

AD-A248662

ACCURATE PROCEDURES FOR APPROXIMATE
BAYESIAN AND CONDITIONAL INFERENCE
WITHOUT THE NEED FOR ORTHOGONAL PARAMETERS

BY

THOMAS J. DICICCIO
JOSEPH B. KELLER
MICHAEL A. MARTIN

TECHNICAL REPORT NO. 450
FEBRUARY 24, 1992

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.



DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

92 4 16 04

**Accurate procedures for approximate Bayesian and conditional inference
without the need for orthogonal parameters**

By Thomas J. DiCiccio

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

Joseph B. Keller

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

and Michael A. Martin

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

Summary

This paper concerns methods to construct approximate confidence limits for a scalar parameter ψ in the presence of nuisance parameters. The methods are based on Bayesian procedures discussed by Peers (1965) and Stein (1985), in which the prior density is chosen so that the posterior quantiles of ψ are approximate confidence limits with coverage error of order $O(n^{-1})$ under repeated sampling. Multidimensional integration of the posterior density is avoided by using approximations of marginal densities and distribution functions; thus, adjustments are obtained that improve the standard normal approximation to the distributions of signed roots of the profile and conditional likelihood ratio statistics for ψ . The necessary prior densities are easy to specify when the nuisance parameters are orthogonal to the parameter of interest, and this simplicity is exploited in developing the methods. However, the need for explicit specification of an orthogonal parameterization is alleviated by approximating the Jacobian of a transformation to orthogonality. The methods are illustrated and compared with other procedures in some examples involving exponential families.

Some key words: Asymptotic normality; Conditional profile likelihood; Confidence limit; Exponential family; Gamma distribution; Marginal density approximation; Noninformative prior; Nuisance parameter; Orthogonal parameterization; Profile likelihood; Signed root likelihood ratio statistic; Tail probability approximation.

1. Introduction

Consider observed random variables X_1, \dots, X_n whose joint distribution depends on a d -dimensional parameter $\phi = (\phi^1, \dots, \phi^d)$, and suppose that inference about a scalar parameter $\psi = \psi(\phi)$ is of interest. Assume that the log likelihood function $l(\phi)$ attains its global maximum at $\hat{\phi} = (\hat{\phi}^1, \dots, \hat{\phi}^d)$ and that the constrained maximum is attained at $\tilde{\phi}(\psi) = (\tilde{\phi}^1, \dots, \tilde{\phi}^d)$ for fixed ψ . Then the maximum likelihood estimator of ψ is $\hat{\psi} = \psi(\hat{\phi})$, and $\tilde{\phi}(\hat{\psi}) = \hat{\phi}$. Expressed in terms of the profile likelihood function $l_p(\psi) = l\{\tilde{\phi}(\psi)\}$, the likelihood ratio statistic for testing $\psi = \psi_0$ is $W(\psi_0) = 2\{l_p(\hat{\psi}) - l_p(\psi_0)\}$, and the signed root of the likelihood ratio statistic is $R(\psi_0) = \text{sgn}(\hat{\psi} - \psi_0)\{W(\psi_0)\}^{1/2}$. Under the null hypothesis, the standard normal approximation to the conditional distribution of $R(\psi_0)$ typically has error of order $O(n^{-1/2})$, where the conditioning is on an exact or approximate ancillary statistic (McCullagh (1984), Barndorff-Nielsen (1986)). Consequently, the value of ψ that satisfies $\Phi\{R(\psi)\} = \alpha$ is an approximate upper $1 - \alpha$ confidence limit having coverage error of order $O(n^{-1/2})$, both conditionally and unconditionally. The primary goal of this paper is to develop related methods for constructing approximate confidence limits that attain higher coverage accuracy.

Bayesian procedures are available to construct improved confidence limits. Various authors have considered how to choose a prior density $\pi(\phi)$ so that, for each α , the posterior $1 - \alpha$ quantile of ψ is an upper confidence limit for the parameter ψ with coverage $1 - \alpha + O(n^{-1})$ in the repeated sampling sense. When there are no nuisance parameters, Welch and Peers (1963) showed that the prior density should be chosen proportional to the square root of the expected information for ψ . When nuisance parameters are present, there is considerable arbitrariness in the choice of prior density, and Peers (1965) and Stein (1985) developed differential equations for $\pi(\phi)$ whose solutions yield limits having coverage error of order $O(n^{-1})$. Unfortunately, two difficulties often arise in implementing the Bayesian methods. Exact calculation of the posterior quantiles of ψ usually requires numerical integration, which can be cumbersome. Moreover, for a convenient parameterization ϕ ,



Codes
and/or
al

A-11

solutions of the Peers and Stein differential equations can be difficult to find.

For Bayesian inference using a prior density $\pi(\phi)$ for ϕ , approximations to the posterior density and distribution function of ψ have been developed that avoid numerical integration. To describe these approximations, some additional notation is necessary. Let $l_i(\phi) = \partial l(\phi)/\partial\phi^i$, $l_{ij}(\phi) = \partial^2 l(\phi)/\partial\phi^i\partial\phi^j$, $\psi_i(\phi) = \partial\psi(\phi)/\partial\phi^i$, $\psi_{ij}(\phi) = \partial^2\psi(\phi)/\partial\phi^i\partial\phi^j$, ($i, j = 1, \dots, d$), and assume the gradient of $\psi(\phi)$ is nonzero. A Lagrange-multiplier argument shows there exists a constant $\tau(\psi)$ such that $\tau(\psi) = l_i(\tilde{\phi})/\psi_i(\tilde{\phi})$ for all i satisfying $\psi_i(\tilde{\phi}) \neq 0$. One value of the index i having this property always exists by assumption, and hence $\tau(\psi) = \{\sum l_i(\tilde{\phi})\}/\{\sum \psi_i(\tilde{\phi})\}$. Set

$$I_{ij}(\psi) = -l_{ij}(\tilde{\phi}) + \tau(\psi)\psi_{ij}(\tilde{\phi}) \quad (i, j = 1, \dots, d).$$

Then $I_{ij}(\hat{\psi}) = -l_{ij}(\hat{\phi})$, since $\tau(\hat{\psi}) = 0$. Define

$$T(\psi) = \frac{1}{\tau(\psi)} \frac{\pi(\tilde{\phi})}{\pi(\hat{\phi})} \left[\frac{\det\{I_{ij}(\hat{\psi})\}}{Q(\psi) \det\{I_{ij}(\psi)\}} \right]^{1/2}, \quad (1)$$

where $Q(\psi) = I^{ij}(\psi)\psi_i(\tilde{\phi})\psi_j(\tilde{\phi})$, $\{I^{ij}(\psi)\}$ is the $d \times d$ matrix inverse of $\{I_{ij}(\psi)\}$, and the standard summation convention is used. The Laplace approximation to the marginal posterior density of ψ given by Tierney, Kass and Kadane (1989) is

$$f_{\psi|X}(\psi) \simeq c \tau(\psi) T(\psi) \exp\{l_p(\psi) - l_p(\hat{\psi})\}, \quad (2)$$

where $X = (X_1, \dots, X_n)$ and c is a normalizing constant. Approximations to the posterior distribution function of ψ that follow from DiCiccio and Martin (1991, 1992) are

$$1 - F_{\psi|X}(\psi) \simeq \Phi(R) + \varphi(R)(R^{-1} - T), \quad 1 - F_{\psi|X}(\psi) \simeq \Phi\{R - R^{-1} \log(RT)\}, \quad (3)$$

where $R = R(\psi)$, $T = T(\psi)$, and Φ and φ are the standard normal distribution function and density, respectively. For arguments ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$, the relative error of approximation (2) is $O(n^{-3/2})$, and the errors of approximations (3) are also $O(n^{-3/2})$.

It follows that if a prior density of the Peers and Stein type is assumed, then the $1 - \alpha$ quantile of approximation (2) and the solutions of the equations

$$\Phi(R) + \varphi(R)(R^{-1} - T) = \alpha, \quad \Phi\{R - R^{-1} \log(RT)\} = \alpha, \quad (4)$$

are upper confidence limits for the parameter ψ having coverage $1 - \alpha + O(n^{-1})$. In some cases, there exists a prior density for which the posterior quantiles of ψ are approximate confidence limits with coverage error of order $O(n^{-3/2})$ or smaller. When such a prior density is used, the confidence limits derived from approximations (2) and (3) have coverage $1 - \alpha + O(n^{-3/2})$.

