Statistica Sinica **5**(1995), 141-160

ON BOOTSTRAP PROCEDURES FOR SECOND-ORDER ACCURATE CONFIDENCE LIMITS IN PARAMETRIC MODELS

Thomas J. DiCiccio and Joseph P. Romano

Stanford University

Abstract. This paper concerns the use of simulation procedures to construct second-order accurate confidence limits having coverage error of order $O(n^{-1})$. An explicit formula for the analytical adjustment required in Efron's (1987) BC_a percentile method is derived, automatic percentile methods that do not require analytical adjustments are proposed, and variance-stabilizing transformations designed to improve the performance of the bootstrap-t method are given. The automatic percentile methods and variance-stabilizing transformations involve a least favorable family construction of Stein (1956), which is related to orthogonal parameters. Connections with approximate limits obtained using profile likelihood methods are also discussed.

Key words and phrases: Approximate confidence limit, bootstrap-t, conditional inference, least favorable family, orthogonal parameters, percentile method, second-order accuracy, signed root adjusted likelihood ratio statistic, variance-stabilizing transformation.

1. Introduction

This paper concerns the use of simulation procedures to construct approximate confidence limits for scalar parameters in parametric settings. The emphasis is on second-order accurate procedures, with second-order accuracy defined primarily in the sense of Hall (1988). Two second-order accurate procedures, discussed by Efron (1981, 1987), are the bootstrap-t (B–t) method and the accelerated bias-corrected (BC_a) percentile method.

The B–t method is especially convenient to use in parametric situations, since standard techniques are readily available to routinely estimate the variance of an estimator. However, the B–t method lacks the property of exact invariance under reparameterization. As Efron (1981) suggests, situations can arise, particularly when dealing with small samples, where this 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 ρ the parameter of interest. Table 1 shows the approximate lower and upper 97.5% confidence limits for ρ obtained by applying the B–t method to the usual sample correlation coefficient r, with $\mathrm{var}(r;\rho) \simeq (1-\rho^2)^2/n$, in the case that $r = 0.5$ is observed. The exact interval for this problem, also shown in Table 1, is described in Section 2.2.

To avoid difficulties that can arise in the B–t method, Efron (1981, 1985, 1987) has developed various percentile bootstrap methods; see DiCiccio and Romano (1988) for a review. These methods are parameterization invariant, and in increasing order of refinement they are the simple percentile (S) method, the bias-corrected (BC) percentile method, and the BC_a method mentioned previously. Among these methods, only the BC_a is second-order accurate in general. Approximations based on (8) below, using the percentile methods for the correlation coefficient example are shown in Table 1. Because of the special features of this example, the BC and BC_a methods coincide.

One drawback of the BC_a method is that it involves a quantity A, known as the acceleration constant, which must be determined by theoretical calculation. Both the B–t and the BC_a methods are discussed in detail in Section 2.1, and an explicit formula is given there for A. This formula generalizes the one given by Efron (1987) for the special case of maximum likelihood estimation, and it can be used, for example, to show that $A = 0$ is appropriate in the correlation coefficient problem.

It is of interest to consider other percentile methods that are second-order accurate, yet do not require any analytical adjustments. Such procedures, called automatic percentile (AP) methods, are described in Section 2.2. Approximations based on (8) using an AP method for the correlation coefficient problem are shown in Table 1. The AP methods involve a least favorable family, similar to the one used by Efron (1987) to develop his formula for A in maximum likelihood estimation. Least favorable families are closely related to the orthogonal parameterizations described by Cox and Reid (1987).

In some cases, the accuracy of the B–t method can be improved substantially by appropriately transforming the parameter of interest. Table 1 shows the improvement achieved in the correlation coefficient example by introducing the reparameterization $\phi = g(\rho) = \tanh^{-1}\rho$, for which $\mathrm{var}\{g(r);\rho\} \simeq n^{-1}$. The use of such variance-stabilizing transformations in connection with the B–t method is discussed in Section 2.3. The least favorable family construction is again involved in these transformations. Related transformations have been considered by Tibshirani (1988) and are also discussed in Section 2.3.

Approximate confidence limits having coverage error of order $O(n^{-1})$ can be obtained using the signed root of the usual or an adjusted likelihood ratio statistic. Generally, these confidence limits are not second-order accurate in the sense of Hall (1988), but instead, they lead naturally to a different definition of second-order accuracy. In Section 2.4, this alternative definition is described, and simulation methods that produce second-order accurate limits in the likelihood sense are discussed. In particular, it is shown that limits satisfying one definition of second-order accuracy can be transformed easily into limits that are second-order accurate according to the other definition.

Further examples are considered in Section 3. An appendix contains technical arguments that justify some of the discussion given in Section 2.

2. Resampling Procedures

2.1. 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 η based on a sample of size n. Suppose that $\theta = \theta(\eta)$ is the real-valued parameter of interest and that the standard deviation of the estimator $\hat{\theta} = \theta(\hat{\eta})$ is $n^{-1/2}\sigma(\eta) + O(n^{-3/2})$. Hall (1988) defined the exact studentized (ES) upper $(1-\alpha)$ confidence limit for θ to be $\hat{\theta}_{\mathrm{ES}}(1-\alpha) = \hat{\theta} - n^{-1/2}\hat{\sigma}K^{-1}(\alpha;\eta)$, where $\hat{\sigma} = \sigma(\hat{\eta})$ and $K(\cdot;\eta)$ is the distribution function of $n^{1/2}(\hat{\theta}-\theta)/\hat{\sigma}$. It is assumed that $K(\cdot;\eta)$ is strictly increasing and that $K(\cdot;\eta)$ and its inverse $K^{-1}(\cdot;\eta)$ have valid Cornish-Fisher expansions with error of order $O(n^{-1})$. Since $\mathrm{pr}\{K^{-1}(\alpha;\eta) \le n^{1/2}(\hat{\theta}-\theta)/\hat{\sigma};\eta\} = 1-\alpha$ by definition, it follows that $\mathrm{pr}\{\theta \le \hat{\theta}_{\mathrm{ES}}(1-\alpha);\eta\} = 1-\alpha$. An approximate upper $(1-\alpha)$ confidence limit $\hat{\theta}(1-\alpha)$ is said to be second-order accurate if it differs from $\hat{\theta}_{\mathrm{ES}}(1-\alpha)$ by order $O_p(n^{-3/2})$, in which case the error in coverage level for $\hat{\theta}(1-\alpha)$ is typically of order $O(n^{-1})$. Although the ES limit is usually unavailable in practice, it does serve to provide a convenient unifying definition of second-order accuracy.

The parametric bootstrap distribution of $n^{1/2}(\hat{\theta}-\theta)/\hat{\sigma}$ is $K(\cdot;\hat{\eta})$, and the B–t approximate upper $(1-\alpha)$ confidence limit for θ is $\hat{\theta}_{B\text{–}t}(1-\alpha) = \hat{\theta} - n^{-1/2}\hat{\sigma}K^{-1}(\alpha;\hat{\eta})$. The B–t method is easily seen to be second-order accurate since $K^{-1}(\alpha;\hat{\eta}) = K^{-1}(\alpha;\eta) + O_p(n^{-1})$. However, cases can arise, as in Table 1, where this method performs poorly.

To review Efron's (1981, 1985, 1987) percentile methods, let $G(\cdot;\eta)$ be the distribution function of the estimator $\hat{\theta}$, so that $G(t;\eta) = \mathrm{pr}(\hat{\theta} \le t;\eta)$. The parametric bootstrap distribution of $\hat{\theta}$ is $\hat{G}(\cdot) = G(\cdot;\hat{\eta})$. The BC$_a$ upper $(1-\alpha)$ confidence limit for θ is

$$\hat{\theta}_{\mathrm{BC}_a}(1-\alpha) = \hat{G}^{-1}\left[\Phi\left\{\frac{z_{1-\alpha}+Z}{1-A(z_{1-\alpha}+Z)} + Z\right\}\right], \tag{1}$$

where Φ is the standard normal distribution function and $z_{1-\alpha} = \Phi^{-1}(1-\alpha)$. The quantities Z and A that appear in (1) are called the bias correction and the acceleration constant, respectively. The BC limit $\hat{\theta}_{\mathrm{BC}}(1-\alpha) = \hat{G}^{-1}\{\Phi(z_{1-\alpha} + 2Z)\}$ is obtained from (1) by setting $A = 0$, and the S limit $\hat{\theta}_S(1-\alpha) = \hat{G}^{-1}(z_{1-\alpha})$ is obtained by setting $A = Z = 0$. Efron (1981) utilized transformation theory to derive the formula $Z = \Phi^{-1}\{\hat{G}(\hat{\theta})\}$. Although the bias correction Z can be obtained directly from the bootstrap distribution $\hat{G}(\cdot)$, the acceleration constant A cannot be so easily derived, and a theoretical calculation is typically required for its evaluation. Efron (1987) presented a formula for the computation of A for the case that $\hat{\eta}$ is the maximum likelihood estimator (MLE).

