Ann. Inst. Statist. Math.
Vol. 45, No. 2, 293–302 (1993)

FREQUENTIST VALIDITY OF HIGHEST POSTERIOR DENSITY REGIONS IN THE MULTIPARAMETER CASE

J. K. GHOSH[1] AND RAHUL MUKERJEE[2,*]

[1] *Stat-Math Division, Indian Statistical Institute, 203, B. T. Road, Calcutta 700 035, India*
[2] *Indian Institute of Management, Joka, Diamond Harbour Road, Post Box No. 16757, Alipore Post Office, Calcutta 700 027, India*

(Received June 1, 1991; revised November 16, 1991)

Abstract. In a multiparameter set-up, this paper characterizes priors which ensure frequentist validity, up to $o(n^{-1})$, of confidence regions based on the highest posterior density. The role of Jeffreys' prior in this regard has also been investigated.

Key words and phrases: Highest posterior density, Jeffreys' prior, non-informative prior, parametric orthogonality.

1. Introduction

The problem of characterizing priors leading to posterior confidence regions with approximate frequentist validity has received a considerable attention in recent years. As noted in Tibshirani (1989), such a study can have practical utility in two ways: (a) it provides a method for constructing accurate frequentist confidence regions and (b) it helps in defining a non-informative prior which could be potentially useful for comparative purposes in Bayesian analysis. Welch and Peers (1963) considered this problem, with reference to one-sided confidence sets, in the one-parameter case. This work was extended by Stein (1985) who gave the explicit form of the difference between the posterior and the frequentist coverage probabilities. Tibshirani (1989) extended the findings in Welch and Peers (1963) to a situation where interest lies in one of several parameters (see also Peers (1965)). Lee (1989) explored the frequentist validity of elliptic confidence regions and also of half spaces in the multiparameter case while Ghosh and Mukerjee (1991, 1992) considered the same problem with reference to posterior regions based on posterior Bartlett-corrected likelihood ratio and conditional likelihood ratio statistics (see also Bickel and Ghosh (1990) in this context). Loh (1988) considered a problem of this kind with reference to confidence sets for a multivariate normal mean. For further references and an excellent review of the literature, we refer to Lee (1989).

* The work of Rahul Mukerjee was supported by a grant from the Centre for Management and Development Studies, Indian Institute of Management Calcutta.

The objective of the present work is to characterize, in the multiparameter case, priors ensuring, up to $o(n^{-1})$, the frequentist validity of confidence regions based on the highest posterior density (HPD). In the Bayesian context, the use of such regions is particularly appealing (see e.g., Lindley (1965), p. 25). The present problem was studied earlier in the one-parameter case by Peers (1968) who considered posterior regions in the form of intervals with equal posterior densities at the extremeties. The approach of Peers (1968) does not seem to work in the multiparameter case and new techniques are called for. This has been attempted in the next section. We have also investigated conditions under which the HPD regions based on Jeffreys' prior have frequentist validity up to $o(n^{-1})$.

2. Results

Let $\{X_i\}$, $i \geq 1$, be a sequence of independent and identically distributed possibly vector-valued random variables with common density $f(x;\theta)$ where $\theta = (\theta_1, \ldots, \theta_p)' \in \mathcal{R}^p$ or some open subset thereof. We make the assumptions in Johnson (1970). Let θ have a prior density $\pi(\cdot)$ which, as in Johnson (1970), is positive and thrice continuously differentiable for all θ. In case $\pi(\cdot)$ 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. In this case, Johnson's (1970) proof for a proper prior goes through. Let P be the joint probability measure of θ and $X = (X_1, \ldots, X_n)'$, where n is the sample size. 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 on compact sets of θ.

