Statistica Sinica **5**(1995), 781-791

ON POSTERIOR CREDIBLE SETS BASED ON THE SCORE STATISTIC

C. Radhakrishna Rao and Rahul Mukerjee

Pennsylvania State University and Indian Institute of Management

Abstract. Conditions for approximate frequentist validity of posterior credible sets based on the score statistic have been derived in the multiparameter case. These conditions can be helpful in supplementing similar conditions obtainable through the likelihood ratio statistic and the highest posterior density region. In the process, explicit expressions are given for the posterior quantiles of the score statistic. Similar results based on Wald's statistic have also been briefly indicated.

Key words and phrases: Jeffreys' prior, non-informative prior, posterior quantile, Rao's score statistic, Wald's statistic.

1. Introduction

In recent years there has been a renewed interest in the study of approximate frequentist validity of posterior credible sets. As noted in Tibshirani (1989), apart from providing a method for constructing accurate frequentist confidence regions, such a study is also helpful in defining a non-informative prior which could be potentially useful for comparative purposes in Bayesian analysis. In other words, posterior credible sets with approximate frequentist validity are meaningful not only from a Bayesian but also from a purely frequentist point of view where no prior is assumed and, as such, priors underlying such credible sets may, in a sense, be looked upon as non-informative. Even if, for inferential purposes in a given context, a Bayesian wishes to use a subjective rather than such a non-informative prior, the latter may help in judging how subjective the former is. We refer to Ghosh and Mukerjee (1992a) for a discussion on this and other approaches for defining non-informative priors.

Welch and Peers (1963) considered the above problem, with reference to one-sided credible sets, in the one-parameter case. Their work was strengthened and extended in various directions by Peers (1965, 1968), Stein (1985), Tibshirani (1989) and Nicolau (1993). Tibshirani (1989) showed that, with a one-dimensional interest parameter, elegant results are available via an orthogonal parametrization (Cox and Reid (1987)). In the multiparameter case, Lee (1989) explored the frequentist validity of elliptic credible regions and half spaces while Ghosh and Mukerjee (1991, 1993) considered posterior regions based on the posterior Bartlett-corrected likelihood ratio (LR) statistic (see Bickel and Ghosh (1990)) and highest posterior density (HPD) regions. Ghosh and Mukerjee (1992b) investigated a similar problem starting from the conditional LR statistic of Cox and Reid (1987). For further references and a review of the literature, we refer to Lee (1989); see also Severini (1991) in this connection.

In the same spirit as with the LR statistic, one may wish to investigate conditions leading to approximate frequentist validity of posterior credible regions based on other popularly used statistics like Rao's score statistic or Wald's statistic. The present work addresses this issue in the multiparameter case with emphasis on the score statistic. The resulting conditions are seen to be stronger than the corresponding condition based on the posterior Bartlett-corrected LR statistic. In fact, it is seen that conditions obtained via the score statistic can supplement those based on the LR statistic and the HPD region and can be useful in making a choice from amongst priors satisfying the latter conditions. Examples have been given in Section 3 to illustrate this point. Here we shall be primarily concerned with a slightly modified version, say T^*, of the score statistic, which seems to be natural in a Bayesian set-up. The consequences of dealing with Wald's statistic or the original version of the score statistic will also be briefly indicated.

It may be of some interest to note that if a posterior region based on T^*, as considered here, has approximate frequentist validity then, apart from having a Bayesian interpretation, such a region will have a higher order frequentist behaviour similar to that of one given by the more traditional approach based on inversion of the LR statistic. Invoking the duality between acceptance regions of tests and credible regions, this follows from known results on higher order comparison of tests (Mukerjee (1993)) which can be utilized to show that in a frequentist set-up a test given by T^* will have the same average power, up to the second order, as the LR test, the average being taken over spherical regions centered at the null hypothetical value with the per observation information matrix at the null hypothetical value used as a Riemannian metric. Thus, in addition to being helpful in the search for non-informative priors as mentioned in the last paragraph, the present results can be useful in obtaining accurate frequentist regions with attractive frequentist properties.

