ON CONFIDENCE INTERVALS ASSOCIATED WITH THE USUAL AND ADJUSTED LIKELIHOODS

Rahul Mukerjee
Indian Institute of Management
Post Box No 16757 Calcutta 700 027, India

N. Reid
Department of Statistics
University of Toronto
Toronto, Canada M5S 3G3

Abstract: In the presence of nuisance parameters, we derive an explicit higher order asymptotic formula to compare the expected lengths of confidence intervals given by likelihood ratio statistics arising from the usual profile likelihood and various adjustments thereof. Highest posterior density regions, with approximate frequentist validity, are also included in the study.

Keywords: Bartlett correction; expected length; highest posterior density region; likelihood ratio statistic; parameter orthogonality; point estimation.

1. Introduction

In recent years, there has been considerable interest in various adjustments to the profile likelihood in the presence of nuisance parameters – see Cox and Reid (1987), McCullagh and Tibshirani (1990), Barndorff-Nielsen (1994) and the references therein. DiCiccio and Stern (1994) investigated, in a general framework, the frequentist and Bayesian Bartlett corrections for the likelihood ratio statistics arising from adjusted likelihoods while Mukerjee (1992) and Ghosh and Mukerjee (1994) considered the related problem of higher order power under contiguous alternatives.

The present paper aims at examining another issue of statistical interest, the expected lengths of confidence intervals given by likelihood ratio statistics arising from the usual profile likelihood and various adjustments thereof. Starting from a very general form of the adjusted profile likelihood, an explicit higher order asymptotic formula is obtained for this purpose and the link with the corresponding point estimation problem is also indicated.

The formula is readily applicable to specific models and, for a given model, helps in finding a good adjustment to the profile likelihood under the criterion of expected length. Our general framework, akin to that in DiCiccio and Stern (1994), allows us to include highest posterior density regions with approximate frequentist validity in our study. Somewhat surprisingly, our explicit formula, when applied to examples, reveals that even under the purely frequentist criterion of expected length, such highest posterior density regions tend to perform better than their competitors.

We refer to Kumon and Amari (1983) and Mukerjee and Reid (1998) for some earlier work on expected lengths of confidence intervals. These authors, however, did not consider models with nuisance parameters and hence the question of considering adjusted likelihoods did not arise in their settings. It may also be noted that, because of well-known duality relations, the findings in Mukerjee (1992) and Ghosh and Mukerjee (1994) on higher order power shed light on the power properties of the associated confidence intervals; cf. Amari (1985; Section 6.5). Nevertheless, the behaviour of the confidence intervals with respect to expected length, which is also of interest and perhaps more natural in the present context, still remains uncertain and this motivates the present work. Incidentally, unlike those on power, our results are non-local.

In this paper, we are primarily concerned with models where the interest and nuisance parameters, namely θ and λ respectively, are both one-dimensional. One reason for considering scalar θ is that some of the adjustments to the profile likelihood that are of special interest to us, i.e., those suggested by McCullagh and Tibshirani (1990) and Barndorff-Nielsen (1994), may not be well-defined for multi-dimensional θ. Restriction to scalar λ is, however, only for notational simplicity and, in the last section, we briefly address the case of the multi-dimensional λ as well.

2. Preliminaries

Let $\{X_s\}$, $s \geq 1$, be a sequence of independent and identically distributed possibly vector-valued random variables with common density $f(x;\theta,\lambda)$, where θ is the interest parameter and λ is the nuisance parameter, both one-dimensional. The parameter space is an open subset of R^2. Let n be the sample size, $\ell(\theta,\lambda)=n^{-1}\sum_{s=1}^{n}\log f(X_s;\theta,\lambda)$, $D_1\equiv\partial/\partial\theta$, $D_2\equiv\partial/\partial\lambda$, and for i, j, i', $j'=0,1,2,\ldots$, define