Let $L(\theta) = \sum_{i=1}^n \log f(X_i;\theta)$, $l(\theta) = n^{-1}L(\theta)$, and for $1 \leq i, j, r, s \leq p$,

$$a_i = \{D_i l(\theta)\}_{\theta=\hat{\theta}}, \qquad a_{ij} = \{D_i D_j l(\theta)\}_{\theta=\hat{\theta}}, \qquad c_{ij} = -a_{ij},$$
$$a_{ijr} = \{D_i D_j D_r l(\theta)\}_{\theta=\hat{\theta}}, \qquad a_{ijrs} = \{D_i D_j D_r D_s l(\theta)\}_{\theta=\hat{\theta}},$$

where D_i is the operator of partial differentiation with respect to θ_i and $\hat{\theta}$ is the maximum likelihood estimator of θ. The $p \times p$ matrix $C = ((c_{ij}))$ is positive definite over S.

With a prior $\pi(\cdot)$ for θ, we consider a HPD region $R_\pi(X)$ for θ of the form

$$R_\pi(X) = \{\theta \colon \pi(\theta \mid X) \geq k(\pi, X)\}, \tag{2.1}$$

where $k(\pi, X)$ is such that

$$P^\pi(\theta \in R_\pi(X) \mid X) = 1 - \alpha + o(n^{-1}), \tag{2.2}$$

$P^\pi(\cdot \mid X)$ is the posterior probability measure for θ under the prior $\pi(\cdot)$, $0 < \alpha < 1$, and

$$\pi(\theta \mid X) = \pi(\theta) \exp\{L(\theta) - L(\hat{\theta})\}/N_\pi(X) \tag{2.3a}$$

is the posterior density of θ, with

$$N_\pi(X) = \int \pi(\theta)\exp\{L(\theta) - L(\hat{\theta})\}d\theta. \tag{2.3b}$$

Let $h \equiv h(\theta) = n^{1/2}(\theta - \hat{\theta})$. Define $\hat{\pi} = \pi(\hat{\theta})$, $\hat{\pi}_i = \{D_i\pi(\theta)\}_{\theta=\hat{\theta}}$ $(1 \le i \le p)$, $\hat{\pi}_{ij} = \{D_iD_j\pi(\theta)\}_{\theta=\hat{\theta}}$ $(1 \le i, j \le p)$ and so on. Then, as noted in Ghosh and Mukerjee ((1991), hereafter abbreviated to GM), from (2.3a), (2.3b) it can be seen by a Taylor's expansion that the posterior density of $h = (h_1, \ldots, h_p)'$ under the prior $\pi(\cdot)$ is given by

$$\begin{aligned}\tilde{\pi}(h \mid X) = \phi(h; C^{-1})\Big[1 &+ n^{-1/2}\Big\{T_{11}(\pi, h) + \frac{1}{6}T_{12}(h)\Big\}\\ &+ n^{-1}\Big\{\frac{1}{2}(T_{21}(\pi, h) - G_1(\pi))\\ &+ \frac{1}{24}(T_{22}(h) - G_2)\\ &+ \frac{1}{6}(T_{11}(\pi, h)T_{12}(h) - G_3(\pi))\\ &+ \frac{1}{72}(T_{12}^2(h) - G_4)\Big\}\Big] + o(n^{-1}),\end{aligned} \tag{2.4}$$

where $\phi(\cdot; C^{-1})$ is the multivariate normal density with null mean vector and dispersion matrix C^{-1}, and

$$\begin{cases} T_{11}(\pi, h) = T_{11}(\pi, X, h) = \hat{\pi}^{-1}\sum_i h_i\hat{\pi}_i,\\ T_{12}(h) = T_{12}(X, h) = \sum_i\sum_j\sum_r h_ih_jh_ra_{ijr},\\ T_{21}(\pi, h) = T_{21}(\pi, X, h) = \hat{\pi}^{-1}\sum_i\sum_j h_ih_j\hat{\pi}_{ij},\\ T_{22}(h) = T_{22}(X, h) = \sum_i\sum_j\sum_r\sum_s h_ih_jh_rh_sa_{ijrs}, \end{cases} \tag{2.5}$$