2. Posterior Distribution of the Score Statistic

Let $\{X_i\}$, $i \geq 1$, be a sequence of independent and identically distributed possibly vector-valued random variables with a common density $f(x;\theta)$ where $\theta = (\theta_1, \ldots, \theta_p)' \in R^p$ or some open subset thereof. We make the asssumptions of Johnson (1970). Let θ have a prior density $\pi(\cdot)$ which is positive and thrice continuously differentiable at all θ. If $\pi(\cdot)$ is not proper, 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. Let $X=(X_1,\ldots,X_n)'$, where n is the sample size, $l(\theta)=n^{-1}\sum_{i=1}^n \log f(X_i;\theta)$ and $\hat\theta$ be the maximum likelihood estimator of θ based on X. Define $\hat\pi=\pi(\hat\theta)$ and for $1\le i,j,r,s\le p$, let $\pi_i(\theta)=D_i\pi(\theta)$, $\pi_{ij}(\theta)=D_iD_j\pi(\theta)$, $\hat\pi_i=\pi_i(\hat\theta)$, $\hat\pi_{ij}=\pi_{ij}(\hat\theta)$, $a_i=\{D_i\,\ell(\theta)\}_{\theta=\hat\theta}$, $a_{ij}=\{D_iD_j\,\ell(\theta)\}_{\theta=\hat\theta}$, $c_{ij}=-a_{ij}$, $a_{ijr}=\{D_iD_jD_r\,\ell(\theta)\}_{\theta=\hat\theta}$, $a_{ijrs}=\{D_iD_jD_rD_s\,\ell(\theta)\}_{\theta=\hat\theta}$, where D_i is the operator of partial differentiation with respect to θ_i. All formal expansions for the posterior, as used here, are valid for sample points in a set S, which may be defined along the line of Bickel and Ghosh (1990; Section 2 with $m=3$), with P_θ-probability $1+O(n^{-2})$ uniformly over compact sets of θ. The $p\times p$ matrix $C=((c_{ij}))$ is positive definite over S. Let $C^{-1}=((c^{ij}))$.

The original version of the score statistic, as defined in Rao (1948), is given by $T\equiv T(X,\theta)=\{\delta(X,\theta)\}'I^{-1}\{\delta(X,\theta)\}$, where $\delta(X,\theta)=n^{\frac12}(D_1\,\ell(\theta),\ldots,D_p\,\ell(\theta))'$ and $I\equiv I(\theta)$ is the per observation information matrix at θ which is assumed to be positive definite at each θ. This is motivated by the fact that under θ, $\delta(X,\theta)$ has a null mean vector and covariance matrix I. However, in the posterior set-up, up to the first order of approximation, $\delta(X,\theta)$ will continue to have a null mean vector but will have covariance matrix C (see (2.2a), (2.3) below). As such, in a posterior set-up, it appears to be natural (cf. Efron and Hinkley (1978)) to consider a slightly modified version of the score statistic given by

$$T^*\equiv T^*(X,\theta)=\{\delta(X,\theta)\}'\,C^{-1}\{\delta(X,\theta)\}. \tag{2.1}$$

Here we shall be primarily concerned with T^*.

We begin by finding an expression for the approximate posterior characteristic function (c.f.) of T^* which helps in the explicit derivation of posterior credible sets based on T^*. Throughout, unless otherwise stated, the summation convention will be followed, i.e., summation will be implied over repeated subscripts or superscripts. For example, $a_{ijr}h_jh_r$ and $c^{ij}\hat\pi_{ij}$ will stand for $\sum_j\sum_r a_{ijr}h_jh_r$ and $\sum_i\sum_j c^{ij}\hat\pi_{ij}$ respectively.

Let $h=(h_1,\ldots,h_p)'=n^{\frac12}(\theta-\hat\theta)$. As noted in Ghosh and Mukerjee (1991), the posterior density of h under the prior $\pi(\cdot)$ is given by

$$\begin{aligned}\tilde\pi(h|X)=\phi(h;C^{-1})\Big[1&+n^{-\frac12}\Big\{U_{11}(\pi,h)+\frac16U_{12}(h)\Big\}+n^{-1}\Big\{\frac12(U_{21}(\pi,h)-G_1(\pi))\\&+\frac1{24}(U_{22}(h)-G_2)+\frac16(U_{11}(\pi,h)U_{12}(h)-G_3(\pi))\\&+\frac1{72}(U_{12}^2(h)-G_4)\Big\}\Big]+o(n^{-1}),\end{aligned} \tag{2.2a}$$

where $\phi(\cdot; C^{-1})$ is the p-variate normal density with null mean vector and covariance matrix C^{-1}, and, with $c^{(1)}_{ijrs} = c^{ij}c^{rs} + c^{ir}c^{js} + c^{is}c^{jr}$,

$$U_{11}(\pi, h) = \hat{\pi}^{-1} h_i \hat{\pi}_i, \quad U_{12}(h) = a_{ijr} h_i h_j h_r, \tag{2.2b}$$