$$\left.\begin{array}{l}\ell_{ij}\equiv\ell_{ij}(\theta,\lambda)=D_1^iD_2^j\ell(\theta,\lambda), Z_{ij}(\theta,\lambda)=n^{1/2}(\ell_{ij}-K_{ij}),\\ K_{ij}\equiv K_{ij}(\theta,\lambda)=E_{\theta,\lambda}\{D_1^iD_2^j\log f(X_1;\theta,\lambda)\}, a_{ij}\equiv a_{ij}(\theta,\lambda)=-K_{ij}(\theta,\lambda)\\ K_{ij\cdot i'j'}\equiv K_{ij\cdot i'j'}(\theta,\lambda)=E_{\theta,\lambda}[\{D_1^iD_2^j\log f(X_1;\theta,\lambda)\}\{D_1^{i'}D_2^{j'}\log f(X_1;\theta,\lambda)\}]\ .\end{array}\right\}\quad(2.1)$$

Similarly, $K_{ij\cdot i'j'\cdot i''j''}$ etc are defined. Among the K_{ij}, $K_{ij\cdot i'j'}$ etc, only those which are used in the sequel are assumed to exist; it is also assumed that these are smooth functions of θ and λ. Since θ is one-dimensional, without loss of generality, we assume global parametric orthogonality, i.e., $K_{10\cdot 01}\equiv 0$ (Cox and Reid, 1987). Then, under standard regularity conditions, the per observation information matrix at (θ,λ) is $\mathrm{diag}(a_{20},a_{02})$ which is assumed to be positive definite for each (θ,λ).

Let $\hat\lambda_\theta$ be the maximum likelihood estimator of λ given θ and $X=(X_1,\ldots,X_n)'$. As in DiCiccio and Stern (1994), we consider a general form of adjusted profile likelihood given by

$$n\tilde\ell(\theta)=n\{\ell(\theta,\hat\lambda_\theta)+n^{-1}B(\theta)\}\ ,\tag{2.2}$$

where we assume that $B(\theta)(\equiv B(\theta,X))$ is smooth and that $B(\theta)$ and its derivatives with respect to θ are at most of order $O_p(1)$. As will be seen, with $B(\theta)$ appropriately chosen, (2.2) not only covers the various adjustments proposed in the literature but also enables us to study highest posterior density regions with approximate frequentist validity. If $B(\theta)\equiv 0$ then (2.2) reduces to the usual unadjusted profile likelihood.

The likelihood ratio statistic arising from (2.2) is given by

$$\tilde T(\theta)=2n\{\tilde\ell(\tilde\theta)-\tilde\ell(\theta)\}\tag{2.3}$$

where $\tilde\ell(\tilde\theta)=\sup_\theta\tilde\ell(\theta)$. Using parametric orthogonality, analogously to equation (7) of DiCiccio and Stern (1994), we have

$$\tilde\theta=\hat\theta-(n\hat\ell_{20})^{-1}B_1(\hat\theta)+O_p(n^{-2})\ ,\tag{2.4}$$

where $\hat\ell_{ij}=\ell_{ij}(\hat\theta,\hat\lambda)$, $\hat\theta$ is the maximum likelihood estimator of θ, $\hat\lambda(=\hat\lambda_{\hat\theta})$ is the maximum likelihood estimator of λ and $B_i(\theta)=D_1^iB(\theta)$. As noted in DiCiccio and

Stern (1994), the statistic $\tilde{T}(\theta)$ admits a Bartlett correction. Denoting this Bartlett correction factor by $1 + n^{-1}\tilde{F}(\theta, \lambda)$, then

$$P_{\theta,\lambda}[\tilde{T}(\theta)/\{1 + n^{-1}\tilde{F}(\hat{\theta}, \hat{\lambda})\} \leq z^2] = 1 - \alpha + o(n^{-1}) \ , \tag{2.5}$$

where z is the upper $\frac{1}{2}\alpha$-point of a standard normal variate. Following DiCiccio and Stern (1994), it can also be seen that

$$\tilde{F}(\theta, \lambda) = F(\theta, \lambda) + \tilde{F}^*(\theta, \lambda) \ , \tag{2.6}$$

where $1 + n^{-1}F(\theta, \lambda)$ represents the Bartlett correction factor for the likelihood ratio statistic arising from the usual unadjusted profile likelihood,

