order accurate.
3. Further Examples

Example 1. Location–scale models. Consider a location–scale family indexed by \( \eta = (\theta, \sigma) \), where \( \theta \) and \( \sigma \) are the location and scale parameters, respectively. Let \( \hat{\theta} \) and \( \hat{\sigma} \) be equivariant estimators, and suppose that \( \theta \) is the parameter of interest. Since \( (\hat{\theta} - \theta)/\hat{\sigma} \) is pivotal in this example, the bootstrap–t method produces exact confidence limits. It is of interest, however, to apply the other procedures for comparison.

The formulae for the bias correction and the acceleration constant of the \( BC_a \) method reduce to simple expressions. There exist constants \( v_{11}, v_{111}, v_{12}, v_{22} \), and \( v_{1111} \) such that, in the notation of (2), \( \lambda^1 = v_{11} \sigma, \lambda^{1,1} = v_{11} \sigma^2, \lambda^{1,2} = v_{12} \sigma^2, \lambda^{2,2} = v_{22} \sigma^2, \) and \( \lambda^{1,1,1} = v_{1111} \sigma^3 \). In terms of these constants, (3) becomes

\[
Z = n^{-1/2} \left( \frac{1}{6} v_{1111} - v_{11} \right) v_{11}^{-3/2},
\]

(15)

\[
A = n^{-1/2} \left( v_{1111} - \frac{1}{3} v_{111} \right) v_{11}^{-3/2}.
\]

For any point \( \eta_0 = (\theta_0, \sigma_0) \) in the parameter space, the least favourable direction is \( (1, v_{12} v_{11}^{-1}) \), and hence the least favourable family through \( \eta_0 \) is \( \eta_0 + \tau(\theta)\mu_0 = (\theta_0, \sigma_0 + (\theta - \theta_0)v_{12} v_{11}^{-1}) \). In terms of this family, using definitions (6) and (8),

\[
\hat{\theta}_{AP}(1 - \alpha) = \hat{\theta}_A(1 - \alpha) = \hat{\theta} - \hat{\sigma} \left\{ \frac{D^{-1}(\alpha)}{1 + D^{-1}(\alpha)v_{12} v_{11}^{-1}} \right\},
\]

(16)

where \( D \) is the distribution function of the pivotal quantity \( (\hat{\theta} - \theta)/\sigma \). Moreover, if the parameter \( \psi \) is defined by \( \sigma = (\theta + \psi)v_{11} v_{11}^{-1} \), then \( (\theta, \psi) \) is an orthogonal parameterization, and \( \hat{\theta}_{AP}(1 - \alpha) \) and \( \hat{\theta}_A(1 - \alpha) \) given by (5) and (9) are both the same as (16).

The variance–stabilizing transformation along the least favourable family (12) is

\[
g_{\eta_0}(\theta) = \frac{v_{11}^{1/2}}{v_{12}} \log \left\{ 1 + \frac{v_{12}}{v_{11}} \frac{(\theta - \theta_0)}{\sigma_0} \right\},
\]

and the variance–stabilized bootstrap–t limit \( \hat{\theta}_{VS}(1 - \alpha) \) based on this transformation also equals (16).

Efron (1987) and Tibshirani (1988) consider the special case where

\[
\frac{(\hat{\theta} - \theta)}{\sigma} \sim \chi^2_{(2n)} \frac{2n}{2n} - 1, \quad \frac{\hat{\sigma}}{\sigma} \frac{(\hat{\theta} - \theta)}{\sigma} \sim \left\{ 1 + \frac{(\hat{\theta} - \theta)}{\sigma} \right\} \left( \frac{\chi^2_{(n-1)}}{n} \right)^{1/2},
\]

(17)

12
for \( n = 15 \). In this case, \( v_1 = 0, v_{11} = v_{12} = 1, v_{22} = \frac{1}{2} \), and \( v_{111} = 2 \), and it follows from (15) that \( A = Z = \frac{1}{3} n^{-1/2} \). Table 2 shows the 90% confidence intervals for \( \theta \) obtained by various bootstrap procedures having observed \( \hat{\theta} = 0, \hat{\sigma} = (14/15)^{1/2} \). The \( BC^\alpha \) and automatic percentile methods produce very similar results.