$$U_{21}(\pi, h) = \hat{\pi}^{-1} h_i h_j \hat{\pi}_{ij}, \quad U_{22}(h) = a_{ijrs} h_i h_j h_r h_s, \tag{2.2c}$$

$$G_1(\pi) = \hat{\pi}^{-1} c^{ij} \hat{\pi}_{ij}, \quad G_2 = a_{ijrs} c^{(1)}_{ijrs}, \quad G_3(\pi) = \hat{\pi}^{-1} a_{ijr} \hat{\pi}_s c^{(1)}_{ijrs}, \tag{2.2d}$$

$$G_4 = a_{ijr} a_{suv} (9c^{ij} c^{rs} c^{uv} + 6c^{is} c^{ju} c^{rv}), \tag{2.2e}$$

each of the implicit summations being over the range 1 to p. Now, for $1 \le i \le p$, an expansion about $\hat{\theta}$ yields

$$\begin{aligned} n^{\frac{1}{2}} D_i\, \ell(\theta) &= n^{\frac{1}{2}} D_i\, \ell(\hat{\theta} + n^{-\frac{1}{2}} h) \\ &= -\, c_{ij} h_j + \frac{1}{2} n^{-\frac{1}{2}} a_{ijr} h_j h_r + \frac{1}{6} n^{-1} a_{ijrs} h_j h_r h_s + o(n^{-1}). \end{aligned} \tag{2.3}$$

Hence by (2.1), (2.2b,c),

$$T^* = h'Ch - n^{-\frac{1}{2}} U_{12}(h) + n^{-1} \Big(\frac{1}{4} U_3(h) - \frac{1}{3} U_{22}(h) \Big) + o(n^{-1}), \tag{2.4a}$$

where

$$U_3(h) = c^{uv} a_{uij} a_{vrs} h_i h_j h_r h_s. \tag{2.4b}$$

With $\xi = (-1)^{\frac{1}{2}} t$, by (2.2a), (2.4a), the approximate posterior c.f. of T^* under the prior $\pi(\cdot)$ is given by

$$\begin{aligned} &\psi_\pi(\xi|X) \\ &= \int \tilde{\pi}(h|X)\, \exp(\xi\, T^*)\, dh \\ &= \int \Big[1 + n^{-\frac{1}{2}} \{ U_{11}(\pi, h) + (\frac{1}{6} - \xi) U_{12}(h) \} \\ &\quad + n^{-1} \Big\{ \frac{1}{2} (U_{21}(\pi, h) - G_1(\pi)) + \frac{1}{24} (U_{22}(h) - G_2) + \frac{1}{6} (U_{11}(\pi, h) U_{12}(h) - G_3(\pi)) \\ &\quad + \frac{1}{72} (U_{12}^2(h) - G_4) + \xi (\frac{1}{4} U_3(h) - \frac{1}{3} U_{22}(h)) - \xi U_{12}(h) (U_{11}(\pi, h) + \frac{1}{6} U_{12}(h)) \\ &\quad + \frac{1}{2} \xi^2 U_{12}^2(h) \Big\} \Big] e^{\xi h' C h} \phi(h; C^{-1})\, dh + o(n^{-1}). \end{aligned}$$

After some simplification with the help of (2.2b-e), (2.4b), the above yields

$$\psi_\pi(\xi|X) = (1-2\xi)^{-\frac{1}{2}p} + n^{-1} \Big\{ \sum_{j=0}^{2} H_j(\pi) (1-2\xi)^{-(\frac{1}{2}p+j)} + H_3 (1-2\xi)^{-(\frac{1}{2}p+3)} \Big\} + o(n^{-1}), \tag{2.5}$$

where

$$H_0(\pi) = -\frac{1}{72}\{36G_1(\pi) + 3G_2 + 12G_3(\pi) + G_4\}, \tag{2.6a}$$

$$H_1(\pi) = \frac{1}{24}\{12G_1(\pi) + 4G_2 + 12G_3(\pi) + 3G_4 - 3G_5\}, \tag{2.6b}$$

$$H_2(\pi) = -\frac{1}{24}\{3G_2 + 8G_3(\pi) + 4G_4 - 3G_5\}, H_3 = \frac{1}{18}G_4, G_5 = c^{ij}a_{irs}a_{juv}c^{(1)}_{rsuv}. \tag{2.6c}$$

