Bayesian and Frequentist Bartlett Corrections for Likelihood Ratio and Conditional Likelihood Ratio Tests

By J. K. GHOSH† and RAHUL MUKERJEE

Indian Statistical Institute, Calcutta, India, and Purdue University, West Lafayette, USA

Indian Institute of Management and Indian Statistical Institute, Calcutta, India

SUMMARY

This paper characterizes priors under which the Bayesian and frequentist Bartlett corrections for the likelihood ratio and the conditional likelihood ratio (CLR) statistics differ by $o(1)$. It is seen that, except for sample points with negligible probability, the CLR statistic has a posterior distribution for which a posterior Bartlett correction exists. This observation leads to an alternative proof of the existence of a frequentist Bartlett correction for the CLR statistic.

Keywords: BARTLETT CORRECTION; CONDITIONAL LIKELIHOOD RATIO TEST; LIKELIHOOD RATIO TEST; NON-INFORMATIVE PRIOR; PARAMETRIC ORTHOGONALITY

1. INTRODUCTION

In a recent paper, Bickel and Ghosh (1990) observed that, except for sample points with negligible probability of order $O(n^{-2})$, the posterior distribution of the likelihood ratio (LR) statistic has a structure such that there is a posterior Bartlett correction which makes the posterior distribution approximable by a χ^2-distribution up to $O(n^{-2})$. They provided a justification for the frequentist Bartlett correction based on this interesting observation. They also posed an open problem relating to the characterization of priors under which the Bayesian and frequentist Bartlett corrections for the LR statistic differ by $o(1)$. For such priors, posterior probability regions, based on a posterior Bartlett corrected LR statistic, are also frequentist regions with error $o(n^{-1})$. In a sense, priors of this kind may be regarded as non-informative priors which, as noted in Tibshirani (1989), can be helpful for comparisons in a Bayesian analysis.

The present work attempts to settle this problem. We also consider a similar problem regarding another important new statistic, namely the conditional likelihood ratio (CLR) statistic introduced by Cox and Reid (1987). The results on the CLR statistic have been derived via those on the LR statistic. In the process, it is seen that, except for sample points with probability of order $O(n^{-2})$, the CLR statistic has a posterior distribution such that a posterior Bartlett correction exists so that results similar to those in Bickel and Ghosh (1990) hold also with the CLR statistic. With reference to a problem posed in Cox and Reid (1987), it is also indicated that this leads to an alternative and possibly simpler proof of the existence of a frequentist Bartlett correction for the CLR statistic (see Mukerjee and Chandra (1991)).

In this paper, primarily for notational simplicity, we consider the situation where both θ_1, the parameter of interest, and θ_2, the nuisance parameter, are one dimensional. The discussion can be extended to multidimensional θ_2 with additional algebra. However, as noted in Cox and Reid (1987), the assumption that θ_1 is one dimensional is non-trivial. In particular, if θ_1 and θ_2 are both multidimensional then we cannot in general employ global parametric orthogonality as noted in Cox and Reid (1987). In this paper, we consider models with a nuisance parameter since one of our main objectives is to include the CLR test in the discussion. It may also be of interest to characterize, in the absence of any nuisance parameter, priors under which the Bayesian and frequentist Bartlett corrections for the LR statistic differ by $o(1)$. This problem will be considered elsewhere.

2. RESULTS ON LIKELIHOOD RATIO STATISTIC

Let $\{X_i\}$, $i \geqslant 1$, be a sequence of independent and identically distributed random variables with common density $f(x; \theta)$, where $\theta = (\theta_1, \theta_2)'$ and, as stated above, θ_1 is the one-dimensional parameter of interest and θ_2 is the one-dimensional nuisance parameter. For $i, j, i', j' = 0, 1, 2, \ldots$, let

$$K_{ij} = E_\theta [\partial^{i+j}\{\log f(X_1; \theta)\}/\partial\theta_1^i \, \partial\theta_2^j],$$

$$K_{ij,i'j'} = E_\theta([\partial^{i+j}\{\log f(X_1; \theta)\}/\partial\theta_1^i \partial\theta_2^j][\partial^{i'+j'}\{\log f(X_1; \theta)\}/\partial\theta_1^{i'} \partial\theta_2^{j'}]).$$

$K_{ij,i'j',i''j''}$ etc. are defined similarly. K_{ij}, $K_{ij,i'j'}$ etc. are functions of θ. Let $a_{20} = -K_{20}$, $a_{02} = -K_{02}$. Since θ_1 and θ_2 are one dimensional, we assume global parametric orthogonality (Cox and Reid, 1987), i.e. $K_{10,01} \equiv 0$, identically in θ. Then under standard regularity conditions the per observation information matrix is given by $\text{diag}(a_{20}, a_{02})$ which is assumed to be positive definite for each θ.