Some further notation is required to present explicit formulae for Z and A. Let $U^i = n^{1/2}(\hat{\eta}^i - \eta^i)$, $i = 1, \ldots, p$, and suppose that

$$E(U^i;\eta) = n^{-1/2}\kappa^i + O(n^{-1}), \quad \mathrm{cov}(U^i, U^j;\eta) = \kappa^{i,j} + O(n^{-1}),$$
$$\mathrm{cum}(U^i, U^j, U^k;\eta) = n^{-1/2}\kappa^{i,j,k} + O(n^{-1}), \tag{2}$$

where the κ's are of order $O(1)$. The fourth- and higher-order cumulants of $U^1, \ldots, U^p$ are assumed to be $O(n^{-1})$ or smaller. It is shown in Section A.1 that the BC_a method is second-order accurate if Z and A differ from $\bar{Z}$ and $\bar{A}$ by order $O_p(n^{-1})$, where

$$\bar{Z} = n^{-1/2}\{(\frac{1}{6}\kappa^{i,j,k} - \kappa^i\kappa^{j,k})\theta_i\theta_j\theta_k + (\frac{1}{2}\kappa^{i,k}\kappa^{j,l} - \frac{1}{2}\kappa^{i,j}\kappa^{k,l})\theta_i\theta_j\theta_{kl}\}/(\kappa^{i,j}\theta_i\theta_j)^{3/2},$$
$$\bar{A} = n^{-1/2}\{(\frac{1}{2}\kappa^{i,j}_l\kappa^{l,k} - \frac{1}{3}\kappa^{i,j,k})\theta_i\theta_j\theta_k\}/(\kappa^{i,j}\theta_i\theta_j)^{3/2}, \tag{3}$$

$\theta_i = \partial\theta(\eta)/\partial\eta^i$, $\theta_{ij} = \partial^2\theta(\eta)/\partial\eta^i\partial\eta^j$ and $\kappa^{i,j}_k = \partial\kappa^{i,j}(\eta)/\partial\eta^k$, $i, j, k = 1, \ldots, p$. These expressions employ the convention whereby summation is assumed over the range $1, \ldots, p$ for every index appearing both as a subscript and as a superscript. Note that $\bar{Z}$ and $\bar{A}$ are $O(n^{-1/2})$, and the choices $Z = \bar{Z}(\hat{\eta})$ and $A = \bar{A}(\hat{\eta})$ ensure second-order accuracy. It is shown in Section A.1 that Z differs from $\bar{Z}$ by $O_p(n^{-1})$.

When $\hat{\eta}$ is the MLE, expression (3) can be expressed in terms of moments of the derivatives of the log-likelihood function $L(\eta)$, based on the entire sample of size n. Let $L_i = \partial L(\eta)/\partial\eta^i$, $L_{ij} = \partial^2 L(\eta)/\partial\eta^i\partial\eta^j$, and let $\lambda_{i,j} = E(L_iL_j)$, $\lambda_{i,jk} = E(L_iL_{jk})$, $\lambda_{i,j,k} = E(L_iL_jL_k)$, etc. $(i, j, k = 1, \ldots, p)$. Note that the λ's are of order $O(n)$. The expected information matrix is $(\lambda_{i,j})$; denote its inverse by $(\lambda^{i,j})$. Second-order accuracy holds provided $Z = \bar{Z} + O_p(n^{-1})$ and $A = \bar{A} + O_p(n^{-1})$, where

$$\bar{Z} = \bar{A} + \frac{1}{2}\left\{(\lambda_{i,j,k} + \lambda_{i,jk})\mu^i - \theta_{jk}\right\}\left\{\lambda_{l,m}\mu^l\mu^m\lambda^{j,k} - \mu^j\mu^k\right\}/(\lambda_{i,j}\mu^i\mu^j)^{3/2},$$
$$\bar{A} = \frac{1}{6}(\lambda_{i,j,k}\mu^i\mu^j\mu^k)/(\lambda_{i,j}\mu^i\mu^j)^{3/2}, \tag{4}$$

and $\mu^i = \lambda^{i,j}\theta_j$. Efron (1987) gave essentially this formula for $\bar{A}$, but rather than recommend the use of $\bar{A}(\hat{\eta})$ for A, he suggested that A be obtained by replacing $\lambda_{i,j}$ in $\bar{A}$ by its observed counterpart $-L_{ij}(\hat{\eta})$.

2.2. Automatic percentile methods

When the distribution of $\hat{\theta}$ depends on η only through θ, it is typically possible to obtain an exact percentile (EP) upper $(1-\alpha)$ confidence limit $\hat{\theta}_{\mathrm{EP}}(1-\alpha)$ from the equation

$$G\left\{\hat{\theta}; \hat{\theta}_{\mathrm{EP}}(1-\alpha)\right\} = \alpha, \tag{5}$$

where $G(t;\theta) = \mathrm{pr}(\hat{\theta} \leq t; \theta)$. This limit is exact in the sense that $\mathrm{pr}\{\theta \leq \hat{\theta}_{\mathrm{EP}}(1-\alpha); \theta\} = 1-\alpha$. While the EP limit can be easily implemented in scalar parameter models with no nuisance parameters, it is usually not available in multiparameter settings. However, one situation involving nuisance parameters for which the EP limit can be constructed is the correlation coefficient example, since the distribution of r depends only on ρ. As in David (1954), for example, the EP limit is generally used for the exact one, and this practice is followed in Table 1.

The AP methods extend the familiar equation (5) for constructing confidence limits to cases where the distribution of $\hat{\theta}$ depends on nuisance parameters. They involve the least favorable family construction introduced by Stein (1956) and later used by Efron (1987). The least favorable direction $\delta = (\delta^1, \ldots, \delta^p)$ at a point η in the parameter space has components defined by $\delta^i = \kappa^{i,j}\theta_j/(\kappa^{i,j}\theta_i\theta_j)$, $i = 1, \ldots, p$. For an arbitrary point η_0 with associated parameter value $\theta_0 = \theta(\eta_0)$ and least favorable direction $\delta_0 = \delta(\eta_0)$, a least favorable family in the parameter space running through η_0 is a curve $\breve{\eta}(\theta;\eta_0)$ indexed by θ such that $\breve{\eta}(\theta_0;\eta_0) = \eta_0$ and

$$\left.\frac{\partial \breve{\eta}^i(\theta;\eta_0)}{\partial\theta}\right|_{\theta=\theta_0} = \delta_0^i + O(n^{-1/2}), \qquad i = 1, \ldots, p. \tag{6}$$

Least favorable families $\breve{\eta}(\theta;\hat{\eta})$ running through the estimator $\hat{\eta}$ are of particular importance for the AP methods. Of course, one such family is the line $\breve{\eta}(\theta;\hat{\eta}) = \hat{\eta} + (\theta - \hat{\theta})\hat{\delta}$, where $\hat{\delta} = \delta(\hat{\eta})$. When $\hat{\eta}$ is the MLE, another such family is $\breve{\eta}(\theta;\hat{\eta}) = \tilde{\eta}(\theta)$, where $\tilde{\eta}(\theta)$ is the constrained MLE of η for given θ. Note $\tilde{\eta}(\hat{\theta}) = \hat{\eta}$ and a calculation shows

$$\left.\frac{\partial \breve{\eta}(\theta)}{\partial\theta}\right|_{\theta=\hat{\theta}} = \frac{I^{ij}\hat{\theta}_j}{I^{ij}\hat{\theta}_i\hat{\theta}_j} = \hat{\delta}^i + O_p(n^{-1/2}),$$

where (I^{ij}) is the inverse of the observed Fisher information matrix (I_{ij}). A key feature of this family is that it transforms correctly under reparameterization. Generally, condition (6) implies that

$$\breve{\eta}^i(\theta;\hat{\eta}) = \hat{\eta} + (\theta - \hat{\theta})\hat{\delta} + O_p(n^{-1}), \qquad i = 1, \ldots, p,$$

for values of θ that differ from $\hat{\theta}$ by order $O_p(n^{-1/2})$.

For the purpose of generalizing (5), let $\breve{G}(t;\theta) = \mathrm{pr}\{\hat{\theta} \le t; \breve{\eta}(\theta;\hat{\eta})\}$. Note that $\breve{G}(\cdot;\hat{\theta}) = \hat{G}(\cdot)$ is the bootstrap distribution of $\hat{\theta}$. The automatic percentile (AP) approximate upper $(1-\alpha)$ confidence limit $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ is defined by the equation

$$\breve{G}\left\{\hat{\theta}; \hat{\theta}_{\mathrm{AP}}(1-\alpha)\right\} = \alpha. \tag{7}$$

The second-order accuracy of the AP method is demonstrated in Section A.2.