Furthermore, it follows from (17) that

\[
E_{\eta_0}(\hat{\sigma}^2|\hat{\theta} = u) = \left( \frac{n - 1}{n} \right) \{ \sigma_0 + (u - \theta_0) \}^2,
\]

and Tibshirani's (1988) version of the variance–stabilized bootstrap–t procedure, based on transformation (14), produces the same limits as (16). Tibshirani approximates the function (18) by simulation, and he reports the approximate 90% interval as \((-0.334, 0.610)\). The difference between his interval and the one reported in Table 2 is presumably attributable to the simulation error arising in his method.

Example 2. The ratio of normal means. Consider a sample \((X_i, Y_i), i = 1, \ldots, n,\) from a bivariate normal distribution having mean \((\mu_1, \mu_2)\) and identity covariance matrix, and suppose that \( \theta = \mu_1/\mu_2 \) is the parameter of interest. Efron (1985, 1987) discusses bootstrap confidence limits for \( \theta \) based on the maximum likelihood estimator \( \hat{\theta} = \bar{X}/\bar{Y} \). He considers specifically the case where \( n = 1 \) and (4,8) is observed, and this situation is assumed for Table 3. Table 3 reports the 95% confidence intervals for \( \theta \) obtained by a variety of the percentile methods described in Section 2. The exact limits given in Table 3 are the Fieller ones, which arise from the standard normal distribution of the pivotal quantity

\[
n^{1/2}(\bar{X} - \theta \bar{Y})/(1 + \theta^2)^{1/2}.
\]

Using formulae (4), it can be shown in this situation that \( Z = A = 0 \), so that the simple percentile, \( BC \), and \( BC^\alpha \) methods all coincide, and they are all second–order accurate. For the observation considered, the simple percentile method gives the Fieller limits. Moreover, an orthogonal parameterization is \( \eta = (\theta, \psi) \), where \( \mu^1 = \theta \psi/(1 + \theta^2)^{1/2}, \mu^2 = \psi/(1 + \theta^2)^{1/2} \), and in terms of this parameterization, \( \hat{\theta}_A(1 - \alpha) \) defined at (9) is also the Fieller limit. As a consequence, it is easy to see that if \( \theta_0 \) is taken to be the simple percentile limit, then \( \hat{\theta}_{AP}(1 - \alpha) \) defined at (5) will again be the exact limit.
To illustrate the results of the automatic percentile method when an orthogonal parameterization is not used, Table 3 also considers the limits defined at (8) for two parameterizations $\eta = (\eta^1, \eta^2)$ defined by $\mu^1 = \eta^1$, $\mu^2 = \eta^2$ and $\mu^1 = \eta^1 \eta^2$, $\mu^2 = \eta^2$. In addition, Table 3 reports the limits obtained by the method described at (11). All three of these intervals are quite close for this example.

Finally, Table 4 demonstrates for this example the convergence of approximate limits obtained by iterating the automatic percentile method (6) to the limit defined at (8). For this table, the parameterization used is $\eta = (\eta^1, \eta^2) = (\mu^1, \mu^2)$, $\theta_0$ is taken to be $\hat{\theta}$, and $\hat{\theta}_i(1 - \alpha)$ denotes the limit obtained after the $i$th iteration.
4. Technical Details

4.1. Introduction

This section provides technical arguments that justify some of the assertions contained in Section 2. The notation introduced in Section 2 is used freely in the present section.

4.2. The bootstrap-\(t\) and \(BC_a\) methods

To error of order \(O(n^{-1})\), the first three cumulants of \(U = n^{1/2}(\hat{\theta} - \theta)/\sigma\) are \(E(U) = n^{-1/2}a\), \(\text{var}(U) = 1\), and \(\kappa_3(U) = n^{-1/2}c\), where

\[
a = \left(\lambda^i\theta_i + \frac{1}{2}\lambda^i,j\theta_{ij}\right)b^{-1},
\]

\[
c = (\lambda^i,j,k\theta_i\theta_j\theta_k + 3\lambda^i,k\lambda^j,\ell\theta_i\theta_j\theta_{k\ell})b^{-3},
\]

and

\[
b = (\lambda^i,j\theta_i\theta_j)^{1/2}.
\]