Let $L(\theta) = L(\theta_1, \theta_2) = \Sigma_{i=1}^n \log f(X_i; \theta)$, $l(\theta) = n^{-1} L(\theta)$, where n is the sample size. Denoting the maximum likelihood estimator (MLE) of θ by $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$, for $i, j = 0, 1, 2, \ldots$, let $l_{ij}(\theta) = \partial^{i+j} l(\theta)/\partial\theta_1^i \partial\theta_2^j$, $b_{ij} = l_{ij}(\hat{\theta})$, $c_{ij} = -b_{ij}$ and

$$C = \begin{pmatrix} c_{20} & c_{11} \\ c_{11} & c_{02} \end{pmatrix}.$$

Recall that the LR statistic is given by

$$\lambda = -2[L\{\theta_1, \hat{\theta}_2(\theta_1)\} - L(\hat{\theta}_1, \hat{\theta}_2)], \tag{2.1}$$

where $\hat{\theta}_2(\theta_1)$ is the MLE of θ_2 given θ_1.

We make the assumptions in Johnson (1970). Let θ have a prior density $\pi(\ )$ which, as in Johnson (1970), is positive and thrice continuously differentiable for all θ. When $\pi(\ )$ is not proper, as assumed by Johnson (1970), we shall require that there is an n_0 (> 0) such that for all $X_1, \ldots, X_{n_0}$ the posterior of θ given $X_1, \ldots, X_{n_0}$ is proper. Then Johnson's (1970) proof holds. Let P be the joint probability measure of θ and $X = (X_1, \ldots, X_n)'$. All formal expansions for the posterior, as used here, are valid for sample points in a set S which may be defined along the lines of Bickel and Ghosh (1990), section 2, with $m = 3$, with P_θ-probability $1 + O(n^{-2})$ uniformly on compact sets of θ. The matrix C is positive definite over S. We also make Edgeworth assumptions as in Bickel and Ghosh (1990), p. 1078.

With a prior density $\pi(\ )$ for θ, we first explicitly calculate the Bayesian Bartlett correction for the posterior distribution of the LR statistic λ. For this, note that the posterior density of $h=(h_1, h_2)' = n^{1/2}(\theta-\hat{\theta})$ is given by

$$\pi(h|X)=\pi(\hat{\theta}+n^{-1/2}h)\exp\{L(\hat{\theta}+n^{-1/2}h)-L(\hat{\theta})\}/N(X), \tag{2.2}$$

where

$$N(X)=\int\pi(\hat{\theta}+n^{-1/2}h)\exp\{L(\hat{\theta}+n^{-1/2}h)-L(\hat{\theta})\}\,\mathrm{d}h. \tag{2.3}$$

Let $\hat{\pi} = \pi(\hat{\theta})$ and for $i,j = 0, 1, 2, \ldots, \pi_{ij}(\theta) = \partial^{i+j}\pi(\theta)/\partial\theta_1^i\partial\theta_2^j$, $\hat{\pi}_{ij} = \pi_{ij}(\hat{\theta})$. Also, let

$$\psi_i(h)= \sum_{j=0}^{i}\binom{i}{j}h_1^{i-j}h_2^j b_{i-jj} \qquad (i=3, 4).$$

Then using Taylor's expansion and observing that $b_{10}=b_{01}=0$ by the definition of $\hat{\theta}$, we obtain

$$\begin{aligned}\pi(\hat{\theta}+n^{-1/2}h)&\exp\{L(\hat{\theta}+n^{-1/2}h)-L(\hat{\theta})\}\\ =&\left[\hat{\pi}+n^{-1/2}\left\{\frac{1}{6}\hat{\pi}\,\psi_3(h)+h_1\hat{\pi}_{10}+h_2\hat{\pi}_{01}\right\}+n^{-1}\left\{\frac{1}{72}\hat{\pi}\,\psi_3(h)^2+\frac{1}{24}\hat{\pi}\,\psi_4(h)\right.\right.\\ &\left.\left.+\frac{1}{6}(h_1\hat{\pi}_{10}+h_2\hat{\pi}_{01})\,\psi_3(h)+\frac{1}{2}(h_1^2\hat{\pi}_{20}+2h_1h_2\hat{\pi}_{11}+h_2^2\hat{\pi}_{02})\right\}\right]\\ &\times\exp\left[-\frac{1}{2}\left\{h_1^2D+c_{02}(h_2+c_{11}c_{02}^{-1}h_1)^2\right\}\right]+o(n^{-1}),\end{aligned} \tag{2.4}$$

where $D=c_{20}-c_{02}^{-1}c_{11}^2$ which is positive over S.