If the least favorable family used in (7) is parameterization invariant, as is the case for the constrained MLE $\tilde{\eta}(\theta)$, then the limit derived from (7) will also have the invariance property. However, the AP limit is not necessarily invariant.

It is of interest to consider the AP method in the orthogonal case. Suppose that $\eta = (\theta, \psi)$, where $\psi = (\eta^2, \ldots, \eta^p)$ is orthogonal to θ, that is, $\kappa^{1,i} = 0$, $i = 2, \ldots, p$. The least favorable direction at each point is the unit vector in direction θ, and a least favorable family through $\eta_0 = (\theta_0, \psi_0)$ is $\breve{\eta}(\theta;\eta_0) = (\theta, \psi_0)$. When this family is used, solving Equation (7) is identical to solving (5) with the nuisance parameter ψ fixed at $\hat{\psi}$, its estimated value.

Since solving Equation (7) for $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ can be computationally demanding, it is useful to have a straightforward method for approximating the solution. One approach to obtaining an approximate solution, described by DiCiccio and Romano (1988), proceeds as follows: let $\hat{\theta}_0(1-\alpha)$ be an initial approximation to $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ that differs from $\hat{\theta}$ by order $O_p(n^{-1/2})$, let $\hat{\theta}_0'(1-\alpha)$ satisfy $\breve{G}\{\hat{\theta}_0'(1-\alpha); \hat{\theta}_0(1-\alpha)\} = \alpha$, so that $\hat{\theta}_0'(1-\alpha)$ is the α quantile of $\hat{\theta}$ for the parameter value $\breve{\eta}\{\hat{\theta}_0(1-\alpha);\hat{\eta}\}$, and finally let

$$\hat{\theta}_1(1-\alpha) = \hat{G}^{-1}\left[\breve{G}\left\{\hat{\theta}_0(1-\alpha); \hat{\theta}_0'(1-\alpha)\right\}\right]. \tag{8}$$

It is shown in Section A.2 that $\hat{\theta}_1(1-\alpha)$ differs from $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ by $O_p(n^{-3/2})$, that is, $\hat{\theta}_1(1-\alpha)$ is second-order accurate. This algorithm requires that the distribution function of $\hat{\theta}$ be determined at the three values $\breve{\eta}\{\hat{\theta}_0(1-\alpha);\hat{\eta}\}$, $\breve{\eta}\{\hat{\theta}_0'(1-\alpha);\hat{\eta}\}$, and $\breve{\eta}(\hat{\theta};\hat{\eta}) = \hat{\eta}$ of the parameter η. If the choice $\hat{\theta}_0(1-\alpha) = \hat{\theta}$ is made, then only two distribution functions are required. Experience with numerical examples suggests, however, that the accuracy of $\hat{\theta}_1(1-\alpha)$ is improved by taking $\hat{\theta}_0(1-\alpha)$ to be a better approximate limit, and it is preferable to use $\hat{\theta}_{\mathrm{S}}(1-\alpha)$ or $\hat{\theta}_{\mathrm{BC}}(1-\alpha)$ for $\hat{\theta}_0(1-\alpha)$. Table 1 shows the limits produced by method (8) in the correlation coefficient example for various choices of $\hat{\theta}_0(1-\alpha)$.

Method (8) can be iterated in an obvious way. For cases where the EP limit given by (5) is available, DiCiccio and Romano (1989) showed that the approximate limit $\hat{\theta}_i(1-\alpha)$ obtained after the ith iteration typically differs from $\hat{\theta}_{\mathrm{EP}}(1-\alpha)$ by order $O_p(n^{-(i+2)/2})$. Consequently, $\hat{\theta}_i(1-\alpha)$ has coverage error of order $O(n^{-(i+1)/2})$ in such cases, since the coverage level of $\hat{\theta}_{\mathrm{EP}}(1-\alpha)$ is exactly

$1-\alpha$. These results indicate that iterating method (8) in general brings the resulting limit closer to the limit $\hat{\theta}_{\rm AP}(1-\alpha)$ given by (7). Indeed, if $\hat{\theta}_0(1-\alpha) = \hat{\theta}_{\rm AP}(1-\alpha)$, then $\hat{\theta}_0'(1-\alpha) = \hat{\theta}$ and $\hat{\theta}_1(1-\alpha) = \hat{\theta}_{\rm AP}(1-\alpha)$. However, since $\hat{\theta}_{\rm AP}(1-\alpha)$ usually has coverage error of order $O(n^{-1})$, iteration of (8) does not generally reduce the order of coverage error. Automatic percentile methods for nonparametric inference have been discussed by DiCiccio and Romano (1990).

2.3. Variance-stabilizing transformations

Consider an arbitrary point η_0 in the parameter space, having least favorable direction $\delta = \delta(\eta_0)$. Let $\theta_0 = \theta(\eta_0)$, and let $\breve{\eta}(\theta;\eta_0)$ be a least favorable family through η_0. Recall that the variance of $n^{1/2}(\hat{\theta}-\theta)$ at the point $\breve{\eta}(\theta;\eta_0)$ is

$$\mathrm{Var}\left\{n^{1/2}(\hat{\theta}-\theta);\breve{\eta}(\theta;\eta_0)\right\} = [\sigma\{\breve{\eta}(\theta;\eta_0)\}]^2 + O(n^{-1}),$$

and consider the reparameterization $\phi = \breve{g}(\theta;\eta_0)$ given by

$$\breve{g}(\theta;\eta_0) = \int^{\theta} [\sigma\{\breve{\eta}(u;\eta_0)\}]^{-1}\, du. \tag{9}$$

Transformation (9) is variance-stabilizing along the least favorable family in the sense

$$\mathrm{Var}\left\{n^{1/2}(\hat{\phi}-\phi);\breve{\eta}(\theta;\eta_0)\right\} = 1 + O(n^{-1})$$

for all θ, where $\hat{\phi} = \breve{g}(\hat{\theta};\eta_0)$. In particular, $\mathrm{Var}\{n^{1/2}(\hat{\phi}-\phi_0);\eta_0\} = 1 + O(n^{-1})$, where $\phi_0 = \breve{g}(\theta_0;\eta_0)$. If the lower limit of integration in (9) is θ_0, then $\phi_0 = 0$. Let $\breve{K}(\cdot;\eta_0)$ be the distribution function of $n^{1/2}(\hat{\phi}-\phi_0)$, that is, $\hat{K}(t;\eta_0) = \mathrm{pr}\{n^{1/2}(\hat{\phi}-\phi_0) \le t;\eta_0)$. Since $\mathrm{pr}\{\breve{K}^{-1}(\alpha;\eta_0) \le n^{1/2}(\hat{\phi}-\phi_0);\eta_0\} = 1-\alpha$, it follows that $\mathrm{pr}\{\phi_0 \le \hat{\phi} - n^{-1/2}\breve{K}^{-1}(\alpha;\eta_0);\eta_0\} = 1-\alpha$, and hence,

$$\mathrm{pr}\left[\theta_0 \le \breve{g}^{-1}\{\hat{\phi} - n^{-1/2}\breve{K}^{-1}(\alpha;\eta_0);\eta_0\};\eta_0\right] = 1-\alpha. \tag{10}$$

Expression (10) shows that the limit

$$\breve{g}^{-1}\left\{\hat{\phi} - n^{-1/2}\breve{K}^{-1}(\alpha;\eta_0);\eta_0\right\} \tag{11}$$

has coverage level exactly $1-\alpha$. Limit (11) typically differs from the exact studentized limit $\hat{\theta}_{\rm ES}(1-\alpha) = \hat{\theta} - n^{-1/2}\hat{\sigma}K^{-1}(\alpha;\eta_0)$ by order $O_p(n^{-3/2})$, and this discrepancy highlights the lack of parameterization invariance in $\hat{\theta}_{\rm ES}(1-\alpha)$.

A bootstrap version of (11) is the variance-stabilized bootstrap-t (VS) approximate upper $(1-\alpha)$ confidence limit, defined by

$$\hat{\theta}_{\rm VS}(1-\alpha) = \breve{g}^{-1}\left\{\hat{\phi} - n^{-1/2}\breve{K}^{-1}(\alpha;\hat{\eta});\hat{\eta}\right\}.$$

It is shown in Section A.3 that $\hat{\theta}_{\mathrm{VS}}(1-\alpha)$ is second-order accurate. In the correlation coefficient example with $\sigma(\eta_0) = 1-\rho^2$ for all η_0, (9) gives $\breve{g}(\rho;\eta_0) = \tanh^{-1}\rho$. Results obtained from using the B–t method for the parameter $\phi = \tanh^{-1}\rho$ are given in Table 1. In this example, the variance-stabilizing transformation significantly improves the B–t limits, so that $\hat{\rho}_{\mathrm{VS}}(1-\alpha)$ is on the whole preferable to $\hat{\rho}_{B\text{–}t}(1-\alpha)$.