Note that in the important special case $\psi = \phi^1$,

$$\tau(\psi) = l_1(\tilde{\phi}), \quad \frac{\det\{I_{ij}(\hat{\psi})\}}{Q(\psi)\det\{I_{ij}(\psi)\}} = \frac{\det\{-l_{ij}(\hat{\phi})\}}{\det\{-l_{i'j'}(\tilde{\phi})\}},$$

where $\{-l_{i'j'}(\phi)\}$ is the $(d - 1) \times (d - 1)$ submatrix of $\{-l_{ij}(\phi)\}$ corresponding to the nuisance parameters ϕ^2, \dots, ϕ^d . Generally, $T = R^{-1} + O_p(n^{-1/2})$ for values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$, so $R^{-1} - T$ and $R^{-1} \log(RT)$ are both $O_p(n^{-1/2})$ in (3) and (4).

Although the Peers and Stein differential equations can be difficult to solve in an arbitrary parameterization ϕ , the equations simplify considerably when orthogonal parameters are used. Tibshirani (1989) noted that if the parameter of interest is $\psi(\phi) = \phi^1$ and the nuisance parameters ϕ^2, \dots, ϕ^d are orthogonal to ϕ^1 in the sense discussed by Cox and Reid (1987), the Peers and Stein equations reduce to

$$\{i_{11}(\phi)\}^{-1/2} \frac{\partial}{\partial \phi^1} \log \pi(\phi) + \frac{\partial}{\partial \phi^1} \{i_{11}(\phi)\}^{-1/2} = 0,$$

which has solutions of the form

$$\pi(\phi) \propto \{i_{11}(\phi)\}^{1/2} g(\phi^2, \dots, \phi^d),$$

where $i_{11}(\phi) = E\{-l_{11}(\phi)\}$ and $g(\phi^2, \dots, \phi^d)$ is an arbitrary positive function of the nuisance parameters.

A key property of orthogonal parameterizations is that the differences $\tilde{\phi}^{i'}(\phi^1) - \hat{\phi}^{i'} (i' = 2, \dots, d)$ between the constrained and overall maximum likelihood estimators of the nuisance parameters are $O_p(n^{-1})$ for values of ϕ^1 such that $\phi^1 - \hat{\phi}^1$ is $O_p(n^{-1/2})$. Consequently, for these values of $\psi = \phi^1$,

$$\frac{\pi(\tilde{\phi})}{\pi(\hat{\phi})} = \left\{ \frac{i_{11}(\tilde{\phi})}{i_{11}(\hat{\phi})} \right\}^{1/2} + O_p(n^{-1}), \quad (5)$$

and

$$T(\psi) = \frac{1}{l_1(\tilde{\phi})} \frac{\pi(\tilde{\phi})}{\pi(\hat{\phi})} \left[\frac{\det\{-l_{ij}(\hat{\phi})\}}{\det\{-l_{i'j'}(\hat{\phi})\}} \right]^{1/2} = \frac{1}{l_1(\tilde{\phi})} \left[\frac{i_{11}(\tilde{\phi}) \det\{-l_{ij}(\hat{\phi})\}}{i_{11}(\hat{\phi}) \det\{-l_{i'j'}(\hat{\phi})\}} \right]^{1/2} + O_p(n^{-1}). \quad (6)$$

The approximation in (6) is exact if the function $g(\phi^2, \dots, \phi^d)$ is constant. When this approximation is used in place of T , confidence limits obtained through (2) and (3) retain coverage error of order $O(n^{-1})$. If there are no nuisance parameters are present and $\psi = \phi$, then

$$T(\phi) = \frac{1}{l^{(1)}(\phi)} \left\{ \frac{i(\phi)}{i(\hat{\phi})} \right\}^{1/2} \{-l^{(2)}(\hat{\phi})\}^{1/2},$$

where $l^{(k)}(\phi) = d^k l(\phi)/d\phi^k$ ($k = 1, 2$) and $i(\phi) = E\{-l^{(2)}(\phi)\}$.

DiCiccio and Martin (1992) considered use of approximation (6) in equations (4) to construct confidence limits. By comparison of this method with related procedures of Barndorff-Nielsen (1986,1991) that improve the standard normal approximation to the conditional distribution of the signed root of the likelihood ratio statistic, they showed the approximate confidence limits derived using the Peers and Stein priors have coverage error of order $O(n^{-1})$ conditionally as well as unconditionally. Indeed, the limits obtained from the Bayesian approach differ from Barndorff-Nielsen's limits by terms of order $O_p(n^{-3/2})$. Although confidence limits produced by Barndorff-Nielsen's approximations have conditional coverage error of order $O(n^{-3/2})$, his procedures generally require specification of statistics that are exactly or approximately ancillary. Ancillary statistics are not necessary for approximation (6). However, an obstacle to the straightforward use of (6) is that it requires orthogonal parameters, and orthogonal parameterizations are often inconvenient

or difficult to find in practice, though they always exist. An approximation to T that does not directly involve such special parameterizations is developed in Section 2 of the present paper. The derivation of this approximation exploits the simple form of the solutions to the Peers and Stein equations in the orthogonal case; however, the need for explicit knowledge of an orthogonal parameterization is avoided by approximating the Jacobian of a transformation to orthogonality.

The use of equations (4) in the absence of nuisance parameters has also been considered by Barndorff-Nielsen and Chamberlain (1992).

Section 3 concerns the construction of improved confidence limits by methods similar to (3) that involve the signed root of the Cox and Reid (1987) conditional likelihood ratio statistic. For certain situations, particularly when the number of nuisance parameters is large, the standard normal approximation to the distribution of the signed root of the profile likelihood ratio statistic can be extremely poor, and approximate confidence limits obtained by solving equations (4) can have true coverage far from the nominal levels. In these cases, the distribution of the signed root of the conditional likelihood ratio statistic tends to be closer to the standard normal, and solving the equations analogous to (4) that are derived in Section 3 tends to produce approximate confidence limits with more accurate coverage. Although the methods developed in Section 3 offer better coverage accuracy, they are also computationally more difficult to implement in general. Cox and Reid (1987) defined their conditional likelihood ratio statistic in terms of a conditional profile likelihood function that requires orthogonal parameters. Section 3 contains an approximation to this function having error of order $O_p(n^{-1})$ that can be computed in any parameterization.

Some examples involving exponential families are considered in Section 4.

2. Approximation of T

Consider a reparameterization $\lambda(\phi) = (\lambda^1, \dots, \lambda^d)$ such that $\lambda^1 = \psi$ is the scalar

parameter of interest and the nuisance parameters $\lambda^2, \dots, \lambda^d$ are orthogonal to λ^1 . Suppose that a prior density $\bar{\pi}(\lambda)$ is assumed for the orthogonal parameterization λ . The corresponding prior density for the original parameterization ϕ is

$$\pi(\phi) = \bar{\pi}\{\lambda(\phi)\} |\det\{\lambda_a^a(\phi)\}| = \bar{\pi}\{\lambda(\phi)\} |\det[\phi_a^i\{\lambda(\phi)\}]|^{-1},$$

where $\lambda_i^a(\phi) = \partial\lambda^a(\phi)/\partial\phi^i$ and $\phi_a^i(\lambda) = \partial\phi^i(\lambda)/\partial\lambda^a$ ($a, i = 1, \dots, d$), so $[\phi_a^i\{\lambda(\phi)\}]$ is the $d \times d$ matrix inverse of $\{\lambda_i^a(\phi)\}$. With this prior density for ϕ , expression (1) for T becomes

$$T(\psi) = \frac{1}{\tau(\psi)} \frac{\bar{\pi}(\tilde{\lambda})}{\bar{\pi}(\hat{\lambda})} \left[\frac{\det\{I_{ij}(\hat{\psi})\}}{Q(\psi) \det\{I_{ij}(\psi)\}} \right]^{1/2} \frac{\det\{\phi_a^i(\hat{\lambda})\}}{\det\{\phi_a^i(\tilde{\lambda})\}}, \quad (7)$$

since $\hat{\lambda} = \lambda(\hat{\phi})$ and $\tilde{\lambda} = \lambda(\tilde{\phi})$. Thus, to approximate T requires knowledge of the ratios $\bar{\pi}(\tilde{\lambda})/\bar{\pi}(\hat{\lambda})$ and $\det\{\phi_a^i(\tilde{\lambda})\}/\det\{\phi_a^i(\hat{\lambda})\}$. Approximations to these ratios are given in formulae (10) and (14), respectively. The approximation to $\bar{\pi}(\tilde{\lambda})/\bar{\pi}(\hat{\lambda})$ applies in the case that $\bar{\pi}(\lambda)$ is a solution to the Peers and Stein differential equations.