Since, by parametric orthogonality, $c_{11}=o(1)$, integrating h_2 out in equation (2.4), it follows from equations (2.2)–(2.4) after considerable algebra that the posterior density of h_1 is given by

$$\begin{aligned}\pi(h_1|X) = \phi(h_1; D^{-1})[1 + n^{-1/2}Q_1^*(h_1) + n^{-1}\{Q_2^*(h_1) - H(X)\\ + R_1(X, h_1)\}] + o(n^{-1}),\end{aligned} \tag{2.5}$$

where $\phi(\ ;D^{-1})$ is the univariate normal density with mean 0 and variance D^{-1},

$$\begin{aligned}Q_1^*(h_1) = \frac{1}{6}h_1^3b_{30}+h_1\left(\frac{1}{2}b_{12}c_{02}^{-1}+\hat{\pi}_{10}\hat{\pi}^{-1}\right)+\left[h_1^3\left(\frac{1}{2}b_{12}c_{11}^2c_{02}^{-2}-\frac{1}{2}b_{21}c_{11}c_{02}^{-1}\right.\right.\\ \left.\left.-\frac{1}{6}b_{03}c_{11}^3c_{02}^{-3}\right)-h_1\left(\hat{\pi}_{01}\hat{\pi}^{-1}c_{11}c_{02}^{-1}+\frac{1}{2}b_{03}c_{11}c_{02}^{-2}\right)\right],\end{aligned} \tag{2.6}$$

$$\begin{aligned}Q_2^*(h_1) =& \frac{1}{72}\{h_1^6b_{30}^2+h_1^4c_{02}^{-1}(6b_{30}b_{12}+9b_{21}^2)+h_1^2c_{02}^{-2}(18b_{03}b_{21}+27b_{12}^2)\\ &+15c_{02}^{-3}b_{03}^2\}+\frac{1}{24}\{h_1^4b_{40}+6h_1^2c_{02}^{-1}b_{22}+3c_{02}^{-2}b_{04}\}\\ &+\frac{1}{2}\hat{\pi}^{-1}(h_1^2\hat{\pi}_{20}+\hat{\pi}_{02}c_{02}^{-1})+\hat{\pi}^{-1}\left\{\frac{1}{6}h_1^4\hat{\pi}_{10}b_{30}+\frac{1}{2}h_1^2c_{02}^{-1}(\hat{\pi}_{10}b_{12}+\hat{\pi}_{01}b_{21})\right.\\ &\left.+\frac{1}{2}\hat{\pi}_{01}c_{02}^{-2}b_{03}\right\},\end{aligned} \tag{2.7}$$

$$H(X)=\int_{-\infty}^{\infty}Q_2^*(h_1)\,\phi(h_1;D^{-1})\,\mathrm{d}h_1, \tag{2.8}$$

and, analogously to the term within squared brackets in equation (2.6), $R_1(X, h_1)$ is a polynomial in h_1 with each coefficient a function of X of order $o(1)$. Note that equation (2.5) agrees with the findings of Tierney and Kadane (1986). For analytical studies like this, the use of equation (2.5) seems more convenient than that of the formula in section 4 of Tierney and Kadane (1986), whereas for numerical approximations it should be the other way round.

Next observe that, by the definition of $\hat{\theta}_2(\theta_1)$,

$$0 = n^{1/2} l_{01}\{\theta_1, \hat{\theta}_2(\theta_1)\}. \tag{2.9}$$

Using a Taylor's expansion for the right-hand side of equation (2.9), it can be seen (see McCullagh (1987)) that

$$\hat{\theta}_2(\theta_1) = \hat{\theta}_2 - n^{-1/2} h_1 c_{02}^{-1} c_{11} + n^{-1} h_1^2 \left(\frac{1}{2} b_{21} c_{02}^{-1} - b_{12} c_{11} c_{02}^{-2} + \frac{1}{2} b_{03} c_{11} c_{02}^{-3}\right) + o(n^{-1}), \tag{2.10}$$