Table 1. Approximate lower and upper 97.5% confidence limits for ρ ($n = 8; r = 0.5$)

Method	Lower limit	Upper limit
Exact (EP)	−0.2940 (2.50)	0.8663 (2.50)
B–t (ρ)	−1.0610	1.1248
S	−0.2716 (2.83)	0.8990 (1.12)
BC, BC$_a$	−0.3636 (1.65)	0.8781 (1.93)
(8) ($\hat{\rho}_0 = \hat{\rho}$)	−0.4157 (1.18)	0.8990 (1.12)
(8) ($\hat{\rho}_0 = \hat{\rho}_{\mathrm{S}}$)	−0.2986 (2.44)	0.8629 (2.68)
(8) ($\hat{\rho}_0 = \hat{\rho}_{\mathrm{BC}}$)	−0.2791 (2.72)	0.8651 (2.56)
B–t ($\phi = \tanh^{-1}\rho$)	−0.3528 (1.77)	0.8803 (1.83)

Each approximate lower limit $\hat{\rho}(0.025)$ is accompanied by $\mathrm{pr}\{r \geq 0.5; \rho = \hat{\rho}(0.025)\}$ as a percentage in parentheses; each approximate upper limit $\hat{\rho}(0.975)$ is accompanied by $\mathrm{pr}\{r \leq 0.5; \rho = \hat{\rho}(0.975)\}$.

Tibshirani (1988) has considered using the transformation

$$g(\theta;\eta_0) = \int^{\theta} \Big\{E(\hat{\sigma}^2 \mid \hat{\theta} = \mu; \eta_0)\Big\}^{-1/2} du \tag{12}$$

in place of (9) to construct $\hat{\theta}_{\mathrm{VS}}(1-\alpha)$. This transformation has the advantage of not requiring direct use of a least favorable family; however, estimating the integrand by simulation, as Tibshirani suggests, can be computationally demanding. It is shown in Section A.3 that the bootstrap-t method based on transformation (12) is second-order accurate. In the case of the correlation coefficient, (12) also gives the $\tanh^{-1}\rho$ transformation.

2.4. Likelihood-based methods

The foregoing discussion has focused on methods that use $n^{-1/2}\sigma$ to standardize $\hat{\theta} - \theta$. Another possibility, which is motivated primarily by considerations of conditional inference, but also has the advantage of avoiding calculation of σ, is to standardize $\hat{\theta} - \theta$ by $(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}$, where $\hat{\theta}_i = \theta_i(\hat{\eta})$. It is natural then to define an exact likelihood (EL) upper $(1-\alpha)$ confidence limit for θ as

$\hat{\theta}_{\rm EL}(1-\alpha) = \hat{\theta} - (I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}J^{-1}(\alpha;\eta)$, where $J(w;\eta) = \text{pr}\{(\hat{\theta}-\theta)/(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2} \leq w;\eta\}$. In general, $\hat{\theta}_{\rm EL}(1-\alpha)$ differs from $\hat{\theta}_{\rm ES}(1-\alpha)$ by order $O_p(n^{-1})$, since $(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2} = n^{-1/2}\hat{\sigma} + O_p(n^{-1})$. Thus, an approximate limit $\hat{\theta}(1-\alpha)$ that differs from $\hat{\theta}_{\rm EL}(1-\alpha)$ by order $O_p(n^{-3/2})$ is not usually second-order accurate in the sense previously defined, although it does have coverage error $O(n^{-1})$. However, in cases where $(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2} = n^{-1/2}\hat{\sigma} + O_p(n^{-3/2})$, the exact limits $\hat{\theta}_{\rm EL}(1-\alpha)$ and $\hat{\theta}_{\rm ES}(1-\alpha)$ agree to error of order $O_p(n^{-3/2})$, and hence, they lead to equivalent definitions of second-order accuracy. This situation arises in exponential families, for example.

A bootstrap version of $\hat{\theta}_{\rm EL}(1-\alpha)$ is $\hat{\theta}_{\rm BL}(1-\alpha) = \hat{\theta} - (I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}J^{-1}(\alpha;\hat{\eta})$. Since $J^{-1}(\alpha;\hat{\eta}) = J^{-1}(\alpha;\eta) + O_p(n^{-1})$, it follows that $\hat{\theta}_{\rm BL}(1-\alpha)$ differs from $\hat{\theta}_{\rm EL}(1-\alpha)$ by order $O_p(n^{-3/2})$ and has coverage error of order $O(n^{-1})$.

There has been considerable interest, recently, in the use of profile and adjusted profile likelihoods to construct approximate confidence limits. The log profile likelihood function is $L\{\tilde{\eta}(\theta)\}$ and attains its maximum value at $\hat{\theta}$. To account for the presence of nuisance parameters when making inferences about θ, various authors, including Barndorff-Nielsen (1983, 1994), Cox and Reid (1987), Kass, Tierney and Kadane (1989), McCullagh and Tibshirani (1990), and Barndorff-Nielsen and Chamberlin (1992), have recommended that the log profile likelihood function be replaced by an objective function of the form $M(\theta) = L\{\tilde{\eta}(\theta)\} + B(\theta)$, where the adjustment function $B(\theta)$ is of order $O_p(1)$ for each θ. The proposed adjustment functions typically have the effect of reducing the bias in the equation for estimating θ; usually, $E[\partial L\{\tilde{\eta}(\theta)\}/\partial\theta]$ is of order $O(1)$, while $E\{\partial M(\theta)/\partial\theta\}$ is of order $O(n^{-1})$. The point $\bar{\theta}$ that maximizes $M(\theta)$ satisfies $\bar{\theta} = \hat{\theta} + O_p(n^{-1})$.

An adjusted likelihood ratio statistic can be defined in terms of the objective function $M(\theta)$. Let $Q(\theta) = 2\{M(\bar{\theta}) - M(\theta)\}$; in wide generality, $Q(\theta)$ is distributed asymptotically as $\chi^2_{(1)}$. The signed root of $Q(\theta)$ is defined by $R(\theta) = \text{sgn}(\bar{\theta}-\theta)\{Q(\theta)\}^{1/2}$. Now, let $m(\eta)$ satisfy $E\{R(\theta)\} = m(\eta) + O(n^{-3/2})$. Then $m(\eta)$ is of order $O(n^{-1/2})$, and the quantity $\hat{\theta}_{\rm SR}(1-\alpha)$ that satisfies the equation

$$R\left\{\hat{\theta}_{\rm SR}(1-\alpha)\right\} + m[\tilde{\eta}\{\hat{\theta}_{\rm SR}(1-\alpha)\}] = z_\alpha \tag{13}$$

is an approximate upper $(1-\alpha)$ confidence limit for θ having coverage error of order $O(n^{-1})$ and differs from $\hat{\theta}_{\rm EL}(1-\alpha)$ by order $O_p(n^{-3/2})$. The particular choice of adjustment function $B(\theta)$ affects the limit $\hat{\theta}_{\rm SR}(1-\alpha)$ at the order $O_p(n^{-3/2})$ level.

Barndorff-Nielsen (1986) showed that the limit $\hat{\theta}_{\rm SR}(1-\alpha)$ obtained when $B(\theta) = 0$ has conditional coverage level $1-\alpha+O(n^{-1})$ given exactly or approximately ancillary statistics (see also McCullagh (1984)). Thus, the exact limit $\hat\theta_{\rm EL}(1-\alpha)$ also has conditional coverage error $O(n^{-1})$, as do the limits $\hat\theta_{\rm SR}(1-\alpha)$ obtained from Equation (13) for general adjustment functions $B(\theta)$. Because of this conditional validity, it can be argued that using $\hat\theta_{\rm EL}(1-\alpha)$ to define second-order accuracy is preferable to using $\hat\theta_{\rm ES}(1-\alpha)$. However, an approximate limit that is second-order accurate with respect to $\hat\theta_{\rm ES}(1-\alpha)$ can easily be converted to a limit that is second-order accurate with respect tp $\hat\theta_{\rm EL}(1-\alpha)$, and vice versa. It is shown in Section A.4 that if $\hat\theta(1-\alpha)$ is an approximate upper $(1-\alpha)$ confidence limit that differs from $\hat\theta_{\rm ES}(1-\alpha)$ by order $O_p(n^{-3/2})$, then

$$\hat\theta + \left\{\hat\theta(1-\alpha)-\hat\theta\right\}\frac{(I^{ij}\hat\theta_i\hat\theta_j)^{1/2}}{n^{-1/2}\hat\sigma} \tag{14}$$

differs from $\hat\theta_{\rm EL}(1-\alpha)$ by order $O_p(n^{-3/2})$. Similarly, if $\hat\theta(1-\alpha)$ is second-order accurate with respect to $\hat\theta_{\rm EL}(1-\alpha)$, then

$$\hat\theta + \left\{\hat\theta(1-\alpha)-\hat\theta\right\}\frac{n^{-1/2}\hat\sigma}{(I^{ij}\hat\theta_i\hat\theta_j)^{1/2}}$$

is second-order accurate with respect to $\hat\theta_{\rm ES}(1-\alpha)$.