For $0 < \alpha < 1$, let z_α^2 be the αth quantile of a central chi-square variate with p degress of freedom. Also, for positive integral ν, Let $K_\nu(\cdot)$ and $k_\nu(\cdot)$ denote respectively the cumulative distribution function and the probability density function of a central chi-square variate with ν degrees of freedom. Let

$$Q^*(\alpha,\pi) = z_\alpha^2 - n^{-1}\{k_p(z_\alpha^2)\}^{-1}\Big\{\sum_{j=0}^{2} H_j(\pi)K_{p+2j}(z_\alpha^2) + H_3K_{p+6}(z_\alpha^2)\Big\}. \tag{2.7}$$

Then, writing $P^\pi\{\cdot|X\}$ for the posterior probability measure of θ under the prior $\pi(\cdot)$ and inverting (2.5), a step which can be justified following Chandra and Ghosh (1979), it follows that

$$P^\pi\{T^*(X,\theta) \le Q^*(\alpha,\pi)|X\} = \alpha + o(n^{-1}). \tag{2.8}$$

Hence $Q^*(\alpha,\pi)$ may be regarded as the αth posterior quantile, up to the order of approximation $o(n^{-1})$, of T^* under the prior $\pi(\cdot)$. By (2.8), the credible set

$$R_{\alpha,\pi}(X) = \{\theta : T^*(X,\theta) \le Q^*(\alpha,\pi)\}, \tag{2.9}$$

given by T^*, has posterior coverage probability $\alpha + o(n^{-1})$.

Remark 1. (a) With reference to a problem posed in Cox (1988), starting from (2.5) and proceeding as in Cordeiro and Ferrari (1991), it is possible to suggest a posterior Bartlett-type adjustment for T^*. However, arguments similar to those used in this and the next section show that priors ensuring frequentist validity, up to $o(n^{-1})$, of posterior credible sets based on such a posterior Bartlett-type adjusted version are precisely the same as those obtained through T^* itself (see (3.4) below). Hence this aspect will not be further considered in the sequel. (b) As noted in Ghosh and Mukerjee (1992b), (2.2a) is in agreement with the findings in Tierney and Kadane (1986). For analytical studies like the present one, the use of (2.2a) seems more convenient than that of the results in Tierney and Kadane (1986), whereas for numerical approximations it should be the other way around.

3. Frequentist Validity of Posterior Regions

We shall now characterize priors ensuring approximate frequentist validity of posterior credible sets given by (2.9). This calls for evaluation of $P_\theta\{T^*(X,\theta) \le Q^*(\alpha,\pi)\}$ up to $o(n^{-1})$. We shall follow the approach in Ghosh and Mukerjee (1991) which is reminiscent of that in Dawid (1991). We take a prior $\bar{\pi}(\cdot)$ satisfying the conditions in Bickel and Ghosh (1990) which are to some extent stronger than those in Johnson (1970) and make Edgeworth assumptions as in Bickel and Ghosh (1990, p.1078). Then, analogously to (2.5), one can obtain the approximate posterior c.f. of T^* under $\bar{\pi}(\cdot)$ and then use (2.6a-c), (2.7) and the facts (i) $K_\nu(z_\alpha^2) - K_{\nu+2}(z_\alpha^2) = 2k_{\nu+2}(z_\alpha^2)$, (ii) $k_{\nu+2}(z_\alpha^2)/k_\nu(z_\alpha^2) = z_\alpha^2/\nu$, to get

$$\begin{aligned} &P^{\bar{\pi}}\{T^*(X,\theta) \le Q^*(\alpha,\pi)|X\} \\ &= \alpha + n^{-1}\sum_{j=0}^{2}\{H_j(\bar{\pi}) - H_j(\pi)\}K_{p+2j}(z_\alpha^2) + o(n^{-1}) \\ &= \alpha + \frac{1}{3}n^{-1}k_{p+2}(z_\alpha^2)[3\{G_1(\pi) - G_1(\bar{\pi})\} + \{1 - 2(p+2)^{-1}z_\alpha^2\}\{G_3(\pi) - G_3(\bar{\pi})\}] \\ &\quad + o(n^{-1}). \end{aligned} \tag{3.1}$$