$$\begin{cases} G_1(\pi) = G_1(\pi, X) = \hat{\pi}^{-1}\sum_i\sum_j c^{ij}\hat{\pi}_{ij},\\ G_2 = G_2(X) = \sum_i\sum_j\sum_r\sum_s c_{ijrs}^{(1)}a_{ijrs},\\ G_3(\pi) = G_3(\pi, X) = \hat{\pi}^{-1}\sum_i\sum_j\sum_r\sum_s a_{ijr}\hat{\pi}_sc_{ijrs}^{(1)},\\ G_4 = G_4(X) = \sum_i\sum_j\sum_r\sum_s\sum_u\sum_v a_{ijr}a_{suv}c_{ijrsuv}^{(2)}, \end{cases} \tag{2.6}$$

$$(2.7)\quad \begin{cases} C^{-1} = ((c^{ij})), \quad c^{(1)}_{ijrs} = c^{ij}c^{rs} + c^{ir}c^{js} + c^{is}c^{jr}, \\ c^{(2)}_{ijrsuv} = c^{ij}c^{rs}c^{uv} + c^{ij}c^{ru}c^{sv} + c^{ij}c^{rv}c^{su} + c^{ir}c^{js}c^{uv} + c^{ir}c^{ju}c^{sv} \\ \qquad + c^{ir}c^{jv}c^{su} + c^{is}c^{jr}c^{uv} + c^{is}c^{ju}c^{rv} + c^{is}c^{jv}c^{ru} + c^{iu}c^{jr}c^{sv} \\ \qquad + c^{iu}c^{js}c^{rv} + c^{iu}c^{jv}c^{rs} + c^{iv}c^{jr}c^{su} + c^{iv}c^{js}c^{ru} + c^{iv}c^{ju}c^{rs}, \end{cases}$$

each of the summations in the above being over the range 1 to p. As an intermediate step in the derivation of (2.4), from (2.3a) one also gets

$$\pi(\theta \mid X) = \hat{\pi}\left[1 + n^{-1/2}\left\{T_{11}(\pi, h) + \frac{1}{6}T_{12}(h)\right\} + n^{-1}\left\{\frac{1}{2}T_{21}(\pi, h) + \frac{1}{24}T_{22}(h) + \frac{1}{6}T_{11}(\pi, h)T_{12}(h) + \frac{1}{72}T_{12}^2(h)\right\}\right] e^{-h'Ch/2}/N_\pi(X) + o(n^{-1}),$$

whence after some algebra one obtains

$$(2.8)\quad \pi(\theta \mid X) = \frac{\hat{\pi}}{N_\pi(X)} \exp\left\{-\frac{1}{2}W(\pi, X, h) + \frac{1}{2}n^{-1}G_5(\pi)\right\} + o(n^{-1}),$$

where

$$(2.9)\quad G_5(\pi) = G_5(\pi, X) = \hat{\pi}^{-2}\sum_i \sum_j \hat{\pi}_i \hat{\pi}_j c^{ij},$$

$$(2.10)\quad W(\pi, X, h) = h'Ch - n^{-1/2}\left\{2T_{11}(\pi, h) + \frac{1}{3}T_{12}(h)\right\} + n^{-1}\left[\left\{T_{11}(\pi, h) + \frac{1}{6}T_{12}(h)\right\}^2 - T_{21}(\pi, h) - \frac{1}{12}T_{22}(h) - \frac{1}{3}T_{11}(\pi, h)T_{12}(h) - \frac{1}{36}T_{12}^2(h) + G_5(\pi)\right],$$

and, as noted above, $h \equiv h(\theta)$. The inclusion of $G_5(\pi)$ in (2.10) simplifies our calculations to some extent.