Results of DiCiccio and Martin (1993) show that the approximate confidence limits obtained from Bayesian calculations, as described by Welch and Peers (1963), Peers (1965), and Stein (1985), are also second-order accurate with respect to $\hat\theta_{\rm EL}(1-\alpha)$.

3. Further Examples

Example 1. Location-scale family. Consider a location-scale family indexed by $\eta=(\theta,\tau)$, where θ and τ are the location and scale parameters, respectively. Let $\hat\theta$ and $\hat\tau$ be equivariant estimators. Since $(\hat\theta-\theta)/\hat\tau$ is pivotal in this example, $\hat\theta_{B-t}(1-\alpha)$ coincides with $\hat\theta_{\rm ES}(1-\alpha)$. It is of interest for comparison to consider the other procedures in this situation. Expression (3) pertaining to the bias correction and the acceleration constant of the BC$_a$ method reduces to simple formulae. There exist constants ν_1, ν_{11}, ν_{12}, ν_{22}, and ν_{111} such that, in the notation of (2), $\kappa^1=\nu_1\tau$, $\kappa^{1,1}=\nu_{11}\tau^2$, $\kappa^{1,2}=\nu_{12}\tau^2$, $\kappa^{2,2}=\nu_{22}\tau^2$, and $\kappa^{1,1,1}=\nu_{111}\tau^3$. It is natural to take $\sigma=\nu_{11}^{1/2}\tau$, and (3) gives

$$\bar Z = n^{-1/2}(\frac{1}{6}\nu_{111}-\nu_1\nu_{11})/\nu_{11}^{3/2}, \quad \bar A = n^{-1/2}(\nu_{11}\nu_{12}-\frac{1}{3}\nu_{111})/\nu_{11}^{3/2}. \tag{15}$$

For any point $\eta_0=(\theta_0,\tau_0)$, the least favorable direction is $(1,\nu_{12}\nu_{11}^{-1})$, and hence a least favorable family through η_0 is $\breve\eta(\theta;\eta_0)=\{\theta,\tau_0+(\theta-\theta_0)\nu_{12}/\nu_{11}\}$. In terms of this family, the solution to Equation (7) is

$$\hat\theta_{\rm AP}(1-\alpha)=\hat\theta-\hat\tau\left\{\frac{D^{-1}(\alpha)}{1+D^{-1}(\alpha)\nu_{12}/\nu_{11}}\right\}, \tag{16}$$

where $D(\cdot)$ is the distribution function of the pivotal quantity $(\hat{\theta}-\theta)/\tau$. Furthermore, the limit $\hat{\theta}_1(1-\alpha)$ obtained from (8) coincides with (16), regardless of the value chosen for $\hat{\theta}_0(1-\alpha)$. Moreover, if the parameter ψ is defined by $\tau=(\theta+\psi)\nu_{12}/\nu_{11}$, then $\eta=(\theta,\psi)$ is an orthogonal parameterization, and with least favorable family $\breve{\eta}(\theta;\eta_0)=(\theta,\psi_0)$, the limits produced by methods (7) and (8) are the same as (16).

The variance-stabilizing transformation (9) along the least favorable family $\breve{\eta}(\theta;\eta_0)=\{\theta,\tau_0+(\theta-\theta_0)\nu_{12}/\nu_{11}\}$ is

$$\breve{g}(\theta;\eta_0)=\frac{\nu_{11}^{1/2}}{\nu_{12}}\log\left\{1+\frac{(\theta-\theta_0)}{\tau_0}\frac{\nu_{12}}{\nu_{11}}\right\},$$

and the limit $\hat{\theta}_{\mathrm{VS}}(1-\alpha)$ based on this transformation also equals (16).

Table 2. Central 90% confidence intervals for θ in Efron's (1987) example

Exact (B–t)	$(-0.336, 0.670)$
S	$(-0.371, 0.444)$
BC	$(-0.339, 0.499)$
BC$_a$	$(-0.304, 0.601)$
(16)	$(-0.304, 0.601)$

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

$$\frac{(\hat{\theta}-\theta)}{\tau}\sim\frac{\chi^2_{(2n)}}{2n}-1,\qquad \frac{\hat{\tau}}{\tau}\Big|\frac{(\hat{\theta}-\theta)}{\tau}\sim\left\{1+\frac{(\hat{\theta}-\theta)}{\tau}\right\}\left(\frac{\chi^2_{(n-1)}}{n}\right)^{1/2},\tag{17}$$

for $n=15$. In this case, $\nu_1=0$, $\nu_{11}=\nu_{12}=1$, $\nu_{22}=1/2$, and $\nu_{111}=2$, and it follows from (15) that $\bar{A}=\bar{Z}=\frac{1}{3}n^{-1/2}$. Table 2 shows the 90% confidence intervals for θ obtained by various bootstrap procedures having observed $\hat{\theta}=0$, and $\hat{\tau}=(14/15)^{1/2}$. The BC$_a$ method and (16) produce very similar results in this example. It follows from (17) that

$$E(\hat{\sigma}^2|\hat{\theta}=u;\eta_0)=\left(\frac{n-1}{n}\right)\{\sigma_0+(u-\theta_0)\}^2,\tag{18}$$

and hence (12) yields the transformation

$$g(\theta;\eta_0)=\left(\frac{n}{n-1}\right)^{1/2}\log\left\{1+\frac{\theta-\theta_0}{\tau_0}\right\}.$$

Tibshirani's (1988) version of the variance-stabilized bootstrap-t procedure, based on this transformation, produces the same limits as (16). Tibshirani approximated the function (17) by simulation, and he reported the approximate 90%

interval as $(-0.334, 0.610)$. The difference between this interval and the one from (16) presented in Table 2 is presumably attributable to simulation error.

Table 3. Central 95% confidence intervals for $\theta = \mu^1/\mu^2$ ($n = 1; x = 4, y = 8$)

Exact (Fieller)	$(0.2476, 0.8162)$
AP ($\mu^1 = \eta^1, \mu^2 = \eta^2$)	$(0.2531, 0.8076)$
AP ($\mu^1 = \eta^1\eta^2, \mu^2 = \eta^2$)	$(0.2413, 0.8261)$
AP ($\breve{\eta}(\theta;\hat{\eta}) = \tilde{\eta}(\theta)$)	$(0.2509, 0.8131)$

Example 2. Ratio of normal means. Consider a sample (x_i, y_i) $(i = 1, \ldots, n)$ from a bivariate normal distribution having mean $\eta = (\mu^1, \mu^2)$ and identity covariance matrix, and suppose that $\theta = \mu^1/\mu^2$ is the parameter of interest. Efron (1985, 1987) discussed bootstrap confidence limits for θ based on the MLE $\hat{\theta} = \bar{x}/\bar{y}$, and he specifically considered the case where $n = 1$ and $(4, 8)$ is observed. This situation is assumed for Table 3, which reports a variety of approximate 95% confidence intervals for θ. The exact limits given in Table 3 are the Fieller ones, which arise from the standard normal distribution of the pivotal quantity $n^{1/2}(\bar{x} - \theta\bar{y})/(1 + \theta^2)^{1/2}$. Formulae (4) gives $\bar{Z} = \bar{A} = 0$, so that the S, BC and BC_a methods all coincide, and they are all second-order accurate. For the observation $(4, 8)$, these methods give the Fieller limits. 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 by using the least favorable family $\breve{\eta}(\theta; \eta_0) = (\theta, \psi_0)$, the AP limit from (7) also agrees with Fieller limit.

To illustrate the results of the AP method for other parameterizations, Table 3 also shows AP intervals obtained using the parameterizations $\eta = (\eta^1, \eta^2)$ defined by $\mu^1 = \eta^1$, $\mu^2 = \eta^2$ and $\mu^1 = \eta^1\eta^2$, $\mu^2 = \eta^2$. For both parameterizations, the least favorable family $\breve{\eta}(\theta; \eta_0) = \eta_0 + (\theta - \theta_0)\delta_0$ is used, where $\delta^i = \lambda^{i,j}\theta_j/(\lambda^{i,j}\theta_i\theta_j)$. In addition, Table 3 reports AP intervals derived using the parameterization invariant least favorable family $\breve{\eta}(\theta; \hat{\eta}) = \tilde{\eta}(\theta)$. All three of these intervals are quite close for this example.

The convergence of $\hat{\theta}_i(1 - \alpha)$ to $\hat{\theta}_{\mathrm{AP}}(1 - \alpha)$ is illustrated for this example in Table 4. For Table 4, the parameterization $\eta = (\eta^1, \eta^2) = (\mu^1, \mu^2)$ with least favorable family $\breve{\eta}(\theta; \eta_0) = \eta_0 + (\theta - \theta_0)\delta_0$ is used, and $\hat{\theta}_0(1 - \alpha)$ is taken to be $\hat{\theta}$. The convergence is quite rapid, and $\hat{\theta}_3(1 - \alpha)$ is nearly identical to $\hat{\theta}_{\mathrm{AP}}(1 - \alpha)$.