Let $I^{-1} = ((I^{ij}))$ and for $1 \le i,j,r,s \le p$, define $I^{(1)}_{ijrs} = I^{ij}I^{rs} + I^{ir}I^{js} + I^{is}I^{jr}$, $V_i = D_i \log f(X_1;\theta)$, $V_{ij} = D_iD_j \log f(X_1;\theta)$, $V_{ijr} = D_iD_jD_r \log f(X_1;\theta)$, $L_{ijr} = E_\theta(V_{ijr})$, $L_{i,jr} = E_\theta(V_iV_{jr})$, $L_{i,j,r} = E_\theta(V_iV_jV_r)$. Note that I^{ij}, L_{ijr}, $L_{i,jr}$, $L_{i,j,r}$ are functions of θ. By (2.2d), (3.1),

$$\begin{aligned} &E_\theta\,[P^{\bar{\pi}}\{T^*(X,\theta) \le Q^*(\alpha,\pi)|X\}] \\ &= \alpha + \frac{1}{3}n^{-1}k_{p+2}(z_\alpha^2)\Big[3I^{ij}\Big(\frac{\pi_{ij}(\theta)}{\pi(\theta)} - \frac{\bar{\pi}_{ij}(\theta)}{\bar{\pi}(\theta)}\Big) + \Big(1 - \frac{2z_\alpha^2}{p+2}\Big)L_{ijr}I^{(1)}_{ijrs}\Big(\frac{\pi_s(\theta)}{\pi(\theta)} - \frac{\bar{\pi}_s(\theta)}{\bar{\pi}(\theta)}\Big)\Big] \\ &\quad + o(n^{-1}), \end{aligned} \tag{3.2}$$

We now choose $\bar{\pi}(\cdot)$ such that $\bar{\pi}(\cdot)$ and its first order partial derivatives vanish on the boundaries of a rectangle containing θ as an interior point. We then integrate $E_\theta[P^{\bar{\pi}}\{T^*(X,\theta) \le Q^*(\alpha,\pi)|X\}]$, as shown in (3.2), by parts with respect to such a $\bar{\pi}(\cdot)$ and finally allow $\bar{\pi}(\cdot)$ to converge weakly to the degenerate measure at θ. After some simplification, this yields

$$\begin{aligned} &P_\theta\{T^*(X,\theta) \le Q^*(\alpha,\pi)\} \\ &= \alpha + n^{-1}k_{p+2}(z_\alpha^2)\{\pi(\theta)\}^{-1}[M_1(\pi) + \{1 - 2(p+2)^{-1}z_\alpha^2\}M_2(\pi)] + o(n^{-1}), \end{aligned} \tag{3.3a}$$

where

$$M_1(\pi) = I^{ij}\pi_{ij}(\theta) - \pi(\theta)D_iD_jI^{ij} = D_iD_j\{I^{ij}\pi(\theta)\} - 2D_i\{\pi(\theta)(D_jI^{ij})\}, \tag{3.3b}$$

$$M_2(\pi) = \frac{1}{3}D_s\{L_{ijr}I^{(1)}_{ijrs}\pi(\theta)\} = D_s\{L_{ijr}I^{ij}I^{rs}\pi(\theta)\}, \tag{3.3c}$$

as L_{ijr} and I^{ij} are invariant under permutation of subscripts and superscripts respectively.

By (2.9), (3.3a), frequentist validity, up to $o(n^{-1})$, holds for posterior credible sets based on T^* if and only if $\pi(\cdot)$ satisfies the partial differential equations

$$M_1(\pi) = 0, \quad M_2(\pi) = 0. \tag{3.4}$$

Remark 2. Proceeding as in Ghosh and Mukerjee (1991) and using (3.3b,c), it can be shown that (3.4) is satisfied by Jeffreys' prior, namely $\pi_0(\theta) \propto \{\det\, I(\theta)\}^{\frac{1}{2}}$, if and only if $D_i[I^{ij}I^{rs}L_{j,r,s}\{\det\, I(\theta)\}^{\frac{1}{2}}] = 0$ and $D_i[I^{ij}I^{rs}L_{jrs}\{\det\, I(\theta)\}^{\frac{1}{2}}] = 0$. The above conditions hold under location, scale and many other models – e.g., with $p = 1$, under the bivariate normal model with zero means, variances 1 and $1 + \theta^2$ and covariance θ. There are also models where Jeffreys' prior does not solve (3.4) but other solutions to (3.4) are available; see Example 1 below in this connection.

Remark 3. As shown in Ghosh and Mukerjee (1991), frequentist validity, up to $o(n^{-1})$, holds for posterior regions based on a posterior Bartlett corrected LR statistic if and only if

$$M_1(\pi) + M_2(\pi) = 0. \tag{3.5}$$

Also, following Ghosh and Mukerjee (1993), frequentist validity, up to $o(n^{-1})$, holds for HPD regions if and only if

$$D_i\left\{I^{ij}\pi_j(\theta)\right\} + D_s\left\{L_{j,ir}I^{ij}I^{rs}\pi(\theta)\right\} = 0. \tag{3.6}$$

Note that (3.4) is stronger than (3.5). The following examples illustrate how (3.4) can supplement (3.5) and (3.6) and thus help in making a choice from amongst priors satisfying the latter conditions. In the process, it will be useful to note from (3.3b) that $M_1(\pi)$ can also be expressed as

$$M_1(\pi) = D_i\left\{I^{ij}\pi_j(\theta)\right\} - D_s\left\{(L_{j,ir} + L_{ijr})I^{ij}I^{rs}\pi(\theta)\right\}, \tag{3.7}$$

which follows since $D_u I^{ij} = I^{ir}I^{js}(L_{u,rs} + L_{urs})$ (cf. Ghosh and Mukerjee (1991)).