Let

$$(2.11)\quad \lambda(X) = z^2\left\{1 + \frac{1}{12}(np)^{-1}\left(G_2 + \frac{1}{3}G_4\right)\right\},$$

where z^2 is the upper α-point of a central chi-square variate with p degrees of freedom and G_2, G_4 are as in (2.6). Then, as shown in the Appendix, from (2.4), (2.10), (2.11),

$$(2.12)\quad P^\pi[W(\pi, X, h(\theta)) \le \lambda(X) \mid X] = 1 - \alpha + o(n^{-1}).$$

Hence defining

$$(2.13)\quad k(\pi, X) = \frac{\hat{\pi}}{N_\pi(X)} \exp\left\{-\frac{1}{2}\lambda(X) + \frac{1}{2}n^{-1}G_5(\pi)\right\},$$

it follows from (2.8) and (2.12) that

$$\begin{aligned} (2.14) \qquad & P^{\pi}(\pi(\theta \mid X) \geq k(\pi, X) \mid X) \\ & = P^{\pi}(W(\pi, X, h(\theta)) \leq \lambda(X) \mid X) + o(n^{-1}) \\ & = 1 - \alpha + o(n^{-1}). \end{aligned}$$

Hence in consideration of (2.2), the HPD region may be taken as in (2.1) with $k(\pi, X)$ given by (2.13).

We now proceed to characterize a prior $\pi(\cdot)$ for which the HPD region $R_\pi(X)$, obtained as above, has frequentist validity up to $o(n^{-1})$, that is, for which the relation

$$(2.15) \qquad P_\theta(\theta \in R_\pi(X)) = 1 - \alpha + o(n^{-1})$$

holds for each θ and each α $(0 < \alpha < 1)$. To that effect, we consider a prior $\pi^*(\cdot)$ satisfying the regularity conditions in Bickel and Ghosh ((1990), Section 2 with $m = 3$) which are slightly stronger than those in Johnson (1970) and make Edgeworth assumptions as in Bickel and Ghosh ((1990), p. 1078). Then, proceeding as in the derivation of (2.14),

$$(2.16) \quad P^{\pi^*}(\theta \in R_\pi(X) \mid X) = 1 - \alpha - 2(np)^{-1} z^2 q_p(z^2) H(X; \pi, \pi^*) + o(n^{-1}),$$

where $q_p(\cdot)$ is the central chi-square density with p degrees of freedom and

$$\begin{aligned} (2.17a) \qquad H(X; \pi, \pi^*) = & \frac{1}{2}\{G_1(\pi^*) - G_1(\pi)\} + \frac{1}{6}\{G_3(\pi^*) - G_3(\pi)\} \\ & + G_5(\pi) - G_6(\pi, \pi^*), \end{aligned}$$

with

$$(2.17b) \qquad G_6(\pi, \pi^*) = G_6(\pi, \pi^*, X) = (\hat{\pi}\hat{\pi}^*)^{-1} \sum_i \sum_j c^{ij} \hat{\pi}_i^* \hat{\pi}_j.$$

The derivation of (2.16), which, as in the Appendix, is based on consideration of the approximate posterior characteristic function of $W(\pi, X, h)$ under the prior $\pi^*(\cdot)$, is omitted here to save space.

For $1 \leq i, j, r \leq p$, let

$$\begin{aligned} & V_i = D_i \log f(X_1; \theta), \qquad V_{ij} = D_i D_j \log f(X_1; \theta), \\ & V_{ijr} = D_i D_j D_r \log f(X_1; \theta), \\ & \mathcal{I}_{ij} = E_\theta(V_i V_j), \qquad L_{ij,r} = E_\theta(V_{ij} V_r), \qquad L_{ijr} = E_\theta(V_{ijr}). \end{aligned}$$

Note that $\mathcal{I}_{ij}$, $L_{ij,r}$, L_{ijr} are functions of θ and that the per observation information matrix at θ is given by $\mathcal{I} \equiv \mathcal{I}(\theta) = ((\mathcal{I}_{ij}))$ which is assumed to be positive definite. Let $\mathcal{I}^{-1} = ((\mathcal{I}^{ij}))$ and for $1 \leq i, j, r, s \leq p$, let $\mathcal{I}^{(1)}_{ijrs}$ be defined analogously to