Note that \(\text{var}\{n^{1/2}(\hat{\theta} - \theta)\} = b^2 + O(n^{-1})\), i.e., \(\sigma^2 = b^2 + O(n^{-1})\). It is assumed then that the distribution function \(H_\eta\) of \(U\) has the expansion

\[
H_\eta(x) = \text{pr}_\eta\{n^{1/2}(\hat{\theta} - \theta)/\sigma \leq x\} = \Phi \left[ x - n^{-1/2} \left\{ \left( a - \frac{1}{6}c \right) + \frac{1}{6}cx^2 \right\} \right] + O(n^{-1}),
\]

in terms of standard normal distribution function \(\Phi\). Similarly, to error of order \(O(n^{-1})\), the first three cumulants of \(V = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}\) are \(E(V) = n^{-1/2}(a - d)\), \(\text{var}(V) = 1\), and \(\kappa_3(V) = n^{-1/2}(c - 6d)\), where \(d = (\lambda^i,jb_i\theta_j)b^{-2}\) and \(b_i = \partial b(\eta)/\partial \eta^i\). A straightforward calculation shows that

\[
d = \left(\frac{1}{2}\lambda^i,j\lambda^{\ell,k}\theta_i\theta_j\theta_k + \lambda^i,k\lambda^j,\ell\theta_i\theta_j\theta_{k\ell}\right)b^{-3},
\]

where \(\lambda^i,j_k = \partial \lambda^i,j(\eta)/\partial \eta^k\). It is also assumed that the distribution function \(K_\eta\) of \(V\) has the expansion

\[
K_\eta(x) = \text{pr}_\eta\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\}
= \Phi \left[ x - n^{-1/2} \left\{ \left( a - \frac{1}{6}c \right) + \left( \frac{1}{6}c - d \right)x^2 \right\} \right] + O(n^{-1}).
\]
The quantile functions $H^{-1}_n(\alpha)$ and $K^{-1}_n(\alpha)$ have expansions

$$H^{-1}_n(\alpha) = z_\alpha + n^{-1/2} \left\{ \left( a - \frac{1}{6} c \right) + \frac{1}{6} c z^2_\alpha \right\} + O(n^{-1}),$$

$$K^{-1}_n(\alpha) = z_\alpha + n^{-1/2} \left\{ \left( a - \frac{1}{6} c \right) + \left( \frac{1}{6} c - d \right) z^2_\alpha \right\} + O(n^{-1}),$$

so that $K^{-1}_n(\alpha) = H^{-1}_n(\alpha) - n^{-1/2} d z^2_\alpha + O(n^{-1})$. The exact upper $1-\alpha$ confidence limit for $\theta$ is

$$\hat{\theta}_{EX}(1-\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} K^{-1}_n(\alpha)$$

$$= \hat{\theta} - n^{-1/2} \hat{\sigma} \{ H^{-1}_n(\alpha) - n^{-1/2} d z^2_\alpha \} + O_p(n^{-3/2}).$$

(21)

Since $\hat{a} = a(\hat{\eta}), \ldots, \hat{d} = d(\hat{\eta})$ differ from $a, \ldots, d$ by terms of order $O_p(n^{-1/2})$, it follows that $H^{-1}_n(\alpha)$ and $K^{-1}_n(\alpha)$ differ from $H^{-1}_n(\alpha)$ and $K^{-1}_n(\alpha)$ respectively by terms of order $O_p(n^{-1})$. This observation establishes the second-order accuracy of the bootstrap-$t$ method.

To demonstrate the second-order accuracy of the $BC_a$ method for the values of the bias correction and the acceleration constant given in (3), note first that

$$G_n(x) = pr_n(\hat{\theta} \leq x) = H_n \{ n^{1/2}(x - \theta)/\sigma \},$$

$$G^{-1}_n(\alpha) = \theta + n^{-1/2} \sigma H^{-1}_n(\alpha).$$

It follows from definition (1) that, with error of order $O_p(n^{-3/2}),$

$$\hat{\theta}_{BC_a}(1-\alpha) = \hat{\theta} + n^{-1/2} \hat{\sigma} H^{-1}_n \{ \Phi \{ z + \left[ (z_{1-\alpha} + Z)/(1 - A(z_{1-\alpha} + Z)) \right] \} \}$$

$$= \hat{\theta} + n^{-1/2} \hat{\sigma} H^{-1}_n \{ \Phi(z_{1-\alpha} + 2Z + Az^2_{1-\alpha}) \}$$

$$= \hat{\theta} + n^{-1/2} \hat{\sigma} \left[ z_{1-\alpha} + 2Z + Az^2_{1-\alpha} + n^{-1/2} \left\{ \left( a - \frac{1}{6} c \right) + \left( \frac{1}{6} c \right) z^2_{1-\alpha} \right\} \right]$$

$$= \hat{\theta} - n^{-1/2} \hat{\sigma} \left[ H^{-1}_n(\alpha) - 2 \left\{ Z + n^{-1/2} \left( a - \frac{1}{6} c \right) \right\} \right.$$

$$\left. - \left( A + \frac{1}{3} n^{-1/2} c \right) z^2_\alpha \right].$$

(22)