Example 1. (Multiparameter exponential family) Let $f(x;\theta)$ be of the form

$$f(x;\theta) = W(x)\,\exp\Big\{\sum_{i=1}^{p}\theta_i W_i(x) - A(\theta)\Big\}. \tag{3.8}$$

For $1 \le i,j,r \le p$, let $A_i(\theta) = D_iA(\theta)$, $A_{ij}(\theta) = D_iD_jA(\theta)$, $A_{ijr}(\theta) = D_iD_jD_rA(\theta)$. Then it is easily seen that here $I_{ij} = A_{ij}(\theta)$, $L_{j,ir} = 0$, $L_{ijr} = -A_{ijr}(\theta)$ and $L_{i,j,r} = A_{ijr}(\theta)$, the last identity being a consequence of the regularity condition

$L_{ijr} + L_{i,jr} + L_{j,ir} + L_{r,ij} + L_{i,j,r} = 0$ $(1 \le i,j,r \le p)$. Hence, by (3.3c), (3.7), the conditions (3.5) and (3.6), arising from the LR statistic and the HPD region respectively, are both equivalent to

$$D_i \left\{ A^{ij}(\theta)\pi_j(\theta) \right\} = 0, \tag{3.9}$$

where $A^{ij}(\theta)$ is the (i,j)th element of the inverse of the $p \times p$ matrix with typical element $A_{ij}(\theta)$. Similarly, the conditions (3.4), obtained via the score statistic reduce to

$$D_i \left\{ A^{ij}(\theta)\pi_j(\theta) \right\} = 0, \quad D_s \left\{ A_{ijr}(\theta) A^{ij}(\theta) A^{rs}(\theta)\pi(\theta) \right\} = 0. \tag{3.10}$$

The equation (3.9) can have many solutions. For example, $\pi(\theta) =$ constant is a solution of (3.9). Furthermore, if for any u, $A_u(\theta)$ is either positive or negative for all θ then $\pi(\theta) \propto A_u(\theta)$ is also an admissible solution of (3.9). The conditions in (3.10), which incorporate (3.9) and hence are stronger than (3.9), can help in making a choice from amongst these and other solutions of (3.9).

For a more specific illustration, consider the inverse Gaussian model with location parameter $(\theta_1/\theta_2)^{\frac{1}{2}}$ and scale parameter θ_1, where $\theta_1, \theta_2 > 0$. Then

$$f(x;\theta) = \theta_1^{\frac{1}{2}}(2\pi x^3)^{-\frac{1}{2}} \exp\left[-(2x)^{-1}\theta_2\{x - (\theta_1/\theta_2)^{\frac{1}{2}}\}^2\right], \quad x > 0,$$

which is of the form (3.8) with $p = 2$ and $A(\theta) = -\frac{1}{2}\log\theta_1 - (\theta_1\theta_2)^{\frac{1}{2}}$. Hence, considering $A_1(\theta)$ and $A_2(\theta)$, as noted above, $\pi^{(1)}(\theta) \propto \{\theta_1^{-1} + (\theta_2/\theta_1)^{\frac{1}{2}}\}$ and $\pi^{(2)}(\theta) \propto (\theta_1/\theta_2)^{\frac{1}{2}}$ are obtained, in addition to $\pi^{(0)}(\theta) =$ constant, as solutions of (3.9). Also, consideration of priors of the form $\pi(\theta) \propto \theta_1^{\beta_1}\theta_2^{\beta_2}$ yield further solutions of (3.9) as $\pi^{(4)}(\theta) \propto \theta_1^{-2}$ and $\pi^{(5)}(\theta) \propto (\theta_1^3\theta_2)^{-\frac{1}{2}}$. It can be checked that among these solutions of (3.9) only $\pi^{(5)}(\theta)$ satisfies (3.10). Thus (3.10), obtained from the score statistic, can help in discriminating among priors reached via consideration of the LR statistic or the HPD region. Incidentally, under the present inverse Gaussian model, $I_{11} = \frac{1}{2}\theta_1^{-2} + \frac{1}{4}(\theta_1^{-3}\theta_2)^{\frac{1}{2}}$, $I_{12} = I_{21} = -\frac{1}{4}(\theta_1\theta_2)^{-\frac{1}{2}}$, $I_{22} = \frac{1}{4}(\theta_1\theta_2^{-3})^{\frac{1}{2}}$, and it may be seen that Jeffreys' prior, given by $\pi_0(\theta) \propto (\theta_1\theta_2)^{-3/4}$ does not satisfy (3.10).