(2.7) with c^{ij}, c^{rs}, etc. replaced by $\mathcal{I}^{ij}$, $\mathcal{I}^{rs}$, etc. Then from (2.6), (2.9), (2.17a), (2.17b), it can be seen that

$$H(X;\pi,\pi^*) = \bar{H}(\theta;\pi,\pi^*) + o(1), \tag{2.18}$$

where

$$\bar{H}(\theta;\pi,\pi^*) = \frac{1}{2}\{\bar{G}_1(\pi^*)-\bar{G}_1(\pi)\}+\frac{1}{6}\{\bar{G}_3(\pi^*)-\bar{G}_3(\pi)\}+\bar{G}_5(\pi)-\bar{G}_6(\pi,\pi^*), \tag{2.19}$$

with

$$\left\{\begin{array}{l} \bar{G}_1(\pi) = \pi^{-1}\sum_i\sum_j \mathcal{I}^{ij}\pi_{ij}, \qquad \bar{G}_1(\pi^*) = (\pi^*)^{-1}\sum_i\sum_j \mathcal{I}^{ij}\pi^*_{ij}, \\ \bar{G}_3(\pi) = \pi^{-1}\sum_i\sum_j\sum_r\sum_s L_{ijr}\pi_s\mathcal{I}^{(1)}_{ijrs}, \\ \bar{G}_3(\pi^*) = (\pi^*)^{-1}\sum_i\sum_j\sum_r\sum_s L_{ijr}\pi^*_s\mathcal{I}^{(1)}_{ijrs}, \\ \bar{G}_5(\pi) = \pi^{-2}\sum_i\sum_j \pi_i\pi_j\mathcal{I}^{ij}, \qquad \bar{G}_6(\pi,\pi^*) = (\pi\pi^*)^{-1}\sum_i\sum_j \pi^*_i\pi_j\mathcal{I}^{ij}, \\ \pi\equiv\pi(\theta), \qquad \pi^*\equiv\pi^*(\theta), \\ \pi_i\equiv\pi_i(\theta) = D_i\pi(\theta), \\ \pi^*_i = \pi^*_i(\theta) = D_i\pi^*(\theta) \qquad (1\le i\le p), \\ \pi_{ij} = \pi_{ij}(\theta) = D_iD_j\pi(\theta), \\ \pi^*_{ij} = \pi^*_{ij}(\theta) = D_iD_j\pi^*(\theta) \qquad (1\le i,j\le p). \end{array}\right. \tag{2.20}$$

As in GM, in consideration of (2.16), (2.18), for a fixed $\pi(\cdot)$,

$$P_\theta(\theta\in R_\pi(X)) = 1-\alpha-2(np)^{-1}z^2q_p(z^2)\Delta_\pi(\theta)+o(n^{-1}), \tag{2.21}$$

where $\Delta_\pi(\theta)$ is obtained by integrating $\bar{H}(\theta;\pi,\pi^*)$ by parts with respect to $\pi^*(\cdot)$ such that $\pi^*(\cdot)$ and its first partial derivatives vanish on the boundary of a rectangle containing θ, and then allowing $\pi^*(\cdot)$ to converge weakly to the degenerate measure at θ. By (2.19), (2.20), an explicit calculation shows that $\Delta_\pi(\theta)$, obtained as above, is given by

$$\begin{aligned} \Delta_\pi(\theta) &= \frac{1}{2}\sum_i\sum_j\{D_iD_j\mathcal{I}^{ij}-\pi^{-1}\mathcal{I}^{ij}\pi_{ij}\} \\ &\quad+\sum_i\sum_j[\pi^{-2}\pi_i\pi_j\mathcal{I}^{ij}+D_i(\pi^{-1}\pi_j\mathcal{I}^{ij})] \\ &\quad-\frac{1}{6}\sum_i\sum_j\sum_r\sum_s\{D_s(L_{ijr}\mathcal{I}^{(1)}_{ijrs})+\pi^{-1}\pi_sL_{ijr}\mathcal{I}^{(1)}_{ijrs}\} \\ &= \frac{1}{2}\pi^{-1}\sum_i\sum_j D_i(\mathcal{I}^{ij}\pi_j)+\frac{1}{2}\pi^{-1}\sum_i\sum_j D_i\{\pi(D_j\mathcal{I}^{ij})\} \end{aligned} \tag{2.22}$$