By choosing the bias correction and acceleration constant so that, with error of order $O(n^{-1}),$

$$Z = -n^{-1/2} \left( a - \frac{1}{6} c \right), \quad A = -n^{-1/2} \left( \frac{1}{3} c - d \right),$$

(23)
expression (22) becomes

\[ \hat{\theta}_{BC_a}(1 - \alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} \left\{ H_{\hat{\theta}}^{-1}(\alpha) - n^{-1/2} d \right\} + O_p(n^{-3/2}). \]

Comparison with (21) shows that the \( BC_a \) method is second-order accurate for these choices of \( Z \) and \( A \). Formulae (3) are obtained by substituting into (23) the expressions for \( a, c, \) and \( d \) given at (19) and (20).

Since \( \Phi^{-1}\{G_{\hat{\theta}}(\hat{\theta})\} = \Phi^{-1}\{H_{\hat{\theta}}(0)\} = \Phi^{-1}\{H_{\hat{\theta}}(0)\} + O_p(n^{-1}) = Z + O_p(n^{-1}) \), Efron's formula for estimating the bias correction from the bootstrap distribution \( G_{\hat{\theta}} \) suffices for second-order accuracy.

For the case of maximum likelihood estimation, it follows from McCullagh (1987, p. 209) that

\[ \lambda^{i,j} = \kappa^{i,j}, \quad \lambda^{i,j} \theta_i \theta_j = \kappa_{i,j} \mu^i \mu^j, \]

\[ \lambda^i \theta_i = - \left( \frac{1}{2} \kappa_{i,j} + \frac{1}{2} \kappa_{i,k} \right) \mu^i \kappa_{j,k}, \]

\[ \lambda^{i,j,k} \theta_i \theta_j \theta_k = - (2 \kappa_{i,j,k} + 3 \kappa_{i,j,k}) \mu^i \mu^j \mu^k, \]

\[ \lambda^{i,j} \lambda^{\ell,k} \theta_i \theta_j \theta_k = - (\kappa_{i,j,k} + 2 \kappa_{i,j,k}) \mu^i \mu^j \mu^k. \]

Substitution of these formulae into (3) yields (4).

4.3. The automatic percentile method

To verify that the automatic percentile method defined by (6) is second-order accurate, suppose that \( \theta_0 \) differs from \( \theta \) by \( O_p(n^{-1/2}) \), and let \( \tau_0 \) be the value of \( \tau \) satisfying \( \theta_0 = \theta\{\eta(\tau)\} \). Then

\[ \theta'_0 = G_{\eta(\tau_0)}^{-1}(\alpha) = \theta_0 + n^{-1/2} \sigma_0 H_{\eta(\tau_0)}^{-1}(\alpha) \]

\[ = \theta_0 + n^{-1/2} \sigma_0 H_{\eta}^{-1}(\alpha) + O_p(n^{-3/2}), \]

where \( \sigma_0 = \sigma\{\eta(\tau_0)\} \). Hence,

\[ n^{1/2}(\theta'_0 - \theta_0) = \sigma_0 H_{\eta}^{-1}(\alpha) + O_p(n^{-1}). \]

Let \( \tau'_0 \) satisfy \( \theta'_0 = \theta\{\eta(\tau'_0)\} \), and let \( \sigma'_0 = \sigma\{\eta(\tau'_0)\} \). Then

\[ G_{\eta(\tau'_0)}(\theta_0) = H_{\eta(\tau'_0)}\{n^{1/2}(\theta_0 - \theta'_0)/\sigma'_0\} \]

\[ = H_{\eta}\{- (\sigma_0/\sigma'_0) H_{\eta}^{-1}(\alpha)\} + O_p(n^{-1}). \]
Now $\sigma_0/\sigma_0' = 1 - n^{-1/2}dz_\alpha + O_p(n^{-1})$, and thus,

$$\hat{\theta}_{AP}(1 - \alpha) = G^{-1}_\eta\{G_{\eta}(\tau_A')(\theta_0)\}$$

$$= \hat{\theta} - n^{-1/2}\hat{\sigma}H^{-1}_\eta(\alpha)(1 - n^{-1}dz_\alpha) + O_p(n^{-3/2})$$

$$= \hat{\theta} - n^{-1/2}\hat{\sigma}\{H^{-1}_\eta(\alpha) - n^{-1/2}dz^2_\alpha\} + O_p(n^{-3/2}).$$