whence by equation (2.1) we obtain

$$\lambda = w^2 + o(n^{-1}), \tag{2.11a}$$

where

$$w = h_1 D^{1/2} \left[1 + \frac{1}{6} n^{-1/2} h_1 D^{-1} (b_{03} c_{11}^3 c_{02}^{-3} - 3 b_{12} c_{11}^2 c_{02}^{-2} + 3 b_{21} c_{11} c_{02}^{-1} - b_{30}) + n^{-1} \left\{R_2(X, h_1) - \frac{1}{8} h_1^2 D^{-1} \left(\frac{1}{3} b_{40} + c_{02}^{-1} b_{21}^2 + \frac{1}{9} D^{-1} b_{30}^2\right)\right\}\right], \tag{2.11b}$$

and, like $R_1(X, h_1)$, $R_2(X, h_1)$ is a polynomial in h_1 with each coefficient a function of X of order $o(1)$. By equation (2.11b),

$$h_1 = w D^{-1/2} - \frac{1}{6} n^{-1/2} w^2 D^{-2} (b_{03} c_{11}^3 c_{02}^{-3} - 3 b_{12} c_{11}^2 c_{02}^{-2} + 3 b_{21} c_{11} c_{02}^{-1} - b_{30}) + n^{-1} \left\{\frac{1}{8} w^3 D^{-5/2} \left(\frac{1}{3} b_{40} + \frac{5}{9} D^{-1} b_{30}^2 + c_{02}^{-1} b_{21}^2\right) + T_1(X, w)\right\} + o(n^{-1}), \tag{2.12}$$

where $T_1(X, w)$ is a polynomial in w with each coefficient a function of X of order $o(1)$. Hence by equations (2.5)–(2.8), it follows after some algebra that the posterior density of w is given by

$$\pi(w|X) = \phi(w; 1)[1 + n^{-1/2} G_1 w + n^{-1}\{G_2^*(w^2 - 1) + T_2(X, w)\}] + o(n^{-1}), \tag{2.13}$$

where

$$G_1 = \frac{1}{3} D^{-3/2} (b_{30} - 3 b_{21} c_{11} c_{02}^{-1} + 3 b_{12} c_{11}^2 c_{02}^{-2} - b_{03} c_{11}^3 c_{02}^{-3}) - D^{-1/2} \left(\hat{\pi}_{01} \hat{\pi}^{-1} c_{11} c_{02}^{-1} - \hat{\pi}_{10} \hat{\pi}^{-1} - \frac{1}{2} b_{12} c_{02}^{-1} + \frac{1}{2} b_{03} c_{11} c_{02}^{-2}\right), \tag{2.14}$$

$$G_2^* = \frac{1}{8} D^{-2} \left(b_{40} + \frac{5}{3} D^{-1} b_{30}^2 + 3 c_{02}^{-1} b_{21}^2 + 2 c_{02}^{-1} b_{12} b_{30}\right) + \frac{1}{8} (D c_{02}^2)^{-1} (2 b_{03} b_{21} + 3 b_{12}^2) + \frac{1}{4} (D c_{02})^{-1} b_{22} + \frac{1}{2} (D \hat{\pi})^{-1} \{\hat{\pi}_{20} + c_{02}^{-1} (\hat{\pi}_{10} b_{12} + \hat{\pi}_{01} b_{21}) + D^{-1} \hat{\pi}_{10} b_{30}\}, \tag{2.15}$$

and $T_2(X, w)$ is a polynomial in w with each coefficient a function of X of order $o(1)$. It can be seen that $T_2(X, w)$ does not involve any odd power of w.

Comparing equation (2.13) with theorem 1 in Bickel and Ghosh (1990), we obtain

$$\pi(w|X) = \phi(w; 1)\{1 + n^{-1/2}G_1 w + n^{-1}G_2(w^2 - 1)\} + o(n^{-1}), \tag{2.16}$$

where

$$G_2 = G_2^* + o(1). \tag{2.17}$$

It should be noted that Bickel and Ghosh (1990) showed that $\pi(w|X)$ is of the form (2.16) without specifying G_1 and G_2 explicitly. Writing

$$\begin{aligned} B &= 2G_2, \\ \lambda^* &= \lambda/(1 + n^{-1}B), \end{aligned} \tag{2.18}$$

it is seen directly from equations (2.11) and (2.16) or by using the lemmas in the appendix of Bickel and Ghosh (1990) that $E(\lambda|X) = 1 + n^{-1}B + o(n^{-1})$, and that for each z ($\geqslant 0$)