$$\tilde{F}^*(\theta, \lambda) = 2D_1(a_{20}^{-1}\gamma_1) - a_{20}^{-1}(a_{20}^{-1}K_{30} + a_{02}^{-1}K_{12})\gamma_1 + a_{20}^{-1}(\gamma_1^2 - \gamma_2) \ , \tag{2.7}$$

and $\gamma_i = \gamma_i(\theta, \lambda)$ $(i = 1, 2)$ are non-stochastic quantities such that

$$B_i(\hat{\theta}) = \gamma_i(\theta, \lambda) + o_p(1) \ . \tag{2.8}$$

Ghosh and Mukerjee (1992) give an explicit expression for $F(\theta, \lambda)$ which, obviously, does not depend on $B(\cdot)$.

3. Some key formulae

3.1 Notation

In order to consider a confidence interval for θ, define

$$\tilde{g}_i(\theta) = D_1^i\tilde{\ell}(\theta) \ , \ \tilde{g}_i = \tilde{g}_i(\tilde{\theta}) \ (i = 1, 2, \ldots) \ , \tag{3.1}$$

$$\tilde{Q} = -\tilde{g}_2 \ , \ \tilde{u} = \frac{1}{6}\tilde{g}_3/\tilde{g}_2 \ , \ \tilde{w} = \frac{1}{72}(3\tilde{g}_4\tilde{g}_2 - \tilde{g}_3^2)/\tilde{g}_2^2 \ . \tag{3.2}$$

Since $\tilde{g}_1 = 0$ by the definition of $\tilde{\theta}$, from (2.3), (3.1) and (3.2),

$$\tilde{T}(\theta) = \{\tilde{Q}^{1/2}(h + n^{-1/2}\tilde{u}h^2 + n^{-1}\tilde{w}h^3)\}^2 + o_p(n^{-1}) \ ,$$

where $h \equiv h(\theta) = n^{1/2}(\theta - \tilde{\theta})$. Hence, considering a "square root version" of $\tilde{T}(\theta)$, it can be seen from (2.5) that a confidence interval for θ, based on $\tilde{T}(\theta)$, with confidence coefficient $1 - \alpha + o(n^{-1})$ is given by $\{C_1(X), C_2(X)\}$, where

$$C_i(X) = \tilde{\theta} + (-1)^i (n\tilde{Q})^{-1/2} z\{1 + \frac{1}{2} n^{-1} \tilde{F}(\hat{\theta}, \hat{\lambda})\} - (n\tilde{Q})^{-1} \tilde{u} z^2 \\ + (-1)^i (n\tilde{Q})^{-3/2} (2\tilde{u}^2 - \tilde{w}) z^3 \quad (i = 1, 2) . \tag{3.3}$$

Denoting the length of this interval by $\tilde{L}$, clearly

$$n^{1/2} \tilde{L} = 2z\tilde{Q}^{-1/2}[1 + n^{-1}\{\frac{1}{2}\tilde{F}(\hat{\theta}, \hat{\lambda}) + (2\tilde{u}^2 - \tilde{w})(z^2/\tilde{Q})\}] \tag{3.4}$$

3.2 Expected length

For $i = 1, 2, \ldots$, let $g_i(\theta) = D_1^i \ell(\theta, \hat{\lambda}_\theta)$ and $g_i = g_i(\hat{\theta})$. Then by (2.2) and (3.1), $\tilde{g}_i(\hat{\theta}) = g_i + n^{-1} B_i(\hat{\theta})$, so that by (2.4) and (3.2),

$$\tilde{Q} = Q - n^{-1}\{B_2(\hat{\theta}) - B_1(\hat{\theta})(g_3/\hat{\ell}_{20})\} + o_p(n^{-1}) , \tag{3.5}$$

$$\tilde{u} = u + o_p(1) , \ \tilde{w} = w + o_p(1) , \tag{3.6}$$

where Q, u and w are as in (3.2) with the $\tilde{g}_i$ there replaced by g_i. From (2.6) and (3.4)-(3.6),