$$-\frac{1}{2}\pi^{-1}\sum_i\sum_j\sum_r\sum_s D_s(\pi L_{ijr}\mathcal{I}^{ij}\mathcal{I}^{rs})$$

$$=\frac{1}{2}\pi^{-1}\left[\sum_i\sum_j D_i(\mathcal{I}^{ij}\pi_j)+\sum_i\sum_j\sum_r\sum_s D_i(\pi\mathcal{I}^{ir}\mathcal{I}^{js}L_{rs,j})\right],$$

noting that $\mathcal{I}^{ij}=\mathcal{I}^{ji}$, L_{ijr} is invariant under permutation of subscripts, recalling the definition of $\mathcal{I}^{(1)}_{ijrs}$, and using the fact (cf. GM) that $D_j\mathcal{I}^{ij}=\sum_r\sum_s\mathcal{I}^{ir}\mathcal{I}^{js}\cdot(L_{rs,j}+L_{jrs})$ $(1\le i,j\le p)$. It can be seen that for $p=1$, the relations (2.21), (2.22) reduce to the corresponding expression in Peers (1968).

By (2.21), (2.22), the relation (2.15), ensuring frequentist validity, up to $o(n^{-1})$, of the HPD region $R_\pi(X)$, holds if and only if $\pi(\theta)$ satisfies the partial differential equation

$$\sum_i\sum_j D_i(\mathcal{I}^{ij}\pi_j)+\sum_i\sum_j\sum_r\sum_s D_i(\pi\mathcal{I}^{ir}\mathcal{I}^{js}L_{rs,j})=0. \tag{2.23}$$

The relation (2.23) gives the main result of this paper. It is interesting to note that (2.23) does not involve z^2 or α.

We now examine the extent to which (2.23) (or equivalently (2.15)) holds for Jeffreys' 'non-informative' prior (Welch and Peers (1963), Dawid (1983)) given by $\pi_0(\theta)\propto\{\det\mathcal{I}(\theta)\}^{1/2}$. It can be seen that (see GM)

$$D_j\pi_0(\theta)=-\frac{1}{2}\pi_0(\theta)\sum_r\sum_s(L_{rs,j}+L_{jrs})\mathcal{I}^{rs}.$$

Hence again noting that the L_{jrs} are invariant under permutation of subscripts, one can check that (2.23) holds for Jeffreys' prior (that is, Jeffreys' prior ensures the frequentist validity, up to $o(n^{-1})$, of HPD regions) if and only if

$$\sum_i\sum_j\sum_r\sum_s D_i\left[\{\det\mathcal{I}(\theta)\}^{1/2}\mathcal{I}^{ir}\mathcal{I}^{js}\left(L_{rs,j}-\frac{1}{2}L_{js,r}-\frac{1}{2}L_{jrs}\right)\right]=0, \tag{2.24}$$

a condition that holds in particular for location or scale families (cf. GM). Even outside location or scale families, the relation (2.23) (or equivalently (2.15)) can hold under Jeffreys' prior. For example, one can consider $\theta\in\mathcal{R}^1$ and $f(x;\theta)$ representing the bivariate normal density with zero means, variances 1 and $1+\theta^2$ and covariance θ.

For the exponential family with $f(x;\theta)$ of the form

$$f(x;\theta)=h(x)\exp\left\{\sum_{i=1}^p\psi^{(i)}(\theta)U_i(x)-A(\theta)\right\},$$

where $\theta=(\theta_1,\ldots,\theta_p)'$ and $E_\theta\{U_i(X)\}=\theta_i$ $(1\le i\le p)$, it can be seen that

$$\mathcal{I}_{ij}=\mathcal{I}_{ji}=D_i\psi^{(j)}(\theta)=D_j\psi^{(i)}(\theta),$$
$$\operatorname{cov}_\theta(U_i(X),U_j(X))=\mathcal{I}^{ij}\quad(1\le i,j\le p),$$
$$L_{rs,j}=D_rD_s\psi^{(j)}(\theta)\quad(1\le r,s,j\le p),$$