It follows from (21) that $\hat{\theta}_{AP}(1 - \alpha)$ is second-order accurate.

To develop an expansion for the limit $\hat{\theta}_A(1 - \alpha)$ defined at (8), note first that for values of $\tau$ which are $O_p(n^{-1/2}),$

$$\theta\{\eta(\tau)\} = \hat{\theta} + \tau + \frac{1}{2}\tau^2\hat{e} + O_p(n^{-3/2}), \quad (24)$$

$$\sigma\{\eta(\tau)\} = \hat{\sigma} + \tau\hat{d} + O_p(n^{-1}),$$

where $e = \theta_{ij}\mu^i\mu^j$. Therefore, the condition $G_{\eta}(\tau_A)(\hat{\theta}) = \alpha$ which defines $\tau_A$ can be written as

$$\hat{\theta} + \tau_A + \frac{1}{2}\tau_A^2\hat{e} + n^{-1/2}(\hat{\sigma} + \tau_A\hat{d})H^{-1}_\eta(\alpha) + O_p(n^{-3/2}) = \hat{\theta},$$

and hence,

$$\tau_A = -n^{1/2}\hat{\sigma}H^{-1}_\eta(\alpha)\left\{1 - n^{-1/2}\left(d - \frac{1}{2}\hat{\sigma}e\right)z_\alpha\right\} + O_p(n^{-3/2}).$$

By substituting this expression into (24) and comparing the result with (21), it is easily seen that $\hat{\theta}_A(1 - \alpha) = \theta\{\eta(\tau_A)\}$ is second-order accurate. The second-order accuracy of the limit defined at (11), pertaining to maximum likelihood estimation, can be shown similarly.

4.4. Variance-stabilizing transformations

Consider the transformation $g_{\eta_0}$ given by (12) which has derivatives $g'_{\eta_0}(\theta_0) = \sigma_0^{-1}$ and $g''_{\eta_0}(\theta_0) = -d_0\sigma_0^{-2}$, where $\sigma_0 = \sigma(\eta_0)$ and $d_0 = d(\eta_0)$. Then

$$n^{1/2}(\hat{\phi} - \phi_0) = U - \frac{1}{2}n^{-1/2}U^2d_0 + O_p(n^{-1}),$$

where $U = n^{1/2}(\hat{\theta} - \theta_0)/\sigma_0$. It follows that

$$K_{\eta_0}(x) = pr_{\eta_0}\{n^{1/2}(\hat{\phi} - \phi_0) \leq x\} = H_{\eta_0}\left\{x + \frac{1}{2}n^{-1/2}d_0x^2\right\} + O(n^{-1}),$$

18
and
\[ \hat{K}_{\eta_0}^{-1}(\alpha) = H_{\eta_0}^{-1}(\alpha) - \frac{1}{2} n^{-1/2} d_0 z_\alpha^2 + O(n^{-1}). \]

Now, for values \( \phi \) that are \( O_p(n^{-1/2}) \) distant from \( \phi_0 \),
\[ g_{\eta_0}^{-1}(\phi) = \theta_0 + (\phi - \phi_0)\sigma_0 + \frac{1}{2} (\phi - \phi_0)^2 d_0 \sigma_0 + O_p(n^{-3/2}). \]

Thus,
\[ \hat{\theta}_{VS}(1 - \alpha) = g_{\eta_0}^{-1}\{ \hat{\phi} - n^{-1/2} \hat{K}_{\eta_0}^{-1}(\alpha) \} \]
\[ = \hat{\theta} - n^{-1/2} \hat{\sigma} H_{\eta_0}^{-1}(\alpha) + \frac{1}{2} n^{-1} \hat{\sigma} d z_\alpha^2 + \frac{1}{2} n^{-1} \sigma_0 d_0 z_\alpha^2 + O_p(n^{-3/2}) \]
\[ = \hat{\theta} - n^{-1/2} \hat{\sigma} \{ H_{\eta_0}^{-1}(\alpha) - n^{-1/2} d z_\alpha^2 \} + O_p(n^{-3/2}). \]

Comparison with (21) shows that the variance stabilized bootstrap–t procedure based on transformation (12) is second–order accurate.