Example 2. (Multiparameter scale family) Let $f(x;\theta)$ be of the form

$$f(x;\theta) = (\theta_1 \cdots \theta_p)^{-1} g\left(x^{(1)}/\theta_1, \ldots, x^{(p)}/\theta_p\right),$$

where $x = (x^{(1)}, \ldots, x^{(p)})'$ and $\theta_1, \ldots, \theta_p > 0$. Then for $1 \le i,j,r \le p$, we have

$$I_{ij} = b_{ij}/(\theta_i\theta_j),\ I^{ij} = b^{ij}\theta_i\theta_j,\ L_{j,ir} = b_{j,ir}/(\theta_i\theta_j\theta_r),\ L_{ijr} = b_{ijr}/(\theta_i\theta_j\theta_r), \tag{3.11}$$

provided they exist, where the quantities b_{ij}, b^{ij}, $b_{j,ir}$, b_{ijr} are constants free from θ and b^{ij} is the (i,j)th element of the inverse of $B = ((b_{ij}))$ which is assumed to be positive definite. Note that the summation convention is not being followed in (3.11). By (3.3c), (3.7), (3.11), the conditions (3.5) and (3.6), arising from the LR statistic and the HPD region respectively, reduce to

$$b^{ij} D_i\{\theta_i \theta_j \pi_j(\theta)\} - \lambda_{1s} D_s\{\theta_s \pi(\theta)\} = 0, \tag{3.12a}$$

and

$$b^{ij} D_i\{\theta_i \theta_j \pi_j(\theta)\} + \lambda_{1s} D_s\{\theta_s \pi(\theta)\} = 0, \tag{3.12b}$$

where $\lambda_{1s} = b_{j,ir} b^{ij} b^{rs}$, $1 \le s \le p$. Similarly, the conditions (3.4), obtained via the score statistic, are now equivalent to

$$b^{ij} D_i\{\theta_i \theta_j \pi_j(\theta)\} - \lambda_{1s} D_s\{\theta_s \pi(\theta)\} = 0, \quad \lambda_{2s} D_s\{\theta_s \pi(\theta)\} = 0, \tag{3.13}$$

where $\lambda_{2s} = b_{ijr} b^{ij} b^{rs}$, $1 \le s \le p$. Unless the vector $(\lambda_{21}, \ldots, \lambda_{2p})'$ is proportional to $(\lambda_{11}, \ldots, \lambda_{1p})'$, (3.13) can strengthen (3.12a,b) and be helpful in making a choice from amongst the priors satisfying the latter conditions.

For a specific illustration, consider the trivariate normal model with a null mean vector, unknown standard deviations θ_1, θ_2, θ_3 and a known correlation matrix

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{pmatrix}.$$

Then $p = 3$, $b_{ij} = b_{ji}$, $b^{ij} = b^{ji}$, $b_{j,ir} = b_{j,ri}$, b_{ijr} is invariant under permutation of subscripts, and an explicit calculation shows that $b_{11} = b_{33} = 5/2$, $b_{22} = 3$, $b_{12} = b_{23} = -1/2$, $b_{13} = 0$, $b^{11} = b^{33} = 29/70$, $b^{22} = 25/70$, $b^{12} = b^{23} = 5/70$, $b^{13} = 1/70$, $b_{1,11} = b_{3,33} = -8$, $b_{2,22} = -10$, $b_{1,22} = b_{2,11} = b_{2,33} = b_{3,22} = 1$, $b_{1,12} = b_{2,12} = b_{2,23} = b_{3,23} = 1/2$, $b_{1,33} = b_{3,11} = b_{1,13} = b_{3,13} = b_{1,23} = b_{2,13} = b_{3,12} = 0$, $b_{111} = b_{333} = 13$, $b_{222} = 16$, $b_{112} = b_{122} = b_{223} = b_{233} = -1$, $b_{113} = b_{133} = b_{123} = 0$, $\lambda_{11} = \lambda_{13} = -739/490$, $\lambda_{12} = -727/490$, $\lambda_{21} = \lambda_{23} = 1187/490$, $\lambda_{22} = 1147/490$. We now consider priors of the form $\pi(\theta) \propto \theta_1^{\beta_1} \theta_2^{\beta_2} \theta_3^{\beta_3}$, where β_1, β_2, β_3 are constants. Such a prior satisfies (3.12a,b) if and only if $29(\beta_1^2 + \beta_3^2) + 25\beta_2^2 + 10\beta_2(\beta_1 + \beta_3) + 2\beta_1\beta_3 + 35(\beta_1 + \beta_2 + \beta_3) = 0$ and $739(\beta_1 + \beta_3 + 2) + 727(\beta_2 + 1) = 0$, which have infinitely many solutions for β_1, β_2, β_3. Out of these solutions, only the one given by $\beta_1 = \beta_2 = \beta_3 = -1$, which corresponds to Jeffreys' prior, is seen to satisfy (3.13) as well. Thus (3.13), given by the score statistic, can help in discriminating among priors obtained via consideration of the LR statistic and the HPD region.