$$P(\lambda^* \leqslant z|X) = \gamma(z) + o(n^{-1}), \tag{2.19}$$

where $\gamma(\;)$ is the cumulative distribution function of the χ^2-distribution with 1 degree of freedom. In fact, following Bickel and Ghosh (1990), the remainder in equation (2.19) is of order $O(n^{-2})$; see Barndorff-Nielsen and Hall (1988) and Chandra and Ghosh (1979) for the corresponding result in the frequentist approach. Thus $1 + n^{-1}B$ is the Bartlett adjustment factor corresponding to the posterior distribution of λ. Relations (2.15), (2.17) and (2.18) specify the Bartlett adjustment factor explicitly. A consideration of the square root version, w, of λ was useful in this explicit derivation —a technique which is helpful in other contexts as well; see DiCiccio *et al.* (1990), Mukerjee and Chandra (1991) and the references therein.

Since $b_{ij} = K_{ij} + o(1)$ for each i, j, using parametric orthogonality, it follows from equations (2.15), (2.17) and (2.18) that

$$B = \tilde{B} + o(1), \tag{2.20a}$$

where

$$\begin{aligned} \tilde{B} = {} & \frac{1}{4}a_{20}^{-2}K_{40} + \frac{5}{12}a_{20}^{-3}K_{30}^2 + \frac{3}{4}(a_{20}^2 a_{02})^{-1}K_{21}^2 + \frac{1}{4}(a_{20}a_{02}^2)^{-1}(2K_{03}K_{21} + 3K_{12}^2) \\ & + \pi(\theta)^{-1}[a_{20}^{-1}\,\pi_{20}(\theta) + (a_{20}a_{02})^{-1}\{\pi_{10}(\theta)K_{12} + \pi_{01}(\theta)K_{21}\} + a_{20}^{-2}K_{30}\,\pi_{10}(\theta)] \\ & + \frac{1}{2}(a_{20}a_{02})^{-1}K_{22} + \frac{1}{2}(a_{20}^2 a_{02})^{-1}K_{12}K_{30}. \end{aligned} \tag{2.20b}$$

However, as noted in Mukerjee and Chandra (1991)—see also Barndorff-Nielsen and Blæsild (1986)—the Bartlett adjusted statistic corresponding to λ in the frequentist set-up is

$$\lambda^{**} = \lambda/(1 + n^{-1}F), \tag{2.21}$$

where

$$F = \tilde{F} + o(1), \tag{2.22a}$$

with

$$\tilde{F} = a_{20}^{-2}\left(\frac{1}{4}K_{40} + K_{20,20} + K_{10,30} + K_{10,10,20}\right)$$
$$+ a_{20}^{-3}\left(\frac{7}{4}K_{10,20}^2 + \frac{11}{6}K_{30}K_{10,20} + \frac{7}{18}K_{30}^2 + \frac{1}{36}K_{10,10,10}^2\right)$$
$$+ (a_{20}a_{02})^{-1}\left(2K_{11,11} + K_{01,21} + 2K_{10,01,11} + K_{10,12} + \frac{1}{2}K_{22}\right)$$
$$- (a_{20}a_{02}^2)^{-1}\left(\frac{1}{4}K_{12}^2 + K_{21}K_{01,02} + \frac{1}{2}K_{21}K_{03} + K_{12}K_{10,02}\right)$$
$$- (a_{20}^2a_{02})^{-1}\left(\frac{1}{4}K_{21}^2 + \frac{1}{3}K_{30}K_{12} + \frac{1}{2}K_{12}K_{10,20} + K_{21}K_{20,01} - \frac{1}{6}K_{12}K_{10,10,10}\right). \quad (2.22b)$$

We now characterize priors for which

$$B - F = o(1). \quad (2.23)$$

For such priors, by equations (2.18) and (2.21), $\lambda^* - \lambda^{**} = o(n^{-1})$, under θ, and it is easy to see that this implies $P_\theta(\lambda^* > \zeta_0) = P_\theta(\lambda^{**} > \zeta_0) + o(n^{-1}) = \alpha + o(n^{-1})$, where ζ_0 is the upper α-point of a χ^2-distribution with 1 degree of freedom. Hence, for priors satisfying equation (2.23), inversion of λ^* leads to a confidence set, namely $\{\theta_1: \lambda^* = \lambda^*(X, \theta_1) \leqslant \zeta_0\}$, with confidence coefficient $1-\alpha$, which has both posterior and frequentist validity up to $o(n^{-1})$ (see Tibshirani (1989), Stein (1985) and Welch and Peers (1963)).