Table 4. Iterations of method (8) for $\theta = \mu^1/\mu^2$ ($\eta^1 = \mu^1, \eta^2 = \mu^2$)

$1 - \alpha$	$\hat{\theta}_0(1 - \alpha)$	$\hat{\theta}_1(1 - \alpha)$	$\hat{\theta}_2(1 - \alpha)$	$\hat{\theta}_3(1 - \alpha)$	$\hat{\theta}_{\mathrm{AP}}(1 - \alpha)$
0.025	0.5	0.2417	0.2537	0.2531	0.2531
0.975	0.5	0.8256	0.8068	0.8077	0.8076

Appendix: Technical Justifications

A.1. Bootstrap-t and BC$_a$ methods

It follows from (2) that the variance of $\hat{\theta}$ is $n^{-1}\kappa^{i,j}\theta_i\theta_j + O(n^{-2})$, so that $\sigma = (\kappa^{i,j}\theta_i\theta_j)^{1/2} + O(n^{-1})$. Moreover, with error of order $O(n^{-1})$, the first three cumulants of $U = n^{1/2}(\hat{\theta} - \theta)/\sigma$ are

$$E(U;\eta) = n^{-1/2}a, \quad \mathrm{Var}(U;\eta) = 1, \quad \mathrm{skew}(U;\eta) = n^{-1/2}c,$$

where

$$a = (\kappa^i\theta_i + \frac{1}{2}\kappa^{i,j}\theta_{ij})b^{-1}, \quad b = (\kappa^{i,j}\theta_i\theta_j)^{1/2},$$
$$c = (\kappa^{i,j,k}\theta_i\theta_j\theta_k + 3\kappa^{i,k}\kappa^{j,l}\theta_i\theta_j\theta_{kl})b^{-3}. \tag{19}$$

It is assumed that the distribution function $H(\cdot;\eta)$ of U satisfies

$$H(u;\eta) = \mathrm{pr}\{n^{1/2}(\hat{\theta} - \theta)/\sigma \le u;\eta\} = \Phi[u - n^{-1/2}\{(a - \frac{1}{6}c) + \frac{1}{6}cu^2\}] + O(n^{-1}).$$

Similarly, with error of order $O(n^{-1})$, the first three cumulants of $V = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ are

$$E(V;\eta) = n^{-1/2}(a - d), \qquad \mathrm{Var}(V;\eta) = 1, \quad \mathrm{skew}(V;\eta) = n^{-1/2}(c - 6d),$$

where $d = (\kappa^{i,j}b_i\theta_j)b^{-2}$ and $b_i = \partial b(\eta)/\partial\eta^i$. Straightforward calculation shows that

$$d = (\frac{1}{2}\kappa_l^{i,j}\kappa^{l,k}\theta_i\theta_j\theta_k + \kappa^{i,k}\kappa^{j,l}\theta_i\theta_j\theta_{kl})b^{-3}, \tag{20}$$

where $\kappa_k^{i,j} = \partial\kappa^{i,j}(\eta)/\partial\eta^k$. It is assumed the distribution function $K(\cdot;\eta)$ of V satisfies

$$K(v;\eta) = \mathrm{pr}\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \le v;\eta\}$$
$$= \Phi[v - n^{-1/2}\{(a - \frac{1}{6}c) + (\frac{1}{6}c - d)v^2\}] + O(n^{-1}).$$

The inverse functions $H^{-1}(\alpha;\eta)$ and $K^{-1}(\alpha;\eta)$ have the Cornish-Fisher expansions

$$H^{-1}(\alpha;\eta) = z_\alpha + n^{-1/2}\{(a - \frac{1}{6}c) + \frac{1}{6}cz_\alpha^2\} + O(n^{-1}),$$
$$K^{-1}(\alpha;\eta) = z_\alpha + n^{-1/2}\{(a - \frac{1}{6}c) + (\frac{1}{6}c - d)z_\alpha^2\} + O(n^{-1}), \tag{21}$$

so that, in particular, $K^{-1}(\alpha;\eta) = H^{-1}(\alpha;\eta) - n^{-1/2}dz_\alpha^2 + O(n^{-1})$. The ES upper $(1-\alpha)$ confidence limit for θ is

$$\begin{aligned}\hat\theta_{\mathrm{ES}}(1-\alpha) &= \hat\theta - n^{-1/2}\hat\sigma K^{-1}(\alpha;\eta)\\ &= \hat\theta - n^{-1/2}\hat\sigma\{H^{-1}(\alpha;\eta) - n^{-1/2}dz_\alpha^2\} + O_p(n^{-3/2}). \qquad (22)\end{aligned}$$

Since a, b, c and d are all of order $O(1)$, the estimates $\hat a = a(\hat\eta), \ldots, \hat d = d(\hat\eta)$ differ by terms of order $O_p(n^{-1/2})$ from $a, \ldots, d$, respectively. It then follows from (21) that $H^{-1}(\alpha;\hat\eta) = H^{-1}(\alpha;\eta) + O_p(n^{-1})$ and $K^{-1}(\alpha;\hat\eta) = K^{-1}(\alpha;\eta) + O_p(n^{-1})$. This observation for $K^{-1}(\alpha;\hat\eta)$ demonstrates the second-order accuracy of the B–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 by (3), note first that

$$G(t;\eta) = \mathrm{pr}(\hat\theta \le t;\eta) = H\{n^{1/2}(t-\theta)/\sigma;\eta\}, \quad G^{-1}(\alpha;\eta) = \theta + n^{-1/2}\sigma H^{-1}(\alpha;\eta).$$

Then, definition (1) gives

$$\begin{aligned}\hat\theta_{\mathrm{BC}_a}(1-\alpha) &= \hat\theta + n^{-1/2}\hat\sigma H^{-1}\left[\Phi\Big\{\frac{z_{1-\alpha}+Z}{1-A(z_{1-\alpha}+Z)} + Z\Big\};\hat\eta\right]\\ &= \hat\theta - n^{-1/2}\hat\sigma[H^{-1}(\alpha;\eta) - 2\{Z + n^{-1/2}(a-\frac{1}{6}c)\} - (A + \frac{1}{3}n^{-1/2}c)z_\alpha^2] + O_p(n^{-3/2}).\end{aligned}$$

Comparison with (22) shows that $\hat\theta_{\mathrm{BC}_a}(1-\alpha)$ differs from $\hat\theta_{\mathrm{ES}}(1-\alpha)$ by order $O_p(n^{-3/2})$ when $Z = \bar Z + O_p(n^{-1})$ and $A = \bar A + O_p(n^{-1})$, where

$$\bar Z = -n^{-1/2}(a - \frac{1}{6}c), \quad \bar A = -n^{-1/2}(\frac{1}{3}c - d). \qquad (23)$$

Substituting expressions (20) and (19) for a, c and d into (23) yields (3).

The second-order accuracy of the BC_a and B–t methods has been discussed in the context of exponential families by DiCiccio and Efron (1992). The accuracy of BC_a limits has also been considered by Bickel (1987).

Efron's (1981) version of Z is easily seen to suffice for second-order accuracy, since

$$\begin{aligned}\Phi^{-1}\{\hat G(\hat\theta)\} &= \Phi^{-1}\{H(0;\hat\eta)\} = \Phi^{-1}\{H(0;\eta)\} + O_p(n^{-1})\\ &= -\,n^{-1/2}(a - \frac{1}{6}c) + O_p(n^{-1}).\end{aligned}$$

In the case of MLEs, the cumulant formulae of McCullagh (1987) p.209 show that

$$\begin{aligned}&\kappa^{i,j} = n\lambda^{i,j}, \quad \kappa^{i,j}\theta_i\theta_j = n\lambda_{i,j}\mu^i\mu^j, \quad \kappa^i\theta_i = -n(\frac{1}{2}\lambda_{i,j,k} + \frac{1}{2}\lambda_{i,jk})\mu^i\lambda^{j,k},\\ &\kappa^{i,j,k}\theta_i\theta_j\theta_k = -n^2(2\lambda_{i,j,k} + 3\lambda_{i,jk})\mu^i\mu^j\mu^k, \quad \kappa^{i,j}\kappa^{j,l}\theta_i\theta_j\theta_{kl} = n^2\theta_{ij}\mu^i\mu^j,\\ &\kappa_l^{i,j}\kappa^{l,k}\theta_i\theta_j\theta_k = -n^2(\lambda_{i,j,k} + 2\lambda_{i,jk})\mu^i\mu^j\mu^k. \qquad (24)\end{aligned}$$

Substitution of formulae (24) into (3) yields (4).