$$n^{1/2}\tilde{L} = n^{1/2}L + n^{-1} z Q^{-1/2}[\tilde{F}^*(\hat{\theta}, \hat{\lambda}) + Q^{-1}\{B_2(\hat{\theta}) - B_1(\hat{\theta})(g_3/\hat{\ell}_{20})\}] + o_p(n^{-1}) , \tag{3.7}$$

where, analogously to (3.4),

$$L = 2z(nQ)^{-1/2}[1 + n^{-1}\{\frac{1}{2}F(\hat{\theta}, \hat{\lambda}) + (2u^2 - w)(z^2/Q)\}]$$

is the length of the confidence interval given by the likelihood ratio statistic arising from the usual unadjusted profile likelihood. Note that L does not depend on $B(\cdot)$.

By (2.1) and the definition of $\hat{\lambda}_\theta$, we have $\ell_{01}(\theta, \hat{\lambda}_\theta) = 0$. Differentiating this with respect to θ and using parametric orthogonality, $(d\hat{\lambda}_\theta/d\theta)_{\hat{\theta}} = o_p(1)$ and $(d^2\hat{\lambda}_\theta/d\theta^2)_{\hat{\theta}} = O_p(1)$. Hence from the definitions of Q and g_3, one can check that

$$Q = -\hat{\ell}_{20} + o_p(1) , \quad g_3 = \hat{\ell}_{30} + o_p(1) . \tag{3.8}$$

From (2.1), (2.7), (2.8), (3.7) and (3.8), a formula for expected length follows as

$$E_{\theta,\lambda}(n^{1/2}\tilde{L}) = E_{\theta,\lambda}(n^{1/2}L) + n^{-1}za_{20}^{-1/2}V(\theta,\lambda) + o(n^{-1}) \ , \tag{3.9}$$

where

$$\begin{aligned} V(\theta,\lambda) &= \tilde{F}^*(\theta,\lambda) + a_{20}^{-1}(\gamma_2 + \gamma_1 K_{30} a_{20}^{-1}) \\ &= 2D_1(a_{20}^{-1}\gamma_1) - (a_{20}a_{02})^{-1}K_{12}\gamma_1 + a_{20}^{-1}\gamma_1^2 \ . \end{aligned} \tag{3.10}$$

Since $E_{\theta,\lambda}(n^{1/2}L)$ does not involve $B(\cdot)$, in view of (3.9), the expected lengths of confidence intervals for various choices of $B(\cdot)$ can be compared on the basis of $V(\theta,\lambda)$ alone.

3.3 Link with point estimation

It is of interest to examine any possible connection between the mean squared error of the estimator $\tilde{\theta}$ given by the adjusted likelihood (2.2) and the expected length (3.9) of the associated confidence interval. To that effect, from (2.4) after lengthy algebra, we obtain

$$nE_{\theta,\lambda}(\tilde{\theta}-\theta)^2 = nE_{\theta,\lambda}(\hat{\theta}-\theta)^2 + (na_{20})^{-1}V^*(\theta,\lambda) + o(n^{-1}) \ , \tag{3.11}$$

where

$$V^*(\theta,\lambda) = V(\theta,\lambda) + a_{20}^{-2}(2K_{10\cdot 20} + K_{30})\gamma_1 \ , \tag{3.12}$$

$V(\theta,\lambda)$ being as in (3.10). In the right-hand side of (3.11), the first term does not depend on $B(\cdot)$ while the second term is influenced by this adjustment. The similarity between (3.9) and (3.11) is noteworthy. In particular, by (3.12), for models with $2K_{10\cdot 20} + K_{30} = 0$ (e.g. the univariate normal model with variance θ and mean λ), this exhibits a direct link between mean squared error and expected length.