Denote the log likelihood function for λ by $\bar{l}(\lambda)$, let $\bar{l}_{ab}(\lambda) = \partial^2 \bar{l}(\lambda)/\partial\lambda^a \partial\lambda^b$, and define $i_{ij}(\phi) = E\{-l_{ij}(\phi)\}$, $\bar{i}_{ab}(\lambda) = E\{-\bar{l}_{ab}(\lambda)\}$, ($a, b, i, j = 1, \dots, d$). Differentiating the identity $\bar{l}(\lambda) = l\{\phi(\lambda)\}$ twice and taking expectations yields

$$\bar{i}_{ab}(\lambda) = i_{ij}\{\phi(\lambda)\} \phi_a^i(\lambda) \phi_b^j(\lambda) \quad (a, b = 1, \dots, d),$$

and hence,

$$\bar{i}^{ab}(\lambda) = i^{ij}\{\phi(\lambda)\} \lambda_i^a\{\phi(\lambda)\} \lambda_j^b\{\phi(\lambda)\} \quad (a, b = 1, \dots, d), \quad (8)$$

where $\{\bar{i}^{ab}(\lambda)\}$ and $\{i^{ij}(\phi)\}$ are the $d \times d$ matrix inverses of $\{\bar{i}_{ab}(\lambda)\}$ and $\{i_{ij}(\phi)\}$, respectively. In particular, for $a = b = 1$, (8) becomes

$$\bar{i}^{11}(\lambda) = i^{ij}\{\phi(\lambda)\} \psi_i\{\phi(\lambda)\} \psi_j\{\phi(\lambda)\}. \quad (9)$$

By definition, orthogonality of $\lambda^2, \dots, \lambda^d$ to $\lambda^1 = \psi$ means that $\bar{i}_{a'1}(\lambda) = \bar{i}_{1a'}(\lambda) = 0$ ($a' = 2, \dots, d$). Two important consequences of orthogonality are $\bar{i}^{a'1}(\lambda) = \bar{i}^{1a'}(\lambda) = 0$ ($a' = 2, \dots, d$) and $\bar{i}^{11}(\lambda) = \{\bar{i}_{11}(\lambda)\}^{-1}$.

Suppose that $\bar{\pi}(\lambda)$ is a prior of the Peers and Stein type. Since $\hat{\phi} = \phi(\hat{\lambda})$ and $\tilde{\phi} = \phi(\tilde{\lambda})$, it follows from (5) and (9) that the ratio $\bar{\pi}(\tilde{\lambda})/\bar{\pi}(\hat{\lambda})$ can be approximated in the ϕ parameterization by

$$\begin{aligned}\frac{\bar{\pi}(\tilde{\lambda})}{\bar{\pi}(\hat{\lambda})} &= \left\{ \frac{i_{11}(\tilde{\lambda})}{i_{11}(\hat{\lambda})} \right\}^{1/2} + O_p(n^{-1}) \\ &= \left\{ \frac{i^{ij}(\hat{\phi})\psi_i(\hat{\phi})\psi_j(\hat{\phi})}{i^{ij}(\tilde{\phi})\psi_i(\tilde{\phi})\psi_j(\tilde{\phi})} \right\}^{1/2} + O_p(n^{-1})\end{aligned}\quad (10)$$

for values of the parameter ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$.

It follows from (8) that

$$i^{ab}(\lambda)\phi_b^i(\lambda) = i^{ij}\{\phi(\lambda)\}\lambda_j^a\{\phi(\lambda)\} \quad (a, i = 1, \dots, d), \quad (11)$$

and setting $a = 1$ in (11) gives

$$i^{11}(\lambda)\phi_1^i(\lambda) = i^{ij}\{\phi(\lambda)\}\psi_j\{\phi(\lambda)\} \quad (i = 1, \dots, d). \quad (12)$$

Combining (9) and (12) shows that for an orthogonal reparameterization $\lambda(\phi)$ with $\lambda^1 = \psi$ the parameter of interest, $\phi_1^i(\lambda) = \partial\phi^i(\lambda)/\partial\lambda^1 = \Gamma^i\{\phi(\lambda)\}$ ($i = 1, \dots, d$), where

$$\Gamma^i(\phi) = \frac{i^{ij}(\phi)\psi_j(\phi)}{i^{kl}(\phi)\psi_k(\phi)\psi_l(\phi)}.$$

The converse of this result also holds: if a reparameterization $\lambda(\phi)$ satisfies $\phi_1^i(\lambda) = \Gamma^i\{\phi(\lambda)\}$ ($i = 1, \dots, d$), then $\lambda^2, \dots, \lambda^d$ are orthogonal to $\lambda^1 = \psi$. When $\psi = \phi^1$ is the scalar parameter of interest, $\Gamma^i(\phi) = i^{11}(\phi)/i^{11}(\phi)$ ($i = 1, \dots, d$). Then the conditions $\phi_1^i(\lambda) = \Gamma^i\{\phi(\lambda)\}$ ($i = 1, \dots, d$) for orthogonality are equivalent to $i_{i'j'}\{\phi(\lambda)\}\phi_1^j(\lambda) = 0$ ($i' = 2, \dots, d$), or

$$\sum_{j'=2}^d i_{i'j'}\{\phi(\lambda)\}\phi_1^{j'}(\lambda) = -i_{i'1}\{\phi(\lambda)\} \quad (i' = 2, \dots, d),$$

which are the orthogonality equations derived by Cox and Reid (1987).

Now differentiation of the identities $\phi_1^i(\lambda) = \Gamma^i\{\phi(\lambda)\}$ ($i = 1, \dots, d$) with respect to λ^a produces the system of equations

$$\phi_{1a}^i(\lambda) = \Gamma_j^i\{\phi(\lambda)\}\phi_a^j(\lambda) \quad (a, i = 1, \dots, d),$$

where $\Gamma_j^i(\phi) = \partial\Gamma^i(\phi)/\partial\phi^j$. Theorem 7.3 of Coddington and Levinson (1955, p.28) shows that the solutions $\phi_a^i(\lambda)$ ($a, i = 1, \dots, d$) to this system have the property

$$\det\{\phi_a^i(\lambda^1, \lambda^2, \dots, \lambda^d)\} = \det\{\phi_a^i(\sigma, \lambda^2, \dots, \lambda^d)\} \exp\left[\int_\sigma^{\lambda^1} \text{tr } \Gamma\{\phi(s, \lambda^2, \dots, \lambda^d)\} ds\right], \quad (13)$$

where σ is an arbitrary constant and $\Gamma(\phi) = \{\Gamma_j^i(\phi)\}$ is a $d \times d$ matrix. By choosing $\lambda^2 = \tilde{\lambda}^2(\lambda^1), \dots, \lambda^d = \tilde{\lambda}^d(\lambda^1)$ and $\sigma = \hat{\lambda}^1$, the left-hand side of (13) is

$$\det[\phi_a^i\{\lambda^1, \tilde{\lambda}^2(\lambda^1), \dots, \tilde{\lambda}^d(\lambda^1)\}] = \det\{\phi_a^i(\tilde{\lambda})\},$$

while the factors on the right-hand side are

$$\det[\phi_a^i\{\hat{\lambda}^1, \tilde{\lambda}^2(\lambda^1), \dots, \tilde{\lambda}^d(\lambda^1)\}] = \det\{\phi_a^i(\hat{\lambda})\} + O_p(n^{-1}),$$

and

$$\begin{aligned} \exp\left(\int_{\hat{\lambda}^1}^{\lambda^1} \text{tr } \Gamma[\phi\{s, \tilde{\lambda}^2(\lambda^1), \dots, \tilde{\lambda}^d(\lambda^1)\}] ds\right) &= \exp\left(\int_{\hat{\lambda}^1}^{\lambda^1} \text{tr } \Gamma[\phi\{\tilde{\lambda}(s)\}] ds\right) + O_p(n^{-1}) \\ &= \exp\left[\int_{\hat{\lambda}^1}^{\lambda^1} \text{tr } \Gamma\{\tilde{\phi}(s)\} ds\right] + O_p(n^{-1}), \end{aligned}$$

for values of λ^1 such that $\lambda^1 - \hat{\lambda}^1$ is $O_p(n^{-1/2})$. Therefore, the ratio $\det\{\phi_a^i(\tilde{\lambda})\}/\det\{\phi_a^i(\hat{\lambda})\}$ can be approximated in the ϕ parameterization by

$$\frac{\det\{\phi_a^i(\tilde{\lambda})\}}{\det\{\phi_a^i(\hat{\lambda})\}} = \exp\left[\int_{\hat{\psi}}^\psi \text{tr } \Gamma\{\tilde{\phi}(s)\} ds\right] + O_p(n^{-1}), \quad (14)$$

for values of the parameter ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$. Since the ratio is $1 + O_p(n^{-1/2})$, other approximations having error of order $O_p(n^{-1})$ are

$$\exp\{(\psi - \hat{\psi}) \text{tr } \Gamma(\tilde{\phi})\}, \quad \exp\{(\psi - \hat{\psi}) \text{tr } \Gamma(\hat{\phi})\}.$$

The advantage of these approximations over (14) is computational simplicity.