A.2. Automatic percentile methods

To verify that the AP method is second-order accurate, recall that $\sigma = b + O(n^{-1})$ and $d = b_i \delta^i$. Hence, for an arbitrary point η_0,

$$\left.\frac{\partial \sigma\{\breve{\eta}(\theta;\eta_0)\}}{\partial \theta}\right|_{\theta=\theta_0} = b_i(\eta_0)\delta^i(\eta_0) + O(n^{-1/2}) = d(\eta_0) + O(n^{-/2}), \tag{25}$$

and for values of θ such that $\hat{\theta} - \theta$ is of order $O_p(n^{-1/2})$, we have

$$\sigma\{\breve{\eta}(\theta;\hat{\eta})\} = \hat{\sigma} + (\theta - \hat{\theta})d + O_p(n^{-1}). \tag{26}$$

The limit $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ defined by Equation (8) satisfies

$$\hat{\theta} = \breve{G}^{-1}\{\alpha;\hat{\theta}_{\mathrm{AP}}(1-\alpha)\} = \hat{\theta}_{\mathrm{AP}}(1-\alpha) + n^{-1/2}\sigma\{\breve{\eta}_{\mathrm{AP}}(1-\alpha)\}H^{-1}\{\alpha;\breve{\eta}_{\mathrm{AP}}(1-\alpha)\},$$

where $\breve{\eta}_{\mathrm{AP}}(1-\alpha) = \breve{\eta}\{\breve{\theta}_{\mathrm{AP}}(1-\alpha);\hat{\eta}\}$, that is,

$$\hat{\theta}_{\mathrm{AP}}(1-\alpha) = \hat{\theta} - n^{-1/2}\sigma\{\breve{\eta}_{\mathrm{AP}}(1-\alpha)\}H^{-1}\{\alpha;\breve{\eta}_{\mathrm{AP}}(1-\alpha)\}. \tag{27}$$

Since (26) implies $\hat{\theta}_{\mathrm{AP}}(1-\alpha) = \hat{\theta} - n^{-1/2}\hat{\sigma}z_\alpha + O_p(n^{-1})$, it follows from (21) that $H^{-1}\{\alpha;\breve{\eta}_{\mathrm{AP}}(1-\alpha)\} = H^{-1}(\alpha;\eta) + O_p(n^{-1})$ and from (26) that

$$\sigma\{\breve{\eta}_{\mathrm{AP}}(1-\alpha)\} = \hat{\sigma} - n^{-1/2}\hat{\sigma}z_\alpha d + O_p(n^{-1}).$$

Therefore, (27) gives

$$\begin{aligned}\hat{\theta}_{\mathrm{AP}}(1-\alpha) &= \hat{\theta} - n^{-1/2}\hat{\sigma}(1 - n^{-1/2}z_\alpha d)H^{-1}(\alpha;\eta) + O_p(n^{-3/2}) \\ &= \hat{\theta} - n^{-1/2}\hat{\sigma}\{H^{-1}(\alpha;\eta) - n^{-1/2}dz_\alpha^2\} + O_p(n^{-3/2}),\end{aligned} \tag{28}$$

and comparison of (28) with (21) shows that $\hat{\theta}_{\mathrm{AP}}(1-\alpha)$ is second-order accurate.

Now consider the limit $\hat{\theta}_1(1-\alpha)$ defined by Equation (8). Suppose that $\hat{\theta}_0(1-\alpha)$ differs from $\hat{\theta}$ by $O_p(n^{-1/2})$, and let $\breve{\eta}_0(1-\alpha) = \breve{\eta}\{\hat{\theta}_0(1-\alpha);\hat{\eta}\}$. Then,

$$\hat{\theta}_0'(1-\alpha) = \breve{G}^{-1}\{\alpha;\hat{\theta}_0(1-\alpha)\} = \hat{\theta}_0(1-\alpha) + n^{-1/2}\sigma\{\breve{\eta}_0(1-\alpha)\}H^{-1}\{\alpha;\breve{\eta}_0(1-\alpha)\},$$

and using (26) gives

$$\begin{aligned}\hat{\theta}_0'(1-\alpha) - \hat{\theta}_0(1-\alpha) &= n^{-1/2}[\hat{\sigma} + \{\hat{\theta}_0(1-\alpha) - \hat{\theta}\}d]H^{-1}(\alpha;\eta) + O_p(n^{-3/2}) \\ &= n^{-1/2}\hat{\sigma}z_\alpha + O_p(n^{-1}).\end{aligned} \tag{29}$$

Now let $\breve{\eta}_0'(1-\alpha) = \breve{\eta}\{\hat{\theta}_0'(1-\alpha);\hat{\eta}\}$, and note from (29) that

$$\begin{aligned}\breve{G}\{\hat{\theta}_0(1-\alpha);\hat{\theta}_0'(1-\alpha)\} =& H\left[\frac{n^{1/2}\{\hat{\theta}_0(1-\alpha)-\hat{\theta}_0'(1-\alpha)\}}{\sigma\{\breve{\eta}_0'(1-\alpha)\}};\breve{\eta}_0'(1-\alpha)\right] \\ =& H\{-(1-n^{-1/2}dz_\alpha)H^{-1}(\alpha;\eta);\eta\} + O_p(n^{-1}). \quad (30)\end{aligned}$$

Since

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

it follows from (30) that

$$\begin{aligned}\hat{\theta}_1(1-\alpha) &= \hat{G}^{-1}[\breve{G}\{\hat{\theta}_0(1-\alpha);\hat{\theta}_0'(1-\alpha)\}] \\ &= \hat{\theta} - n^{-1/2}\{H^{-1}(\alpha;\eta) - n^{-1/2}dz_\alpha^2\} + O_p(n^{-3/2}). \quad (31)\end{aligned}$$

Comparison of (31) with (21) establishes the second-order accuracy of $\hat{\theta}_1(1-\alpha)$.

A.3. Variance-stabilizing transformations

By (25), the transformation $\breve{g}(\theta;\eta_0)$ defined at (9) has derivatives

$$\frac{\partial\breve{g}(\theta;\eta_0)}{\partial\theta}\Big|_{\theta=\theta_0} = \frac{1}{\sigma_0}, \quad \frac{\partial^2\breve{g}(\theta;\eta_0)}{\partial\theta^2}\Big|_{\theta=\theta_0} = -\frac{d_0}{\sigma_0^2} + O(n^{-1/2}), \quad (32)$$

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

$$\breve{K}(t;\eta_0) = \mathrm{pr}\{n^{1/2}(\hat{\phi}-\phi_0) \le t;\eta_0\} = H\{t + \frac{1}{2}n^{-1/2}d_0t^2;\eta_0\} + O(n^{-1}),$$

and hence, $\breve{K}^{-1}(\alpha;\eta_0) = H^{-1}(\alpha;\eta_0) - \frac{1}{2}n^{-1/2}d_0z_\alpha^2 + O(n^{-1})$. Thus,

$$\breve{K}^{-1}(\alpha;\hat{\eta}) = H^{-1}(\alpha;\eta) - \frac{1}{2}n^{-1/2}dz_\alpha^2 + O_p(n^{-1}). \quad (33)$$

The derivatives (32) yield

$$\frac{\partial\breve{g}^{-1}(\phi;\eta_0)}{\partial\phi}\Big|_{\phi=\phi_0} = \sigma_0, \quad \frac{\partial^2\breve{g}^{-1}(\phi;\eta_0)}{\partial\phi^2}\Big|_{\phi=\phi_0} = \sigma_0d_0 + O(n^{-1/2}),$$

so that

$$g_{\eta_0}^{-1}(\phi) = \theta_0 + (\phi-\phi_0)\sigma_0 + \frac{1}{2}(\phi-\phi_0)^2\sigma_0d_0 + O(n^{-3/2}) \quad (34)$$

for values of ϕ that differ from ϕ_0 by order $O(n^{-1/2})$. By (33) and (34), the variance-stabilized bootstrap-t limit satisfies

$$\begin{aligned}\hat{\theta}_{\mathrm{VS}}(1-\alpha) &= \breve{g}^{-1}\{\hat{\phi} - n^{-1/2}\breve{K}^{-1}(\alpha;\hat{\eta});\hat{\eta}\} \\ &= \hat{\theta} - n^{-1/2}\hat{\sigma}\{H^{-1}(\alpha;\eta) - n^{-1/2}dz_\alpha^2\} + O_p(n^{-3/2}). \qquad (35)\end{aligned}$$

Comparison of (35) with (21) shows that $\hat{\theta}_{\mathrm{VS}}(1-\alpha)$ is second-order accurate.