4. Application and examples

4.1 Some frequentist adjustments

As indicated in the Introduction, of special interest to us are the adjustments to the profile likelihood as proposed in Cox and Reid (1987), McCullagh and Tibshirani

(1990) and Barndorff-Nielsen (1994). The last paper (Section 3), in fact, proposes two adjustments. Proceeding as in Ghosh and Mukerjee (1994) or DiCiccio and Stern (1994) and using parametric orthogonality, it can be seen that these adjusted likelihoods are all of the form (2.2) with (see (2.8))

$$\gamma_1 = \frac{1}{2} a_{02}^{-1} K_{12} \ . \tag{4.1}$$

This, combined with (3.10), shows that all the adjustments mentioned above are equivalent up to the order of approximation stated in (3.9), with respect to expected lengths of confidence intervals. Therefore, for comparative purposes, hereafter we consider only the Cox and Reid (1987) adjustment. Incidentally, $a_{02}^{-1} K_{12}$ is invariant with respect to the choice of the orthogonal nuisance parameter λ and hence, by (4.1), so is the right hand side of (3.10) for the Cox and Reid (1987) adjustment.

4.2 Expected length: examples

We now illustrate the use of (3.9) and (3.10) considering adjusted likelihoods of the form (2.2) with

$$B(\theta) = b \log\{-\ell_{02}(\theta, \hat{\lambda}_\theta)\} \ , \tag{4.2}$$

the constant b being free from θ, λ and n. The form (4.2) corresponds to the Cox and Reid (1987) adjustment if $b = -1/2$ and the usual unadjusted profile likelihood if $b = 0$. Under (4.2), by (2.8), $\gamma_1 = -b a_{02}^{-1} K_{12}$, so that the expression for $V(\theta, \lambda)$, as given in (3.10), reduces to

$$V_b(\theta, \lambda) = (b + b^2)(a_{20} a_{02}^2)^{-1} K_{12}^2 - 2b D_1\{(a_{20} a_{02})^{-1} K_{12}\} \ . \tag{4.3}$$

Choices of $B(\cdot)$, which are not of the form (4.2) can as well be of interest. These include choices dictated by Bayesian considerations, an issue which will be discussed in the next subsection where the examples will also be revisited.

Example 1. For the univariate normal model with variance θ and mean λ, parametric orthogonality holds and $a_{20} = \frac{1}{2}\theta^{-2}$, $a_{02} = \theta^{-1}$, $K_{12} = \theta^{-2}$. By (4.3), then $V_b(\theta, \lambda) = 2(b^2 - b) = \psi(b)$, say, which is minimum at $b = 1/2$. Note that $\psi(1/2) < \psi(0) < \psi(-1/2)$.

Example 2. For the univariate Cauchy model with scale parameter θ and location parameter λ, we again have $V_b(\theta,\lambda)=2(b^2-b)$ and the same conclusion as in the last example holds.

Example 3. This relates to the ratio of independent exponential means. Let

$$f(x;\theta,\lambda)=(\mu_1\mu_2)^{-1}\exp\{-(\mu_1^{-1}x^{(1)}+\mu_2^{-1}x^{(2)})\},\quad x^{(1)},x^{(2)}>0\ ,$$

where $x=(x^{(1)},x^{(2)})'$, $\mu_1=\theta^{1/2}\lambda$, $\mu_2=\theta^{-1/2}\lambda$, and $\theta,\lambda>0$. Note that $\theta=\mu_1/\mu_2$. Then parametric orthogonality holds and $K_{12}=0$, i.e., by (4.3) all choices of b are equivalent with respect to expected length.

Example 4. For the bivariate normal model with correlation coefficient θ, both means λ and both variances unity, parametric orthogonality holds and $a_{20}=(1+\theta^2)(1-\theta^2)^{-2}$, $a_{02}=2(1+\theta)^{-1}$, $K_{12}=2(1+\theta)^{-2}$. Hence by (4.3),

$$V_b(\theta,\lambda)=(1+\theta^2)^{-2}(1-\theta)\{(b+b^2)(1-\theta)(1+\theta^2)+2b(1+5\theta+\theta^2+\theta^3)\}=\psi(\theta;b)\ ,$$

say, and no choice of b minimizing $\psi(\theta;b)$, uniformly in θ, is possible. In particular, for $b=0$ and $-1/2$, which correspond to the usual and the Cox and Reid (1987) adjusted likelihoods respectively, we have

$$\psi(\theta;0)=0\ ,\quad \psi(\theta;-\frac{1}{2})=-\frac{1}{4}(1+\theta^2)^{-2}(1-\theta)(5+19\theta+5\theta^2+3\theta^3)\ . \qquad (4.4)$$

Note that $\psi(\theta;-\frac{1}{2})<\psi(\theta;0)$ if and only if $-0.28\le\theta<1$.