Using regularity conditions like

$$\frac{\partial K_{ij}}{\partial\theta_1} = K_{ij,10} + K_{i+1j}, \qquad \frac{\partial K_{ij,i'j'}}{\partial\theta_1} = K_{ij,i'j',10} + K_{i+1j,i'j'} + K_{ij,i'+1j'},$$
$$\frac{\partial K_{ij}}{\partial\theta_2} = K_{ij,01} + K_{ij+1}, \qquad \frac{\partial K_{ij,i'j'}}{\partial\theta_2} = K_{ij,i'j',01} + K_{ij+1,i'j'} + K_{ij,i'j'+1},$$
$$K_{30} + 3K_{10,20} + K_{10,10,10} = 0,$$
$$K_{11,10} + K_{21} = 0, \qquad K_{12,10} + K_{11,11} + K_{11,10,01} + K_{21,01} + K_{22} = 0,$$

some of which are consequences of parametric orthogonality, it follows after some simplification from equations (2.20b) and (2.22b) that

$$\tilde{B} - \tilde{F} = \pi(\theta)^{-1}\left[\frac{\partial}{\partial\theta_1}\left\{\frac{\pi_{10}(\theta)}{a_{20}} - \frac{K_{10,20}\,\pi(\theta)}{a_{20}^2} + \frac{K_{12}\,\pi(\theta)}{a_{20}a_{02}}\right\} + \frac{\partial}{\partial\theta_2}\left\{\frac{K_{21}\,\pi(\theta)}{a_{20}a_{02}}\right\}\right]. \quad (2.24)$$

Hence by equations (2.20a) and (2.22a), $B - F = o(1)$ holds if and only if $\pi(\theta)$ satisfies the differential equation

$$\frac{\partial}{\partial\theta_1}\left\{\frac{\pi_{10}(\theta)}{a_{20}} - \frac{K_{10,20}\,\pi(\theta)}{a_{20}^2} + \frac{K_{12}\,\pi(\theta)}{a_{20}a_{02}}\right\} + \frac{\partial}{\partial\theta_2}\,\frac{K_{21}\,\pi(\theta)}{a_{20}a_{02}} = 0. \quad (2.25)$$

An illustrative example will be presented in the next section.

3. RESULTS ON CONDITIONAL LIKELIHOOD RATIO STATISTIC

Following Cox and Reid (1987), the CLR statistic is given by

$$\lambda_c = 2\{\rho(\theta_1^*) - \rho(\theta_1)\}, \tag{3.1}$$

where $\rho(\theta_1) = L\{\theta_1, \hat{\theta}_2(\theta_1)\} - \frac{1}{2}\log[-n\ l_{02}\{\theta_1, \hat{\theta}_2(\theta_1)\}]$, and $\rho(\theta_1^*) = \sup_{\theta_1}\{\rho(\theta_1)\}$. Proceeding along the lines of Mukerjee and Chandra (1991) and using parametric orthogonality, from equations (2.1), (2.11a), (2.12) and (3.1), we can show that

$$\lambda_c = w^2 - n^{-1/2}\beta_1(X)w + n^{-1}\{\tfrac{1}{4}\beta_1^2(X) + \beta_2(X)w^2\} + o(n^{-1}) = w_c^2 + o(n^{-1}), \tag{3.2}$$

where

$$w_c = w - \tfrac{1}{2}n^{-1/2}\beta_1(X) + \tfrac{1}{2}n^{-1}w\,\beta_2(X), \tag{3.3}$$

$$\beta_1(X) = D^{-1/2}(c_{02}^{-1}b_{12} - c_{11}c_{02}^{-2}b_{03}), \tag{3.4}$$

$$\beta_2(X) = -\tfrac{1}{2}c_{20}^{-1}\{c_{02}^{-2}(b_{12}^2 + b_{21}b_{03}) + c_{02}^{-1}b_{22}\} - \tfrac{1}{6}(c_{20}^2c_{02})^{-1}b_{30}b_{12} + o(1). \tag{3.5}$$