In summary, the approximation to $T(\psi)$ that emerges from (7), (10) and (14) is

$$T(\psi) = \frac{1}{\tau(\psi)} \left\{ \frac{i^{ij}(\hat{\phi})\psi_i(\hat{\phi})\psi_j(\hat{\phi})}{i^{ij}(\tilde{\phi})\psi_i(\tilde{\phi})\psi_j(\tilde{\phi})} \right\}^{1/2} \left[\frac{\det\{I_{ij}(\hat{\psi})\}}{Q(\psi)\det\{I_{ij}(\psi)\}} \right]^{1/2} \times \exp \left[- \int_{\hat{\psi}}^{\psi} \text{tr } \Gamma\{\tilde{\phi}(s)\} ds \right] + O_p(n^{-1}), \quad (15)$$

and the same order of error holds if the integral is replaced by either $(\hat{\psi} - \psi) \text{tr } \Gamma(\tilde{\phi})$ or $(\hat{\psi} - \psi) \text{tr } \Gamma(\hat{\phi})$. If approximation (15) is used for T , then the confidence limits obtained through (2) and (3) have coverage error of order $O(n^{-1})$, both conditionally and unconditionally. When the parameterization ϕ is orthogonal, $\text{tr } \Gamma(\phi) = 0$ and (15) reduces to approximation (6).

An alternative formula for $\text{tr } \Gamma(\phi)$ is now developed that may be useful in computing (15). Differentiation of the identities $\Gamma^i(\phi) = i^{ij}(\phi)\kappa_j(\phi)$ ($i = 1, \dots, d$) with respect to ϕ^l yields

$$\Gamma_l^i(\phi) = i^{ij}(\phi)\{\kappa_{j,l}(\phi) - i_{jk,l}(\phi)\Gamma^k(\phi)\} \quad (i, l = 1, \dots, d),$$

where $\kappa_j(\phi) = \psi_j(\phi)/\{i^{kl}(\phi)\psi_k(\phi)\psi_l(\phi)\}$, $\kappa_{j,l}(\phi) = \partial\kappa_j(\phi)/\partial\phi^l$, $i_{jk,l}(\phi) = \partial i_{jk}(\phi)/\partial\phi^l$, and the result $i_l^{ij}(\phi) = -i^{ik}(\phi)i^{jm}(\phi)i_{km,l}(\phi)$ is used, ($i, j, k, l = 1, \dots, d$). Thus,

$$\text{tr } \Gamma(\phi) = i^{ij}(\phi)\{\kappa_{i,j}(\phi) - i_{ik,j}(\phi)\Gamma^k(\phi)\}. \quad (16)$$

When $\psi = \phi^1$ is the scalar parameter of interest, $\kappa_j(\phi) = \delta_j^1/i^{11}(\phi)$ ($j = 1, \dots, d$), where δ_j^1 is Kronecker's delta, and (16) yields

$$\text{tr } \Gamma(\phi) = i_{ij,k}(\phi)i^{i1}(\phi)\{i^{j1}(\phi)i^{k1}(\phi) - i^{11}(\phi)i^{jk}(\phi)\}/\{i^{11}(\phi)\}^2. \quad (17)$$

A desirable property of any procedure for inference about ψ is that it not depend on the particular choice of underlying parameter ϕ . Confidence limits obtained from the standard normal approximation for the signed root of the likelihood ratio statistic have this property, since $R(\psi)$ is invariant under reparameterization. Furthermore, for fixed prior density

$\pi(\phi)$, the quantities $T(\psi)$ and $\tau(\psi)T(\psi)$ defined at (1) are invariant, and the approximate posterior quantiles of ψ derived from (2) and (3) also do not depend on choice of underlying parameter. When (2) and (3) are used to construct approximate confidence limits, with $T(\psi)$ given by (1) and $\pi(\psi)$ taken to be a solution to the Peers and Stein differential equations, some arbitrariness can arise because the solutions to these equations are not unique in general. However, the confidence limits obtained from using different solutions agree at least to order $O_p(n^{-1})$, that is, they differ by at most $O_p(n^{-3/2})$. Approximation (6) to $T(\psi)$, requiring an orthogonal parameterization, is not invariant since it depends on the orthogonal parameterization used. This approximation is independent of the choice of orthogonal parameters to error of order $O_p(n^{-1})$, however, and when it is used in (2) and (3) to construct approximate confidence limits, the resulting limits are parameterization invariant to error of order $O_p(n^{-3/2})$. Similar comments apply to approximation (15). This approximation to $T(\psi)$ depends on the choice of parameterization ϕ at order $O_p(n^{-1})$, and it can be used in (2) and (3) to produce approximate confidence limits that are parameterization dependent only at order $O_p(n^{-3/2})$.

It is worthwhile to remark that the quantities $\Gamma^i(\phi)$ ($i = 1, \dots, d$) have another interpretation. A straightforward calculation involving Lagrange multipliers shows that

$$\frac{d}{d\psi} \tilde{\phi}^i(\psi) = \frac{I^{ij}(\psi)\psi_j(\tilde{\phi})}{I^{kl}(\psi)\psi_k(\tilde{\phi})\psi_l(\tilde{\phi})} \quad (i = 1, \dots, d).$$

In particular,

$$\left. \frac{d}{d\psi} \tilde{\phi}^i(\psi) \right|_{\psi=\hat{\psi}} = \frac{l^{ij}(\hat{\phi})\psi_j(\hat{\phi})}{l^{kl}(\hat{\phi})\psi_k(\hat{\phi})\psi_l(\hat{\phi})} \quad (i = 1, \dots, d),$$

and thus, for values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$,

$$\frac{d}{d\psi} \tilde{\phi}^i(\psi) = \Gamma^i(\tilde{\phi}) + O_p(n^{-1/2}) \quad (i = 1, \dots, d).$$

3. Procedures based on conditional profile likelihood

Suppose, as in Section 2, that a prior density $\bar{\pi}(\lambda)$ is considered for an orthogonal reparameterization $\lambda = \lambda(\phi)$. Then $T(\psi)$ is given by (7), and the Laplace approximation (2) to the posterior density of ψ can be written as

$$f_{\psi|X}(\psi) \simeq c \frac{\bar{\pi}(\tilde{\lambda})}{\bar{\pi}(\hat{\lambda})} \exp\{l_c(\psi) - l_c(\check{\psi})\}, \quad (18)$$

where

$$l_c(\psi) = l_p(\psi) - \frac{1}{2} \log \left[\frac{Q(\psi) \det\{I_{ij}(\psi)\}}{Q(\check{\psi}) \det\{I_{ij}(\check{\psi})\}} \right] - \log \left[\frac{\det\{\phi_a^i(\tilde{\lambda})\}}{\det\{\phi_a^i(\hat{\lambda})\}} \right],$$

$l_c(\psi)$ is maximized at $\check{\psi}$, and c is a normalizing constant. Applying a tail probability approximation of DiCiccio and Martin (1991) to (18) yields

$$1 - F_{\psi|X}(\psi) \simeq \Phi(R_c) + \varphi(R_c)(R_c^{-1} - T_c), \quad 1 - F_{\psi|X}(\psi) \simeq \Phi\{R_c - R_c^{-1} \log(R_c T_c)\}, \quad (19)$$

where $R_c = R_c(\psi) = \text{sgn}(\check{\psi} - \psi)[2\{l_c(\check{\psi}) - l_c(\psi)\}]^{1/2}$,

$$T_c = T_c(\psi) = \frac{\bar{\pi}\{\tilde{\lambda}(\psi)\}}{\bar{\pi}\{\tilde{\lambda}(\check{\psi})\}} \frac{\{-l_c^{(2)}(\check{\psi})\}^{1/2}}{l_c^{(1)}(\psi)},$$

and $l_c^{(k)}(\psi) = d^k l_c(\psi)/d\psi^k$, ($k = 1, 2$). Both approximations in (19) have errors of order $O_p(n^{-3/2})$ for arguments ψ such that $\psi - \check{\psi}$ is $O_p(n^{-1/2})$. It follows that if $\bar{\pi}(\lambda)$ is a solution of the Peers and Stein differential equations, then the solutions of the equations

$$\Phi(R_c) + \varphi(R_c)(R_c^{-1} - T_c) = \alpha, \quad \Phi\{R_c - R_c^{-1} \log(R_c T_c)\} = \alpha, \quad (20)$$

are approximate upper $1 - \alpha$ confidence limits having coverage error of order $O(n^{-1})$ conditionally as well as unconditionally. If the prior density $\bar{\pi}(\lambda)$ produces posterior quantiles for ψ that are approximate confidence limits having coverage error of order $O(n^{-3/2})$ or smaller, then the confidence limits obtained from equations (20) have coverage error of order $O(n^{-3/2})$.