The above examples reveal that, under the criterion of expected length, the confidence interval arising from the Cox and Reid (1987) adjustment does not necessarily perform better than that given by the unadjusted profile likelihood. Surprisingly, in some situations, as illustrated by Examples 1 and 2, an adjustment in the reverse direction might help in this regard.

Closed form expressions for the point estimator $\tilde{\theta}$, associated with the adjustment (4.2), are available in Examples 1 and 3. In Example 1, $\tilde{\theta}=\sum_{s=1}^{n}(X_s-\overline{X})^2/(n+2b)$, where $\overline{X}$ is the sample mean, and the mean squared error of $\tilde{\theta}$ is also minimum at $b=\frac{1}{2}$, which gives the minimum risk equivariant estimator under squared error loss. In Example

3, $\tilde{\theta}$ equals the corresponding ratio of sample means for every b, i.e., all choices of b are again equivalent.

4.3. Highest posterior density intervals

The present results can be used to study expected lengths of confidence intervals which arise from consideration of highest posterior density and have frequentist validity up to $o(n^{-1})$. Denoting the possibly improper prior density corresponding to such an interval by $\pi \equiv \pi(\theta, \lambda)$, for this purpose one has to take (cf. DiCiccio and Stern (1994))

$$B(\theta) = -\frac{1}{2}\log\{-\ell_{02}(\theta, \hat{\lambda}_\theta)\} + \log \pi(\theta, \hat{\lambda}_\theta)$$

in (2.2) and hence

$$\gamma_1 = \frac{1}{2}a_{02}^{-1}K_{12} + \pi^{-1}D_1\pi \tag{4.5}$$

in (3.10). We refer to DiCiccio and Stern (1994) and Ghosh and Mukerjee (1995) for characterization, under appropriate regularity conditions, of priors ensuring approximate frequentist validity of the highest posterior density region.

For the purpose of illustration, we revisit Examples 1-4 considering the priors (i) $\pi(\theta, \lambda) = \theta^{-1}$, (ii) $\pi(\theta, \lambda) = \theta^{-1}$, (iii) $\pi(\theta, \lambda) = (\theta\lambda)^{-1}$ and (iv) $\pi(\theta, \lambda) = (1-\theta^2)^2/(1+\theta^2)$, which yield highest posterior density intervals with frequentist validity, up to $o(n^{-1})$, in these examples respectively; cf. Ghosh and Mukerjee (1995). In Examples 1 or 2, with π as in (i) or (ii) respectively, by (3.10) and (4.5) we have $V(\theta, \lambda) = -1/2$ so that the corresponding highest posterior density intervals have shorter expected lengths than the ones arising from the unadjusted profile likelihood and the Cox and Reid (1987) adjustment thereof. The same conclusion is seen to hold in Example 3 with π as in (iii). Finally, in Example 4, with π as in (iv), by (3.10) and (4.5), for the highest posterior density interval

$$V(\theta, \lambda) = \psi(\theta; -1/2) + 4(1+\theta^2)^{-3}(2\theta^6 + 9\theta^4 + 24\theta^2 - 3)$$

where $\psi(\theta; -1/2)$ is an in (4.4). Comparison with (4.4) shows that the highest posterior density interval has shorter expected length than the one arising from the usual profile likelihood if and only if $-0.34 \le \theta \le 0.37$ and the one arising from the Cox and Reid (1987) adjustment thereof if and only if $|\theta| \le 0.34$.

Thus, in the examples, the highest posterior density interval performs well even under the frequentist criterion of expected length. Indeed, in the posterior setup the highest posterior density interval is the shortest one. It appears that, under approximate frequentist validity, this property is carried over, to some extent, to the frequentist setup.