By equation (3.3),

$$w = w_c\{1 - \tfrac{1}{2}n^{-1}\beta_2(X)\} + \tfrac{1}{2}n^{-1/2}\beta_1(X) + o(n^{-1}),$$

and hence by equation (2.16) the posterior density of w_c is given by

$$\pi(w_c|X) = \phi(w_c; 1)\{1 + n^{-1/2}G_{1c}w_c + n^{-1}G_{2c}(w_c^2 - 1)\} + o(n^{-1}), \tag{3.6}$$

where, with G_1 and G_2 as in equations (2.14) and (2.17),

$$\begin{aligned} G_{1c} &= G_1 - \tfrac{1}{2}\beta_1(X), \\ G_{2c} &= G_2 + \tfrac{1}{2}\beta_2(X) + \tfrac{1}{8}\beta_1^2(X) - \tfrac{1}{2}G_1\beta_1(X). \end{aligned} \tag{3.7}$$

As in Section 2, by equations (3.2) and (3.6), the posterior distribution of λ_c is such that a posterior Bartlett correction exists and the posterior Bartlett corrected statistic corresponding to λ_c is $\lambda_c^* = \lambda_c/(1 + n^{-1}B_c)$, where by equations (2.18) and (3.7)

$$B_c = 2G_{2c} = B + \beta_2(X) + \tfrac{1}{4}\beta_1^2(X) - G_1\beta_1(X). \tag{3.8}$$

From this we can further deduce, exactly as in Bickel and Ghosh (1990) (see the proof of their theorem 3) the existence of a frequentist Bartlett adjustment for λ_c and this settles a problem posed in Cox and Reid (1987). Mukerjee and Chandra (1991) provided another proof of the existence of a frequentist Bartlett correction for λ_c. Their proof appears to be more complicated but their results are more detailed and include an explicit expression for the frequentist Bartlett correction (see equation (3.10) below).

We now characterize priors for which the Bayesian and frequentist Bartlett corrections for the CLR statistic differ by $o(1)$. As in the context of Section 2, with such priors, inversion of λ_c^* leads to a confidence set which has both posterior and frequentist validity up to $o(n^{-1})$. Since $b_{ij} = K_{ij} + o(1)$, using parametric orthogonality, it follows from equations (2.14), (2.17), (2.20a), (3.4), (3.5) and (3.8) that

$$\begin{aligned} B_c = \tilde{B} &- (a_{20}a_{02}^2)^{-1}\{\tfrac{3}{4}K_{12}^2 + \tfrac{1}{2}K_{21}K_{03}\} - \tfrac{1}{2}(a_{20}^2a_{02})^{-1}K_{30}K_{12} \\ &- (a_{20}a_{02})^{-1}\{\tfrac{1}{2}K_{22} + \pi(\theta)^{-1}\pi_{10}(\theta)K_{12}\} + o(1), \end{aligned} \tag{3.9}$$

where $\tilde{B}$ is as in equation (2.20b). Also, with the results in Mukerjee and Chandra (1991), the Bartlett adjusted statistic corresponding to λ_c in the frequentist set-up is $\lambda_c^{**} = \lambda_c/(1 + n^{-1}F_c)$, where

$$F_c = \tilde{F} + (a_{20}a_{02})^{-1}(K_{10.12} + \tfrac{1}{2}K_{22}) + (a_{20}a_{02}^2)^{-1}(K_{12}K_{10.02} - \tfrac{1}{2}K_{21}K_{03} + \tfrac{1}{4}K_{12}^2) + (a_{20}^2 a_{02})^{-1}K_{12}(\tfrac{1}{2}K_{10.20} + \tfrac{1}{3}K_{30} - \tfrac{1}{6}K_{10.10.10}) + o(1), \tag{3.10}$$

with $\tilde{F}$ as in equation (2.22b). With the regularity conditions stated in Section 2, it follows from equations (2.24), (3.9) and (3.10) that $B_c - F_c = o(1)$ if and only if $\pi(\theta)$ satisfies the differential equation

$$\frac{\partial}{\partial\theta_1}\left\{\frac{\pi_{10}(\theta)}{a_{20}} - \frac{K_{10.20}\,\pi(\theta)}{a_{20}^2}\right\} + \frac{\partial}{\partial\theta_2}\left\{\frac{K_{21}\,\pi(\theta)}{a_{20}a_{02}}\right\} = 0. \tag{3.11}$$