To investigate the transformation $g(\theta;\eta_0)$ defined at (12), note that

$$E(\hat{\theta};\eta_0) = \theta_0 + O(n^{-1}), \quad E(\hat{\sigma}^2;\eta_0) = \sigma_0^2 + O(n^{-1}),$$

$$\mathrm{Cov}(\hat{\theta},\hat{\sigma}^2;\eta_0) = n^{-1}2\sigma_0^3 d_0 + O(n^{-1}),$$

and the standard formula for conditional expectation (McCullagh (1987), p.164) gives

$$E_{\eta_0}(\hat{\sigma}^2|\hat{\theta}=\theta) = \sigma_0^2 + 2(\theta-\theta_0)\sigma_0 d_0 + O(n^{-2})$$

for values of θ that are $O(n^{-1/2})$ distant from θ_0. Thus,

$$\frac{\partial g(\theta;\eta_0)}{\partial\theta}\Big|_{\theta=\theta_0} = \frac{1}{\sigma_0} + O(n^{-1}), \quad \frac{\partial^2 g(\theta;\eta_0)}{\partial\theta^2}\Big|_{\theta=\theta_0} = -\frac{d_0}{\sigma_0^2} + O(n^{-1/2}). \qquad (36)$$

By comparison of (36) with (32), it is evident that the second-order accuracy of the variance-stabilized bootstrap-t procedure based on transformation (12) can be demonstrated by an argument identical to the preceding one given for transformation (9).

A.4. Likelihood-based methods

To error of order $O(n^{-1})$, the first three cumulants of $W = (\hat{\theta}-\theta)/(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}$ are

$$E(W;\eta) = n^{-1/2}(a-d), \quad \mathrm{Var}(W;\eta) = 1, \quad \mathrm{skew}(W;\eta) = n^{-1/2}(c-6d), \quad (37)$$

where a, c and d are obtained by substituting formulae (24) into (19) and (20). Since the cumulants of W and $V = n^{1/2}(\hat{\theta}-\theta)/\hat{\sigma}$ agree to the order of error considered, it follows that $J(w;\eta)$ and $J^{-1}(\alpha;\eta)$ differ from $K(w;\eta)$ and $K^{-1}(\alpha;\eta)$ by terms of order $O(n^{-1})$. In particular, (21) implies that $J^{-1}(\alpha;\eta)$ has the Cornish-Fisher expansion

$$J^{-1}(\alpha;\eta) = z_\alpha + n^{-1/2}\{(a-\frac{1}{6}c) + (\frac{1}{6}c-d)z_\alpha^2\} + O(n^{-1}),$$

and $J^{-1}(\alpha;\hat{\eta}) = J^{-1}(\alpha;\eta) + O_p(n^{-1})$. Thus, the limit $\hat{\theta}_{\mathrm{BL}}(1-\alpha)$ differs from $\hat{\theta}_{\mathrm{EL}}(1-\alpha)$ by order $O_p(n^{-3/2})$ and has coverage error of order $O(n^{-1})$.

To investigate the limit $\hat{\theta}_{\rm SR}(1-\alpha)$ defined by Equation (13), let $\beta_1 = E\{\partial B(\theta)/\partial\theta\}$. Typically, β_1 is of order $O(1)$ and $\partial B(\theta)/\partial\theta$ differs from β_1 by order $O_p(n^{-1/2})$. Standard expansions show that the signed root $R = R(\theta)$ satisfies

$$R = W - n^{-1/2}(\frac{1}{6}c - d)W^2 + n^{-1/2}\sigma\beta_1 + O_p(n^{-1});$$

so (37) gives

$$E(R;\eta) = n^{-1/2}(a - \frac{1}{6}c + \sigma\beta_1), \quad {\rm Var}(R;\eta) = 1, \quad {\rm skew}(R;\eta) = 0,$$

to error of order $O(n^{-1})$. Thus, $m(\eta) = -\bar{Z} + n^{-1/2}\sigma\beta_1 + O(n^{-1})$ and $R(\theta) - m\{\tilde{\eta}(\theta)\}$ is an approximate pivotal quantity having the standard normal distribution to error of order $O(n^{-1})$. It follows that the limit $\hat{\theta}_{\rm SR}(1-\alpha)$ has coverage error of order $O(n^{-1})$.

Now, as a function of $\theta, R(\theta) - m\{\tilde{\eta}(\theta)\}$ has the expansion

$$\begin{aligned} R(\theta) - m\{\tilde{\eta}(\theta)\} &= (\hat{\theta} - \theta)/(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2} - n^{-1/2}(\frac{1}{6}c - d)(\hat{\theta} - \theta)^2/(I^{ij}\hat{\theta}_i\hat{\theta}_j) \\ &\quad - n^{-1/2}(a - \frac{1}{6}c) + O_p(n^{-1}), \end{aligned}$$

for values of θ such that $\hat{\theta} - \theta$ is $O_p(n^{-1/2})$. Consequently, $\hat{\theta}_{\rm SR}(1-\alpha)$ satisfies

$$\begin{aligned} \hat{\theta}_{\rm SR}(1-\alpha) &= \hat{\theta} - (I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}[z_\alpha + n^{-1/2}\{(a - \frac{1}{6}c) + (\frac{1}{6}c - d)z_\alpha^2\}] + O_p(n^{-3/2}) \\ &= \hat{\theta}_E(1-\alpha) + O_p(n^{-3/2}). \end{aligned}$$

Related expansions have been derived by DiCiccio and Stern (1992).

Finally, to show that the limit defined by (14) differs from $\hat{\theta}_{\rm EL}(1-\alpha)$ by order $O_p(n^{-3/2})$, suppose $\hat{\theta}(1-\alpha)$ is an approximate limit that is second-order accurate with respect to $\hat{\theta}_{\rm ES}(1-\alpha)$. Since $\hat{\theta}(1-\alpha) = \hat{\theta} - n^{-1/2}\hat{\sigma}K^{-1}(\alpha;\eta) + O_p(n^{-3/2})$,

$$\begin{aligned} \hat{\theta} + \{\hat{\theta}(1-\alpha) - \hat{\theta}\}\frac{(I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}}{n^{-1/2}\hat{\sigma}} &= \hat{\theta} - (I^{ij}\hat{\theta}_i\hat{\theta}_j)^{1/2}K^{-1}(\alpha;\eta) + O_p(n^{-3/2}) \\ &= \hat{\theta}_{\rm EL}(1-\alpha) + O_p(n^{-3/2}). \end{aligned}$$

References

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. Biometrika **70**, 343-365.

Barndorff-Nielsen, O. E. (1994). Adjusted versions of profile likelihood and directed likelihood and extended likelihood. J. Roy. Statist. Soc. Ser.B **56**, 125-140.

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. Biometrika **73**, 307-322.

Barndorff-Nielsen, O. E. and Chamberlin, S. R. (1992). Stable and invariant adjusted likelihood roots. Technical Report 259. Department of Theoretical Statistics, Aarhus University.

Bickel, P. J. (1987). Comment on "Better bootstrap confidence intervals" by B. Efron. J. Amer. Statist. Assoc. **82**, 191.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). J. Roy. Statist. Soc. Ser.B **49**, 1-39.

David, F. N. (1954). Tables of the Correlation Coefficient. Cambridge University Press, New York.

DiCiccio, T. J. and Efron, B. (1992). More accurate confidence intervals in exponential families. Biometrika **79**, 231-245.

DiCiccio, T. J. and Martin, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics. J. Roy. Statist. Soc. Ser.B **55**, 305-316.

DiCiccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals. J. Roy. Statist. Soc. Ser.B **50**, 338-354.

DiCiccio, T. J. and Romano, J. P. (1989). The automatic percentile method: Accurate confidence limits in parametric models. Canad. J. Statist. **17**, 155-169.

DiCiccio, T. J. and Romano, J. P. (1990). Nonparametric confidence limits by resampling methods and least favorable families. Internat. Statist. Rev. **58**, 59-76.

DiCiccio, T. J. and Stern, S. E. (1992). On using adjusted profile likelihoods to construct accurate one-sided confidence limits for scalar parameters. Technical Report 411. Department of Statistics, Stanford University.

Efron, B. (1981). Nonparametric standard errors and confidence intervals (with discussion). Canad. J. Statist. **9**, 139-172.

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. Biometrika **72**, 45-58.

Efron, B. (1987). Better bootstrap confidence intervals (with discussion). J. Amer. Statist. Assoc. **82**, 171-200.

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

Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. Biometrika **76**, 663-674.

McCullagh, P. (1984). Local sufficiency. Biometrika **71**, 233-244.

McCullagh, P. (1987). Tensor Methods in Statistics. Chapman and Hall, London.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. J. Roy. Statist. Soc. Ser.B **52**, 325-344.

Peers, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. J. Roy. Statist. Soc. Ser.B **27**, 9-16.

Stein, C. (1956). Efficient nonparametric testing and estimation. In Proc. 3rd Berkeley Symp. Math. Statist. Probab. **1** (Edited by J. Neymann), 187-195. University of California Press.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In Sequential Methods in Statistics, Banach center publications, **16**. PWN-Polish scientific publishers, Warsaw.

Tibshirani, R. (1988). Variance stabilization and the bootstrap. Biometrika **75**, 433-444.

Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. J. Roy. Statist. Soc. Ser.B **25**, 318-329.

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

(Received February 1993; accepted September 1994)