Remark 4. Along the line of (2.1), one may wish to consider a version of Wald's statistic given by $T_1^* = n(\hat{\theta} - \theta)' C(\hat{\theta} - \theta)$. Proceeding as in the derivation of (3.4), it can be shown that posterior credible sets based on T_1^* have frequentist validity, up to $o(n^{-1})$, if and only if the conditions in (3.4) hold. For the special case $p = 1$, we have also studied the corresponding conditions arising from the original version of the score statistic, namely T, and the original version of Wald's statistic given by $T_1 = n(\hat{\theta} - \theta)' I(\hat{\theta})(\hat{\theta} - \theta)$. Consideration of T_1 again leads to (3.4) while consideration of T yields the conditions

$$I^{-1}(d\,\pi(\theta)/d\theta) - I^{-2} L_{1,11} \pi(\theta) = \text{constant}, \quad I^{-2} L_{1,1,1} \pi(\theta) = \text{constant},$$

noting that, with $p = 1$, $I = I(\theta)$ is a scalar.

Acknowledgement

We thank the Chair Editor, the Associate Editor and the referees for very constructive suggestions. This work was supported by the US Army Research Office and the Centre for Management and Development Studies, Indian Institute of Management, Calcutta.

References

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. Sankhyā Ser.A **41**, 22-47.

Cordeiro, G. M. and Ferrari, S. L. P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. Biometrika **78**, 573-582.

Cox, D. R. (1988). Some aspects of conditional and asymptotic inference: A review. Sankhyā Ser.A **50**, 314-337.

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

Dawid, A. P. (1991). Fisherian inference in likelihood and prequential frames of reference (with discussion). J. Roy. Statist. Soc. Ser.B **53**, 79-109.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). Biometrika **65**, 457-487.

Ghosh, J. K. and Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. J. Multivariate Anal. **38**, 385-393.

Ghosh, J. K. and Mukerjee, R. (1992a). Non-informative priors (with discussion). Bayesian Statistics 4 (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 195-210, Oxford University Press.

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

Ghosh, J. K. and Mukerjee, R. (1993). Frequentist validity of highest posterior density regions in the multiparameter case. Ann. Inst. Statist. Math. **45**, 293-302.

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

Lee, C. B. (1989). Comparison of frequentist coverage probability and Bayesian posterior coverage probability, and applications. Ph.D. thesis, Purdue University.

Mukerjee, R. (1993). Rao's score test: Recent asymptotic results. Handbook of Statistics **11** (Edited by G. S. Maddala, C. R. Rao and H. D. Vinod), 363-379, North-Holland, Amsterdam.

Nicolau, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. J. Roy. Statist. Soc. Ser.B **55**, 377-390.

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

Peers, H. W. (1968). Confidence properties of Bayesian interval estimates. J. Roy. Statist. Soc. Ser.B **30**, 535-544.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. Proc. Cambridge Philos. Soc. **44**, 50-57.

Severini, T. A. (1991). On the relationship between Bayesian and non-Bayesian interval estimates. J. Roy. Statist. Soc. Ser.B **53**, 611-618.

Stein, C. (1985). On coverage probability of confidence sets based on a prior distribution. Sequential Methods in Statist. 16 (Banach Center Publications), 485-514, Polish Scientific Publishers, Warsaw.

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. Amer. Statist. Assoc. **81**, 82-86.

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

Department of Statistics, The Pennsylvania State University, University Park, PA 16802, U.S.A.

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

(Received April 1993; accepted September 1994)