Remark 1. In general, from (3.9), (3.10), (4.1) and (4.5), it may be seen that the highest posterior density interval, corresponding to a prior π , has a shorter expected length than the one arising from the Cox and Reid (1987) adjusted likelihood if and only if

$$2D_1\left(a_{20}^{-1}\frac{D_1\pi}{\pi}\right)+a_{20}^{-1}\left(\frac{D_1\pi}{\pi}\right)^2<0\ . \tag{4.6}$$

The conclusions made while revisiting Examples 1-4 are in conformity with (4.6).

Remark 2. It is of interest to examine how far our asymptotic results match the small sample behaviour of the confidence intervals. To this effect, we again consider Example 1 where exact computations are possible. Continuing with adjustments of the form (4.2), intricate algebra shows that the expected length of an exact confidence interval arising from the corresponding likelihood ratio statistic is minimum if and only if $b=\frac{1}{2}$, which is precisely the same as our asymptotic findings. Table 1 below shows $n^{1/2}\theta^{-1}$ times the exact expected lengths of the confidence intervals, with confidence coefficient 0.95 , for $n=3,5,7$ and $b=-\frac{1}{2}$, 0 and $\frac{1}{2}$. Recall that $b=0$ and $-\frac{1}{2}$ correspond to the unadjusted profile likelihood and the Cox and Reid (1987) adjustment thereof. Also, by (4.5), the case $b=\frac{1}{2}$ corresponds to the highest posterior density interval for the prior $\pi(\theta,\lambda)=\theta^{-1}$. Table 1 shows that the choice $b=0$ performs close to the optimal choice $b=\frac{1}{2}$ and that either of these choices entails a nonignorable gain over the choice $b=-\frac{1}{2}$. In the setup of Example 1, it is also possible to compute the exact coverage of the approximate confidence limits specified by (3.3). For $1-\alpha=0.95$, such exact computations show that the coverage of these approximate limits lies in the range 0.95 ± 0.02 whenever n exceeds 4 , 10 or 18 for $b=-\frac{1}{2}$, 0 or $\frac{1}{2}$ respectively. Exact computations are possible also in the context of Example 3 and these too can be seen to be in agreement with the asymptotic findings reported in this and the preceding subsection.

Table 1. $n^{1/2}\theta^{-1}$ times exact expected lengths $(1-\alpha=0.95)$

n \ b	$-\frac{1}{2}$	0	$\frac{1}{2}$
3	40.5220	33.9046	33.6213
5	14.0365	12.3897	12.2041
7	10.1414	9.2265	9.0811

Remark 3.

In all the examples above the function $V(\theta,\lambda)$ is free of the nuisance parameter λ, and is thus the same for any version of the nuisance parameter that is orthogonal to θ. This is true in general for the nonBayesian adjustments discussed in Section 4.1 and in equation (4.2), as pointed out in the last sentence of that subsection. It is also true for highest posterior density intervals. From equation (3.6) of Ghosh and Mukerjee (1995), one can check that if the prior $\pi(\theta,\lambda)$ ensures frequentist validity of the highest posterior density interval, up to $o(n^{-1})$, in the (θ,λ)-parametrization then so does the transformed version of $\pi(\theta,\lambda)$ under any reparametrization which keeps theta unaffected, replaces λ by a one-one transformation thereof and thus retains parametric orthogonality. Hence from (3.10) and (4.5) it follows that the quantity γ_1 and consequently the expression for $V(\theta,\lambda)$ for the highest posterior density interval remain invariant over the choice of the orthogonal nuisance parameter λ. Hence any comparison among intervals associated with these adjustments or between any of these intervals and the highest posterior density interval with approximate frequentist validity remains unaffected by the choice of the orthogonal nuisance parameter λ.

Nevertheless, as pointed out by a referee, there may be other adjustments for which $V(\theta,\lambda)$ is not invariant, although one might consider that the adjustments studied above are more appealing for that reason. However, this is only an asymptotic invariance property: an unsatisfactory aspect of the adjusted likelihood of Cox and Reid (1987) is its potential sensitivity to the choice of nuisance parameter in any given problem with finite sample size.