The function $l_c(\psi)$ is identical to the conditional profile likelihood function of Cox and Reid (1987). In terms of an orthogonal parameterization λ with log likelihood function $\bar{l}(\lambda)$, Cox and Reid have recommended replacing the objective function $l_p(\psi)$ by

$$l_p(\psi) - \frac{1}{2} \log \left[\frac{\det\{-\bar{l}_{a'b'}(\tilde{\lambda})\}}{\det\{-\bar{l}_{ab}(\hat{\lambda})\}} \right], \quad (21)$$

where $\{-\bar{l}_{a'b'}(\lambda)\}$ is the $(d-1) \times (d-1)$ submatrix of $\{-\bar{l}_{ab}(\lambda)\}$ corresponding to the nuisance parameters $\lambda^2, \dots, \lambda^d$. To demonstrate the equality of $l_c(\psi)$ and (21), note that

$$\det\{-\bar{l}_{a'b'}(\tilde{\lambda})\} = -\bar{l}^{11}(\tilde{\lambda}) \det\{-\bar{l}_{ab}(\tilde{\lambda})\}, \quad (22)$$

where $\{-\bar{l}^{ab}(\lambda)\}$ is the $d \times d$ matrix inverse of $\{-\bar{l}_{ab}(\lambda)\}$. Differentiation of the identity $\bar{l}(\lambda) = l\{\phi(\lambda)\}$ gives

$$-\bar{l}_{ab}(\tilde{\lambda}) = I_{ij}(\psi) \phi_a^i(\tilde{\lambda}) \phi_b^j(\tilde{\lambda}), \quad -\bar{l}^{ab}(\tilde{\lambda}) = I^{ij}(\psi) \lambda_i^a(\tilde{\lambda}) \lambda_j^b(\tilde{\lambda}) \quad (a, b = 1, \dots, d);$$

hence,

$$\det\{-\bar{l}_{ab}(\tilde{\lambda})\} = \det\{I_{ij}(\psi)\} [\det\{\phi_a^i(\tilde{\lambda})\}]^2, \quad -\bar{l}^{11}(\tilde{\lambda}) = Q(\psi). \quad (23)$$

Combining (22) and (23) gives

$$\left[\frac{\det\{-\bar{l}_{a'b'}(\tilde{\lambda})\}}{\det\{-\bar{l}_{ab}(\tilde{\lambda})\}} \right]^{1/2} = \left[\frac{Q(\psi) \det\{I_{ij}(\psi)\}}{Q(\hat{\psi}) \det\{I_{ij}(\hat{\psi})\}} \right]^{1/2} \frac{\det\{\phi_a^i(\tilde{\lambda})\}}{\det\{\phi_a^i(\hat{\lambda})\}},$$

and it follows that $l_c(\psi)$ and (21) are the same.

Calculation of $l_c(\psi)$ requires knowledge of the orthogonal parameterization λ . However, by using (14), $l_c(\psi)$ can be approximated by a function that does not explicitly involve orthogonal parameters. For values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$,

$$l_c(\psi) = l_p(\psi) - \frac{1}{2} \log \left[\frac{Q(\psi) \det\{I_{ij}(\psi)\}}{Q(\hat{\psi}) \det\{I_{ij}(\hat{\psi})\}} \right] - \int_{\hat{\psi}}^{\psi} \text{tr } \Gamma\{\tilde{\phi}(s)\} ds + O_p(n^{-1}), \quad (24)$$

and the integral can be replaced by either $(\hat{\psi} - \psi) \text{tr } \Gamma(\tilde{\phi})$ or $(\hat{\psi} - \psi) \text{tr } \Gamma(\hat{\phi})$ without changing the order of the error term. Similarly, when $\bar{\pi}(\lambda)$ is a prior of the Peers and Stein type, it follows from (9) that $T_c(\psi)$ can be approximated by

$$T_c(\psi) = \left\{ \frac{i^{ij}(\check{\phi}) \psi_i(\check{\phi}) \psi_j(\check{\phi})}{i^{ij}(\tilde{\phi}) \psi_i(\tilde{\phi}) \psi_j(\tilde{\phi})} \right\}^{1/2} \frac{\{-l_c^{(2)}(\check{\psi})\}^{1/2}}{l_c^{(1)}(\psi)} + O_p(n^{-1}), \quad (25)$$

where $\check{\phi} = \tilde{\phi}(\check{\psi})$. Since $\check{\psi} = \hat{\psi} + O_p(n^{-1})$, $l_c^{(2)}(\check{\psi}) = -l_p^{(2)}(\hat{\psi}) + O_p(1)$, and $-l_p^{(2)}(\hat{\psi}) = \{Q(\hat{\psi})\}^{-1}$, the computationally simpler approximation

$$T_c(\psi) = \left\{ \frac{i^{ij}(\hat{\phi})\psi_i(\hat{\phi})\psi_j(\hat{\phi})}{i^{ij}(\tilde{\phi})\psi_i(\tilde{\phi})\psi_j(\tilde{\phi})} \right\}^{1/2} \frac{\{Q(\hat{\psi})\}^{-1/2}}{l_c^{(1)}(\psi)} + O_p(n^{-1}) \quad (26)$$

also holds. The solutions to equations (20) remain upper confidence limits with conditional coverage $1 - \alpha + O(n^{-1})$ when either approximation (25) or (26) is used for $T_c(\psi)$ and approximation (24) is used for $l_c(\psi)$ in calculating $\check{\psi}$, $R_c(\psi)$ and $T_c(\psi)$.

The function $l_c(\psi)$ depends on the choice of orthogonal parameterization; however, for arguments ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$, the values of $l_c(\psi)$ obtained under different orthogonal parameterizations vary by at most $O_p(n^{-1})$. The same property holds for the function $T_c(\psi)$, when the prior density is completely specified. Approximate confidence limits obtained by solving the equations in (20) depend on both the orthogonal parameterization and the solution to the Peers and Stein equations used. However, all the limits thus derived agree up to order $O_p(n^{-1})$; that is, they differ from one another by terms of order $O_p(n^{-3/2})$. Similar comments apply for approximations (24), (25) and (26). Approximation (24) to $l_c(\psi)$ depends on the underlying parameterization ϕ at order $O_p(n^{-1})$ for values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$. When (24) is used in (25) or (26) to approximate $T_c(\psi)$, the resulting approximation is also parameterization invariant to error of order $O_p(n^{-1})$. Using these approximations for $l_c(\psi)$ and $T_c(\psi)$ in (20) produces approximate confidence limits that depend on the parameterization ϕ only at order $O_p(n^{-3/2})$.

4. Examples

The methods of Section 2 and 3 are now illustrated in some simple situations where orthogonal parameterizations are readily available, so that comparisons with other available procedures are possible.

Consider a sample X_1, \dots, X_n from a d -dimensional exponential family. Assume the log likelihood function for the canonical parameter $\phi = (\phi^1, \dots, \phi^d)$ is $l(\phi) = n\{\sum \phi^i \bar{X}^i -$

$\beta(\phi)$ }, where $\bar{X} = (\bar{X}^1, \dots, \bar{X}^d) = n^{-1} \sum X_i$, and let $\beta_i(\phi) = \partial\beta(\phi)/\partial\phi^i$, $\beta_{ij}(\phi) = \partial^2\beta(\phi)/\partial\phi^i\partial\phi^j$ ($i, j = 1, \dots, d$). Then $E(\bar{X}^i) = \beta_i(\phi)$ ($i = 1, \dots, d$), and in the notation of Section 2, $-l_{ij}(\phi) = i_{ij}(\phi) = n\beta_{ij}(\phi)$ and $i^{ij}(\phi) = n^{-1}\beta^{ij}(\phi)$ ($i, j = 1, \dots, d$), where $\{\beta^{ij}(\phi)\}$ is the $d \times d$ matrix inverse of $\{\beta_{ij}(\phi)\}$.