3.1. *Examples*

Let $f(x;\theta)$ represent the univariate normal density with mean θ_2 and variance θ_1, the parameter space being $\Theta = \{\theta: \theta_1 > 0, -\infty < \theta_2 < \infty\}$. It can be seen that under this parameterization parametric orthogonality holds. Also, $a_{20} = \frac{1}{2}\theta_1^{-2}$, $a_{02} = \theta_1^{-1}$, $K_{10.20} = -\theta_1^{-3}$, $K_{21} = 0$ and $K_{12} = \theta_1^{-2}$. Hence, solving equation (2.25), $B - F = o(1)$ if and only if $\pi(\theta)$ is of the form

$$\pi(\theta) = d_1(\theta_2)\theta_1^{-1} + d_2(\theta_2)\theta_1^{-3}, \tag{3.12}$$

where $d_1(\theta_2)$ and $d_2(\theta_2)$ are non-negative functions of θ_2 such that at least one of them is positive. Similarly, solving equation (3.11), $B_c - F_c = o(1)$ if and only if $\pi(\theta)$ is of the form

$$\pi(\theta) = d_3(\theta_2)\theta_1^{-1} + d_4(\theta_2)\theta_1^{-2}, \tag{3.13}$$

where $d_3(\theta_2)$ and $d_4(\theta_2)$ are non-negative functions of θ_2, at least one of them being positive. By equations (3.12) and (3.13), the relations $B - F = o(1)$ and $B_c - F_c = o(1)$ hold simultaneously if and only if $\pi(\theta) = d(\theta_2)\theta_1^{-1}$, for some function $d(\theta_2)$ (> 0) of θ_2.

Recently, Tibshirani (1989) showed (see also Peers (1965)) that priors of the form $\pi_0(\theta) = d(\theta_2)a_{20}^{1/2}$ ensure, up to $o(n^{-1/2})$, the frequentist validity of posterior quantiles of θ_1. It is interesting that in many situations of importance $\pi_0(\theta)$, with suitably chosen $d(\theta_2)$, satisfies both equation (2.25) and equation (3.11). This happens, for example, in the above example and also in the context of

(a) a univariate normal model where interest lies in the population mean, the population variance being the nuisance parameter,
(b) the exponential regression model of Cox and Reid (1987) where the regression coefficient is the parameter of interest and
(c) a model relating to two independent univariate normal populations with unknown means and known variances, the ratio of means being the parameter of interest (Cox and Reid, 1987).

However, there are also models where $\pi_0(\theta)$ satisfies neither equation (2.25) nor equation (3.11). This happens, for example, in a bivariate normal model with correlation coefficient θ_1, means each θ_2 and variances each 1, where other solutions to equations (2.25) and (3.11) are not difficult to obtain.

As this discussion indicates, for the models usually arising in practice, solutions to equations (2.25) and (3.11) are readily available and often, though not always, $\pi_0(\theta)$ is a solution. This has been checked with many other examples that are not reported here. Equations (2.25) and (3.11) become identical if $K_{12} = 0$, something that arises in situations (a)–(c) above. As a referee has pointed out, we could attempt to compare λ

and λ_c on the basis of qualitative comparisons between equations (2.25) and (3.11). This is an interesting problem to which we do not know the answer at this stage.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for very constructive suggestions. The work of the second author was supported by a grant from the Centre for Management and Development Studies, Indian Institute of Management, Calcutta.

REFERENCES

Barndorff-Nielsen, O. E. and Blæsild, P. (1986) A note on the calculation of Bartlett adjustments. *J. R. Statist. Soc.* B, **48**, 353–358.

Barndorff-Nielsen, O. E. and Hall, P. (1988) On the level error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374–378.

Bickel, P. J. and Ghosh, J. K. (1990) A decomposition for the likelihood ratio statistic and the Bartlett correction—a Bayesian argument. *Ann. Statist.*, **18**, 1070–1090.

Chandra, T. K. and Ghosh, J. K. (1979) Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhya* A, **41**, 22–47.

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

DiCiccio, T. J., Field, C. A. and Fraser, D. A. S. (1990) Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77–95.

Johnson, R. A. (1970) Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.*, **41**, 851–864.

McCullagh, P. (1987) *Tensor Methods in Statistics*, ch. 7. London: Chapman and Hall.

Mukerjee, R. and Chandra, T. K. (1991) Bartlett-type adjustment for the conditional likelihood ratio statistic of Cox and Reid. *Biometrika*, **78**, 365–372.

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.

Stein, C. (1985) On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach Centre Publications, vol. 16, pp. 485–514. Warsaw: PWN.

Tibshirani, R. (1989) Noninformative priors for one parameter of many. *Biometrika*, **76**, 604–608.

Tierney, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Ass.*, **81** 82–86.

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