Concerning the transformation \( g_{\eta_0} \) given by (14), it follows from McCullagh (1987, p. 164) that for values of \( \theta \) which are \( O(n^{-1/2}) \) distant from \( \theta_0 \),
\[ E_{\eta_0}(\hat{\theta}^2|\hat{\theta} = \theta) = \sigma_0^2 + 2(\theta - \theta_0)\sigma_0 d_0 + O(n^{-1}). \]

In particular, for transformation (14), \( g_{\eta_0}'(\theta_0) = \sigma_0^{-1} + O(n^{-1}) \) and \( g_{\eta_0}''(\theta_0) = -d_0 \sigma_0^{-2} + O(n^{-1}) \). The second–order accuracy of the variance–stabilized bootstrap–t procedure based on transformation (14) can now be established by an argument similar to the one given above for transformation (12).

\textbf{Acknowledgements}

The authors are grateful to Bradley Efron and Peter Hall for helpful discussions and suggestions.
References


20
Table 1. Central 95% confidence intervals for the correlation coefficient, having observed $r = 0.5$ ($n = 8$)

<table>
<thead>
<tr>
<th>Method</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact</td>
<td>-0.2940 (2.50)</td>
<td>0.8663 (2.50)</td>
</tr>
<tr>
<td>Bootstrap-t ($\rho$)</td>
<td>-1.0610</td>
<td>1.1248</td>
</tr>
<tr>
<td>Simple percentile</td>
<td>-0.2716 (2.83)</td>
<td>0.8990 (1.12)</td>
</tr>
<tr>
<td>BC, BC$_{a}$</td>
<td>-0.3636 (1.65)</td>
<td>0.8781 (1.93)</td>
</tr>
<tr>
<td>Automatic percentile</td>
<td>-0.2986 (2.44)</td>
<td>0.8629 (2.68)</td>
</tr>
<tr>
<td>Bootstrap-t ($\phi = \tanh^{-1}\rho$)</td>
<td>-0.3528 (1.77)</td>
<td>0.8803 (1.83)</td>
</tr>
</tbody>
</table>

Each lower limit $\hat{\beta}_L$ is accompanied by $\text{pr}(r \geq 0.5| \rho = \hat{\beta}_L)$ in parentheses as a percentage; each upper limit $\hat{\beta}_U$ is accompanied by $\text{pr}(r \leq 0.5| \rho = \hat{\beta}_U)$.  

---

21
<table>
<thead>
<tr>
<th>Method</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact, Bootstrap-t</td>
<td>[-0.336, 0.670]</td>
</tr>
<tr>
<td>Simple percentile</td>
<td>[-0.371, 0.444]</td>
</tr>
<tr>
<td>BC</td>
<td>[-0.339, 0.499]</td>
</tr>
<tr>
<td>BC$_a$</td>
<td>[-0.304, 0.601]</td>
</tr>
<tr>
<td>Automatic percentile</td>
<td>[-0.304, 0.601]</td>
</tr>
<tr>
<td>Table 3. Central 95% confidence intervals for ( \theta = \mu^1/\mu^2 ) in Example 2</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Exact (Fieller)</td>
<td></td>
</tr>
<tr>
<td>(8); ( \mu^1 = \eta^1 ), ( \mu^2 = \eta^2 )</td>
<td></td>
</tr>
<tr>
<td>[0.2476, 0.8162]</td>
<td></td>
</tr>
<tr>
<td>(8); ( \mu^1 = \eta^1 \eta^2 ), ( \mu^2 = \eta^2 )</td>
<td></td>
</tr>
<tr>
<td>[0.2531, 0.8076]</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>[0.2413, 0.8261]</td>
<td></td>
</tr>
<tr>
<td>[0.2509, 0.8131]</td>
<td></td>
</tr>
<tr>
<td>$1-\alpha$</td>
<td>$\theta_0$</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>0.025</td>
<td>0.5</td>
</tr>
<tr>
<td>0.975</td>
<td>0.5</td>
</tr>
</tbody>
</table>