Example 1. Suppose that the scalar parameter of interest is $\psi = \phi^1$. Cox and Reid (1987) have noted that $\lambda = (\lambda^1, \dots, \lambda^d) = \{\phi^1, \beta_2(\phi), \dots, \beta_d(\phi)\}$ is an orthogonal parameterization, and $T(\psi)$ can be calculated from (7) using this parameterization. By orthogonality and (9), the Peers and Stein prior densities are of the form $\bar{\pi}(\lambda) \propto [\beta^{11}\{\phi(\lambda)\}]^{-1/2} g(\lambda^2, \dots, \lambda^d)$. If the function $g(\lambda^2, \dots, \lambda^d)$ is chosen to be constant, then $\bar{\pi}(\tilde{\lambda})/\bar{\pi}(\hat{\lambda}) = \{\beta^{11}(\hat{\phi})/\beta^{11}(\tilde{\phi})\}^{1/2}$. Now define $B(\phi) = \{\beta_{ij}(\phi)\}$ and $B_1(\phi) = \{\beta_{i'j'}(\phi)\}$ ($i', j' = 2, \dots, d$), so that $B_1(\phi)$ is the $(d-1) \times (d-1)$ submatrix of $B(\phi)$ corresponding to the nuisance parameters ϕ^2, \dots, ϕ^d . Since $\det\{\lambda_i^a(\phi)\} = B_1(\phi)$, it follows that $\det\{\phi_a^i(\tilde{\lambda})\}/\det\{\phi_a^i(\hat{\lambda})\} = \det B_1(\hat{\phi})/\det B_1(\tilde{\phi})$. Direct calculation shows that $\tau(\psi) = l_1(\tilde{\phi}) = n\{\beta_1(\hat{\phi}) - \beta_1(\tilde{\phi})\}$, $\det\{I_{ij}(\psi)\} = n^d \det B(\tilde{\phi})$, and $Q(\psi) = n^{-1}\beta^{11}(\tilde{\phi})$. Hence, (7) gives

$$T(\psi) = \frac{1}{n^{1/2}\{\beta_1(\hat{\phi}) - \beta_1(\tilde{\phi})\}} \left\{ \frac{\det B(\tilde{\phi})}{\det B_1(\hat{\phi})} \right\}^{1/2}. \quad (27)$$

For other choices of $g(\lambda^2, \dots, \lambda^d)$ in the prior density $\bar{\pi}(\lambda)$, (10) ensures that (27) is an approximation to (7) having error of order $O_p(n^{-1})$ for values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$.

Approximation (15) to $T(\psi)$, which does not require knowledge of orthogonal parameters, coincides with (27), because approximation (14) is exact in this case. The identity

$$\exp \left[\int_{\hat{\psi}}^{\psi} \text{tr } \Gamma\{\tilde{\phi}(s)\} ds \right] = \frac{\det B_1(\hat{\phi})}{\det B_1(\tilde{\phi})}, \quad (28)$$

follows from expression (17). Since $i_{ijk}(\phi) = \beta_{ijk}(\phi)$ ($i, j, k = 1, \dots, d$), where $\beta_{ijk}(\phi) = \partial^3\beta(\phi)/\partial\phi^i\partial\phi^j\partial\phi^k$, and since $d\tilde{\phi}^i(\psi)/d\psi = \beta^{i1}(\tilde{\phi})/\beta^{11}(\tilde{\phi})$ ($i = 1, \dots, d$), (17) yields

$$\text{tr } \Gamma\{\tilde{\phi}(\psi)\} = -\{\beta^{ij}(\tilde{\phi}) - \beta^{i1}(\tilde{\phi})\beta^{j1}(\tilde{\phi})/\beta^{11}(\tilde{\phi})\} \beta_{ijk}(\tilde{\phi}) \beta^{k1}(\tilde{\phi})/\beta^{11}(\tilde{\phi})$$

$$\begin{aligned}
&= -\text{tr} \left[B_1(\tilde{\phi})^{-1} \frac{d}{d\psi} B_1\{\tilde{\phi}(\psi)\} \right] \\
&= -\frac{d}{d\psi} \log \det B_1\{\tilde{\phi}(\psi)\}.
\end{aligned} \tag{29}$$

For this choice of orthogonal parameters, the conditional profile likelihood function is

$$l_c(\psi) = l_p(\psi) + \log \left\{ \frac{\det B_1(\tilde{\phi})}{\det B_1(\hat{\phi})} \right\}^{1/2},$$

which was derived by Barndorff-Nielsen and Cox (1979). It follows from (28) that approximation (24) to $l_c(\psi)$ is exact. If the prior density is $\bar{\pi}(\lambda) \propto [\beta^{11}\{\phi(\lambda)\}]^{-1/2}$, then

$$T_c(\psi) = \left\{ \frac{\beta^{11}(\check{\phi})}{\beta^{11}(\tilde{\phi})} \right\}^{1/2} \frac{\{-l_c^{(2)}(\check{\psi})\}^{1/2}}{l_c^{(1)}(\psi)}; \tag{30}$$

for other prior densities of the Peers and Stein type, (30) is an approximation to $T_c(\psi)$ having error of order $O_p(n^{-1})$ for values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$. Approximation (25) to $T_c(\psi)$ agrees with (30), and approximation (26) is

$$T_c(\psi) = \frac{n^{1/2}}{\{\beta^{11}(\tilde{\phi})\}^{1/2} l_c^{(1)}(\psi)} + O_p(n^{-1}). \tag{31}$$

It is of interest in the present context to compare the methods from Sections 2 and 3 with similar procedures that have appeared in the literature. For this example, if $T(\psi)$ and $T_c(\psi)$ are replaced by

$$U(\psi) = \frac{1}{n^{1/2}(\hat{\psi} - \psi)} \left\{ \frac{\det B_1(\tilde{\phi})}{\det B(\hat{\phi})} \right\}^{1/2}, \quad U_c(\psi) = \frac{1}{(\check{\psi} - \psi)\{-l_c^{(2)}(\check{\psi})\}^{1/2}} \tag{32}$$

in equations (4) and (20), then the solutions to these equations are approximate upper confidence limits with coverage $1 - \alpha + O(n^{-3/2})$. This use of $U(\psi)$ and $U_c(\psi)$ has been discussed by Barndorff-Nielsen (1986), Skovgaard (1987), Davison (1988), DiCiccio and Martin (1991), Fraser (1991), Fraser, Reid, and Wong (1991), and Pierce and Peters (1992). For values of ψ such that $\psi - \hat{\psi}$ is $O_p(n^{-1/2})$, $T(\psi)$ and $T_c(\psi)$ differ from $U(\psi)$ and $U_c(\psi)$ by terms of order $O_p(n^{-1})$. Note that $U(\psi)$ is obtained from $T(\psi)$ by replacing $l_1(\tilde{\phi})\{i^{11}(\tilde{\phi})\}^{1/2} = n^{1/2}\{\beta_1(\hat{\phi}) - \beta_1(\tilde{\phi})\}\{\beta^{11}(\tilde{\phi})\}^{1/2}$ with the asymptotically equivalent

quantity $(\hat{\phi}^1 - \phi^1)\{l^{11}(\hat{\phi})\}^{-1/2} = n^{1/2}(\hat{\psi} - \psi)\{\beta^{11}(\hat{\phi})\}^{-1/2}$, and that $U_c(\psi)$ is obtained from $T_c(\psi)$ by replacing $l_c^{(1)}(\psi)\{i^{11}(\tilde{\phi})\}^{1/2} = n^{-1/2}l_c^{(1)}(\psi)\{\beta^{11}(\tilde{\phi})\}^{1/2}$ with $(\check{\psi} - \psi)\{-l_c^{(2)}(\check{\psi})\}^{1/2}$ in (31).