Remark 4. Extension of our results to the case of a multi-dimensional nuisance parameter λ is possible with heavier notation and algebra. As an illustration, we indicate how the formulae (3.9) and (3.10) change with a multi-dimensional $\lambda = (\lambda_2, \ldots, \lambda_q)'$. Let $I = ((I_{ij}))$ be the $q \times q$ per observation information matrix at (θ, λ), where I_{11} corresponds to θ, $I_{ij} (2 \leq i, j \leq q)$ correspond to λ, and parametric orthogonality holds, i.e. $I_{1j} = 0 \ (2 \leq j \leq q)$. Furthermore, let $I^{-1} = ((I^{ij}))$, $D_1 \equiv \partial/\partial\theta$, $D_j = \partial/\partial\lambda_j \quad (2 \leq j \leq q)$, and

$$M_{tij} = E_{\theta,\lambda}\{D_t D_i D_j \log f(X_1; \theta, \lambda)\}\ , \quad 1 \leq t, i, j \leq q\ .$$

Then the formula for expected length will again be given by (3.9), with a_{20} there replaced by I_{11} and

$$V(\theta, \lambda) = 2D_1(I_{11}^{-1}\gamma_1) - I_{11}^{-1}\gamma_1 \sum_{i,j=2}^{q}\sum I^{ij} M_{1ij} + I_{11}^{-1}\gamma_1^2\ ,$$

γ_1 being as before given by (2.8).

Acknowledgement Thanks are due to the Editor, Associate Editor and a referee for very constructive suggestions. The work was partially supported by the Natural Sciences and Engineering Research Council of Canada, and that of RM was also supported by a grant from the Centre of Management and Development Studies, Indian Institute of Management Calcutta.

References

Amari, S. (1985). *Differential Geometrical Methods in Statistics.* Springer Verlag, Berlin.

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

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

DiCiccio, T.J. and Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. Roy. Statist. Soc. B* **56**, 397-408.

Ghosh, J.K. and Mukerjee, R. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. Roy. Statist. Soc. B* **54**, 867-875.

Ghosh, J.K. and Mukerjee, R. (1994). Adjusted versus conditional likelihood: power properties and Bartlett-type adjustment. *J. Roy. Statist. Soc. B* **56**, 185-188.

Ghosh, J.K. and Mukerjee, R. (1995). Frequentist validity of highest posterior density regions in the presence of nuisance parameters. *Statistics & Decisions* **13**, 131-139.

Kumon, M. and Amari, S. (1983). Geometrical theory of higher order asymptotics of test, interval estimator and conditional inference. *Proc. Roy. Soc. London A* **387**, 429-458.

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

Mukerjee, R. (1992). Conditional likelihood and power: higher order asymptotics. *Proc. Roy. Soc. London A* **438**, 433-446.

Mukerjee, R. and Reid, N. (1998). Comparison of test statistics via expected lengths of associated confidence intervals. *J. Statist. Plann. Inf.* (Special issue on score statistic), to appear.

List of symbols

B : adjustment to profile likelihood
C : confidence limit
D : derivative operator
E : expectation
F : Bartlett correction factor
H : as in (2.1)
I : information matrix (in Section 5)
K : as in (2.1)
L : likelihood
M : as in Section 5
O : order
P : probability
Q : as in (3.2)
R : real space
T : LR statistic
U : as in (3.14)
V : expected length of confidence interval
X : observation

a : in the information matrix
b : scalar, as in (4.2)
f : density
g : as in (3.1)
h : $h = \sqrt{n}(\theta - \tilde{\theta})$
i : dummy
j : dummy
ℓ : likelihood, per observation
m : nuisance parameter
n : sample size
o : order
p : in O_p or o_p
q : dimension of nuisance parameter, in (Section 5)
s : dummy
t : dummy
u : as in (3.2)
w : as in (3.2)
x : in $f(x; \theta, \lambda)$
z : normal percentile

α : in confidence coefficient
γ : as in (2.8)
λ : length of confidence interval
θ : parameter of interest
μ : exponential mean, in Example 3
ψ : as in Examples 1 and 4
π : prior density