Example 2. Suppose the parameter of interest is $\psi = E(\bar{X}^1) = \beta_1(\phi)$. An orthogonal parameterization is $\lambda = (\lambda^1, \dots, \lambda^d) = \{\beta_1(\phi), \phi^2, \dots, \phi^d\}$, and for a specified prior density $\bar{\pi}(\lambda)$, $T(\psi)$ can be calculated from (7). The Peers and Stein prior densities have the form $\bar{\pi}(\lambda) \propto [\beta_{11}\{\phi(\lambda)\}]^{-1/2}g(\lambda^2, \dots, \lambda^d)$, and if the function $g(\lambda^2, \dots, \lambda^d)$ is taken to be constant, then $\bar{\pi}(\tilde{\lambda})/\bar{\pi}(\hat{\lambda}) = \{\beta_{11}(\hat{\phi})/\beta_{11}(\tilde{\phi})\}^{1/2}$. It is easily seen that $\det\{\lambda_i^a(\phi)\} = \beta_{11}(\phi)$, whence $\det\{\phi_a^i(\tilde{\lambda})\}/\det\{\phi_a^i(\hat{\lambda})\} = \beta_{11}(\hat{\phi})/\beta_{11}(\tilde{\phi})$, and that $\tau(\psi) = n\{\hat{\psi} - \psi\}/\beta_{11}(\tilde{\phi})$, where $\hat{\psi} = \bar{X}^1$. Therefore, (7) yields

$$T(\psi) = \frac{\{\beta_{11}(\tilde{\phi})\}^{3/2}}{n(\hat{\psi} - \psi)\{\beta_{11}(\hat{\phi})\}^{1/2}} \left[\frac{\det\{I_{ij}(\hat{\psi})\}}{Q(\psi)\det\{I_{ij}(\psi)\}} \right]^{1/2}, \quad (33)$$

with $I_{ij}(\psi) = n\{\beta_{ij}(\tilde{\phi}) + (\hat{\psi} - \psi)\beta_{1ij}(\tilde{\phi})/\beta_{11}(\tilde{\phi})\}$ and $Q(\psi) = I^{ij}(\psi)\beta_{1i}(\tilde{\phi})\beta_{1j}(\tilde{\phi})$.

Approximation (15) to $T(\psi)$ does not coincide with (33) because (14) fails to be exact in this case. Approximation (14) can be verified directly, since $\Gamma^i(\phi) = \delta_1^i/\beta_{11}(\phi)$ ($i = 1, \dots, d$) and $d\tilde{\phi}^i(\psi)/d\psi = \Gamma^i(\tilde{\phi}) + O_p(n^{-1/2})$ for values of ψ such that $\hat{\psi} - \psi$ is $O_p(n^{-1/2})$. Thus,

$$\begin{aligned} \text{tr } \Gamma\{\tilde{\phi}(\psi)\} &= -\frac{\beta_{111}(\tilde{\phi})}{\{\beta_{11}(\tilde{\phi})\}^2} = -\frac{1}{\beta_{11}(\tilde{\phi})}\beta_{11i}(\tilde{\phi})\frac{d}{d\psi}\tilde{\phi}^i(\psi) + O_p(n^{-1/2}) \\ &= -\frac{d}{d\psi} \log \beta_{11}\{\tilde{\phi}(\psi)\} + O_p(n^{-1/2}), \end{aligned}$$

which yields

$$\int_{\hat{\psi}}^{\psi} \text{tr } \Gamma\{\tilde{\phi}(s)\} ds = \log \frac{\beta_{11}(\hat{\phi})}{\beta_{11}(\tilde{\phi})} + O_p(n^{-1}),$$

and (14) follows.

For this case,

$$l_c(\psi) = l_p(\psi) - \frac{1}{2} \log \left[\frac{Q(\psi)\det\{I_{ij}(\psi)\}}{Q(\hat{\psi})\det\{I_{ij}(\hat{\psi})\}} \right] - \log \left\{ \frac{\beta_{11}(\hat{\phi})}{\beta_{11}(\tilde{\phi})} \right\},$$

and

$$T_c(\psi) = \left\{ \frac{\beta_{11}(\hat{\phi})}{\beta_{11}(\check{\phi})} \right\}^{1/2} \frac{\{-l_c^{(2)}(\check{\psi})\}^{1/2}}{l_c^{(1)}(\psi)}, \quad (34)$$

when $\bar{\pi}(\lambda) \propto [\beta_{11}\{\phi(\lambda)\}]^{-1/2}$. Approximation (24) to $l_c(\psi)$ is

$$l_c(\psi) = l_p(\psi) - \frac{1}{2} \log \left[\frac{Q(\psi) \det\{I_{ij}(\psi)\}}{Q(\hat{\psi}) \det\{I_{ij}(\hat{\psi})\}} \right] + \int_{\hat{\psi}}^{\psi} \frac{\beta_{111}\{\tilde{\phi}(s)\}}{[\beta_{11}\{\tilde{\phi}(s)\}]^2} ds + O_p(n^{-1}),$$

and approximation (25) to $T_c(\psi)$ coincides with (34).

Example 3. Consider a sample X_1, \dots, X_n from the gamma distribution having density $f(x) = \{(\nu/\mu)^\nu / \Gamma(\nu)\} x^{\nu-1} \exp\{-(\nu/\mu)x\}$, $x < 0$, with mean μ and shape parameter ν . The parameters μ and ν are orthogonal. Let $\bar{X} = n^{-1} \sum X_i$ and $\bar{X}' = (\prod X_i)^{1/n}$. Then $\hat{\mu} = \bar{X}$, and $\hat{\nu}$ satisfies

$$\rho(\hat{\nu}) = \log \frac{\bar{X}'}{\bar{X}}, \quad \rho(\nu) = \frac{d}{d\nu} \log \Gamma(\nu) - \log \nu.$$

Suppose the mean μ is the parameter of interest. The constrained maximum likelihood estimator $\tilde{\nu}(\mu)$ is given by

$$\rho(\tilde{\nu}) = \log \frac{\bar{X}'}{\bar{X}} + \log \frac{\bar{X}}{\mu} - \frac{\bar{X}}{\mu} + 1,$$

and the signed root of the likelihood ratio statistic is

$$R(\mu) = \text{sgn}(\bar{X} - \mu)[2n\{\zeta(\hat{\nu}) - \zeta(\tilde{\nu})\}]^{1/2}, \quad \zeta(\nu) = \nu \frac{d}{d\nu} \log \Gamma(\nu) - \log \Gamma(\nu) - \nu.$$

The Peers and Stein prior densities are of the form $\pi(\mu, \nu) \propto g(\nu)/\mu$, so

$$T(\mu) = \frac{\mu}{n^{1/2}(\bar{X} - \mu)} \frac{\hat{\nu}^{1/2}}{\tilde{\nu}} \frac{g(\tilde{\nu})}{g(\hat{\nu})} \left\{ \frac{\rho^{(1)}(\hat{\nu})}{\rho^{(1)}(\tilde{\nu})} \right\}^{1/2},$$

where $\rho^{(1)}(\nu) = d\rho(\nu)/d\nu = d^2 \log \Gamma(\nu)/d\nu^2 - 1/\nu$.

For the choice $g(\nu) = \nu \rho^{(1)}(\nu)$,

$$T(\nu) = \frac{\mu}{n^{1/2}(\bar{X} - \mu)} \frac{1}{\hat{\nu}^{1/2}} \left\{ \frac{\rho^{(1)}(\tilde{\nu})}{\rho^{(1)}(\hat{\nu})} \right\}^{1/2}. \quad (35)$$

It follows from Barndorff-Nielsen (1986, 1987) that for this particular $T(\mu)$ the solutions to equations (4) have coverage error of order $O(n^{-3/2})$. Thus, when the prior density is $\pi(\mu, \nu) \propto \nu \rho^{(1)}(\nu)/\mu$, the posterior quantiles of μ are approximate confidence limits having coverage $1 - \alpha + O(n^{-3/2})$. A drawback of (35) is that it requires calculation of the trigamma function. Note that the choice $g(\nu) = \{\rho^{(1)}(\nu)\}^{1/2}$ produces

$$T(\mu) = \frac{\mu}{n^{1/2}(\bar{X} - \mu)} \frac{\hat{\nu}^{1/2}}{\tilde{\nu}}, \quad (36)$$

which is computationally simpler than (35). Table 1 shows simulated coverage probabilities for approximate confidence limits obtained from equations (4) with T given by (35) and (36). Although (35) yields slightly better coverage accuracy than (36), both versions of T perform well.

Apart from an additive constant, $l_c(\mu)$ is given by

$$l_c(\mu) = n\zeta(\tilde{\nu}) - \frac{1}{2} \log \rho^{(1)}(\tilde{\nu}),$$

and $T_c(\mu)$ is given by

$$T_c(\mu) = \frac{\check{\mu}}{\mu} \frac{g(\tilde{\nu})}{g(\check{\nu})} \frac{\{-l_c^{(2)}(\check{\mu})\}^{1/2}}{l_c^{(1)}(\mu)}.$$

For the choice $g(\nu) = \nu \rho^{(1)}(\nu)$, $T_c(\mu)$ becomes

$$T_c(\mu) = \frac{\check{\mu}}{\mu} \frac{\check{\nu} \rho^{(1)}(\tilde{\nu})}{\check{\nu} \rho^{(1)}(\check{\nu})} \frac{\{-l_c^{(2)}(\check{\mu})\}^{1/2}}{l_c^{(1)}(\mu)},$$

and the approximate confidence limits obtained by solving equations (20) with this $T_c(\mu)$ have coverage error of order $O(n^{-3/2})$.

Now suppose the shape parameter ν is of interest. Then $\hat{\mu}(\nu) = \bar{X}$, and

$$R(\nu) = \text{sgn}(\hat{\nu} - \nu) (2n[\zeta(\hat{\nu}) - \zeta(\nu) - \nu\{\rho(\hat{\nu}) - \rho(\nu)\}])^{1/2}.$$

The Peers and Stein prior densities are $\pi(\mu, \nu) \propto \{\rho^{(1)}(\nu)\}^{1/2} g(\mu)$, and

$$T(\nu) = \frac{1}{n^{1/2}\{\rho(\hat{\nu}) - \rho(\nu)\}} \left(\frac{\hat{\nu}}{\nu} \right)^{1/2} \{\rho^{(1)}(\nu)\}^{1/2}.$$

It follows from (32) that if $T(\nu)$ is replaced by

$$U(\nu) = \frac{1}{n^{1/2}(\hat{\nu} - \nu)} \left(\frac{\hat{\nu}}{\nu} \right)^{1/2} \frac{1}{\{\rho^{(1)}(\hat{\nu})\}^{1/2}},$$

then the approximate confidence limits obtained from equations (4) have coverage $1 - \alpha + O(n^{-3/2})$. Note that $T(\nu)$ is independent of the choice of $g(\mu)$, since $\tilde{\mu}(\nu) = \hat{\mu}$. Hence it is impossible to find a prior density of the Peers and Stein type so that $T(\nu)$ and $U(\nu)$ agree, which suggests that confidence limits with coverage error of order $O(n^{-3/2})$ cannot be constructed for ν using the Bayesian approach.

For this case,

$$l_c(\nu) = n[\zeta(\nu) + \nu\{\rho(\hat{\nu}) - \rho(\nu)\}] - \frac{1}{2}\log\nu, \quad T_c(\nu) = \left\{ \frac{\rho^{(1)}(\nu)}{\rho^{(1)}(\check{\nu})} \right\}^{1/2} \frac{\{-l_c^{(2)}(\check{\nu})\}^{1/2}}{l_c^{(1)}(\nu)}.$$

Again, it follows from (32) that if $T_c(\nu)$ is replaced by

$$U_c(\nu) = \frac{1}{(\check{\nu} - \nu)\{-l_c^{(2)}(\check{\nu})\}^{1/2}},$$

then the confidence limits obtained from equations (20) have coverage $1 - \alpha + O(n^{-3/2})$.

References

- Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307-22.
- Barndorff-Nielsen, O. E. (1987). Discussion of paper by D. R. Cox and N. Reid. *J. R. Statist. Soc. B* **49**, 18-20.
- Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557-63.
- Barndorff-Nielsen, O. E. & Chamberlain, S. R. (1992). An ancillary invariant modification of the signed log likelihood ratio. *Scand. J. Statist.* To appear.
- Barndorff-Nielsen, O. E. & Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc. B* **41**, 279-38.

- Coddington, E. A. & Levinson, N. (1955). *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.
- Cox, D. R. & Reid, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. R. Statist. Soc. B* **49**, 1-39.
- Davison, A.C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc. B* **50**, 445-61.
- DiCiccio, T. J. & Martin, M. A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika* **78**, 891-902.
- DiCiccio, T. J. & Martin, M. A. (1992). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc. B*. To appear.
- Fraser, D. A. S. (1991). Statistical inference: Likelihood to significance. *J. Am. Statist. Assoc.* **86**, 258-65.
- Fraser, D. A. S., Reid, N. & Wong, A. (1991). Exponential linear models: a two-pass procedure for saddlepoint approximation. *J. R. Statist. Soc. B* **53**, 483-92.
- McCullagh, P. (1984). Local sufficiency. *Biometrika* **71**, 233-44.
- Peers, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc. B* **27**, 9-16.
- Pierce, D. A. & Peters, D. (1992). Practical use of higher-order asymptotics for multiparameter exponential families. *J. R. Statist. Soc. B*. To appear.
- Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions. *J. Appl. Probab.* **24**, 875-87.
- Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach center publications, **16**. Warsaw: PWN-Polish scientific publishers.
- Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-8.

Tierney, L., Kass, R. E. & Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**, 425-33. Correction note. *Biometrika* (1991). **76**, 233-34.

Welch, B. L. & Peers, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *J. R. Statist. Soc. B* **25**, 318-29.

**Table 1. Simulated coverage probabilities of approximate upper
 $1 - \alpha$ confidence limits for gamma mean μ**

$(\mu, \nu) = (5, 0.25)$

	$n = 5$				$n = 10$			
$1 - \alpha$	0.01	0.05	0.95	0.99	0.01	0.05	0.95	0.99
A	4.070	12.706	95.496	98.802	2.349	9.070	95.907	99.106
$B(35)$	1.335	5.502	94.844	98.843	1.040	5.061	94.964	98.964
(36)	1.434	5.767	94.686	98.777	1.080	5.155	94.888	98.931
$C(35)$	1.365	5.669	94.866	98.869	1.055	5.102	94.969	98.967
(36)	1.460	5.912	94.707	98.804	1.091	5.200	94.893	98.934

$(\mu, \nu) = (1, 1)$

	$n = 5$				$n = 10$			
$1 - \alpha$	0.01	0.05	0.95	0.99	0.01	0.05	0.95	0.99
A	3.351	10.555	93.560	98.114	1.976	7.702	94.811	98.783
$B(35)$	1.378	5.646	94.549	98.688	1.055	5.128	94.737	98.954
(36)	1.459	5.903	94.347	98.616	1.099	5.260	94.831	98.912
$C(35)$	1.380	5.716	94.586	98.738	1.054	5.141	94.943	98.963
(36)	1.457	5.952	94.385	98.667	1.098	5.270	94.836	98.921

$(\mu, \nu) = (0.5, 10)$

	$n = 5$				$n = 10$			
$1 - \alpha$	0.01	0.05	0.95	0.99	0.01	0.05	0.95	0.99
A	3.039	9.391	91.990	97.508	1.738	6.890	93.896	98.488
$B(35)$	1.519	5.878	94.301	98.528	1.095	5.154	94.867	98.900
(36)	1.592	5.919	94.264	98.526	1.103	5.179	94.844	98.892
$C(35)$	1.541	5.881	94.330	98.578	1.081	5.149	94.877	98.916
(36)	1.551	5.924	94.291	98.567	1.089	5.173	94.854	98.907

A , B and C refer to limits obtained by solving the equations $\Phi(R) = \alpha$, $\Phi(R) + \varphi(R)(R^{-1} - T) = \alpha$ and $\Phi\{R - R^{-1} \log(RT)\} = \alpha$, respectively. Table entries are percentages based on 1,000,000 trials.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 450	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Accurate Procedures for Approximate Bayesian and Conditional Inference Without the Need for Orthogonal Parameters		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas J. DiCiccio, Joseph B. Keller, and Michael A. Martin		8. CONTRACT OR GRANT NUMBER(s) N00025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE February 24, 1992
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 30, if different from Report)		
18. SUPPLEMENTARY NOTES Also issued as Technical Report 391, Department of Statistics, Stanford, NSF Grant No. DMS-89-05974		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Asymptotic normality; conditional profile likelihood; confidence limit; exponential family; gamma distribution; marginal density approximation; noninformative prior; nuisance parameter; orthogonal parameterization; profile likelihood; signed root likelihood statistic; tail probability approximation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) see attached		

Unclassified

~~SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)~~

This paper concerns methods to construct approximate confidence limits for a scalar parameter ψ in the presence of nuisance parameters. The methods are based on Bayesian procedures discussed by Peers (1965) and Stein (1985), in which prior density is chosen so that the posterior quantiles of ψ are approximate confidence limits with coverage error of order $O(n^{-1})$ under repeated sampling. Multidimensional integration of the posterior density is avoided by using approximations of marginal densities and distribution functions; thus, adjustments are obtained that improve the standard normal approximation to the distributions of signed roots of the profile and conditional likelihood ratio statistics for ψ . The necessary prior densities are easy to specify when the nuisance parameters are orthogonal to the parameter of interest, and this simplicity is exploited in developing the methods. However, the need for explicit specification of an orthogonal parameterization is alleviated by approximating the Jacobian of a transformation to orthogonality. The methods are illustrated and compared with other procedures in some examples involving exponential families.

Unclassified