and hence one can check that (2.23) holds with $\pi(\theta) \propto \{\det \mathcal{I}(\theta)\}^{-1}$. Also, as in GM, for the exponential family with $f(x;\theta)$ of the form

$$f(x;\theta) = h(x) \exp\left\{\sum_{i=1}^{p} \theta_i U_i(x) - A(\theta)\right\},$$

the relation (2.23) holds with $\pi(\theta) = \rho$, where ρ is a positive constant.

Consider now the location-scale model with $p = 2$, $\theta = (\theta_1, \theta_2)'$ and $f(x;\theta)$ given by $f(x;\theta) = \theta_2^{-1} f^*(\theta_2^{-1}(x - \theta_1))$, where $-\infty < \theta_1 < \infty$, $\theta_2 > 0$. Then for each i, j, r, $\mathcal{I}_{ij} \propto \theta_2^{-2}$ and $L_{ij,r} \propto \theta_2^{-3}$, provided they exist. Hence assuming that $\mathcal{I}$ is positive definite for each θ, it can be seen that (2.23) holds with $\pi(\theta) \propto \theta_2^{-1}$.

We now give examples of two specific situations where (2.23) (or equivalently (2.15)) does not hold for Jeffreys' prior but other solutions to (2.23) are available.

Example 1. Let $p = 2$ and $f(x;\theta)$ represent the univariate normal density with mean θ_1 and variance θ_2 $(-\infty < \theta_1 < \infty,\ \theta_2 > 0)$. Then using some calculations in GM, the partial differential equation (2.23) reduces to

$$\theta_2 D_1^2 \pi + 2\theta_2^2 D_2^2 \pi - 2\theta_2 D_2 \pi - 6\pi = 0. \tag{2.25}$$

Here $\mathcal{I}_{11} = \theta_2^{-1}$, $\mathcal{I}_{21} = \mathcal{I}_{12} = 0$, $\mathcal{I}_{22} = \theta_2^{-2}/2$, and it can be seen that Jeffreys' prior $\pi_0(\theta) \propto \theta_2^{-3/2}$ does not satisfy (2.25). However, (2.25) has solutions. In particular, (2.25) holds with $\pi(\theta) \propto \theta_2^3$ or with $\pi(\theta) \propto \theta_2^{-1}$.

Example 2. Consider the ratio of normal means model (Cox and Reid (1987)) given by $f(x;\theta) = \phi_1(x^{(1)} - \mu_1)\phi_1(x^{(2)} - \mu_2)$, where $\phi_1(\cdot)$ represents the standard univariate normal density, $x = (x^{(1)}, x^{(2)})'$, $\theta = (\theta_1, \theta_2)'$, $\mu_1 = \theta_1\theta_2/(1 + \theta_1^2)^{1/2}$, $\mu_2 = \theta_2/(1 + \theta_1^2)^{1/2}$, and θ_1, $\theta_2 > 0$. Then μ_1, $\mu_2 > 0$, $\theta_1 = \mu_1/\mu_2$, $\theta_2 = (\mu_1^2 + \mu_2^2)^{1/2}$. It can be seen that here $\mathcal{I}_{11} = \theta_2^2/(1 + \theta_1^2)^2$, $\mathcal{I}_{21} = \mathcal{I}_{12} = 0$, $\mathcal{I}_{22} = 1$, $L_{11,1} = -2\theta_1\theta_2^2/(1 + \theta_1^2)^3$, $L_{12,2} = 0$, $L_{21,1} = \theta_2/(1 + \theta_1^2)^2$, $L_{22,2} = 0$. Hence it can be seen that Jeffreys' prior $\pi_0(\theta) \propto \theta_2/(1 + \theta_1^2)$ does not satisfy (2.23) but that (2.23) holds under $\pi(\theta) \propto \theta_2(1 + \theta_1^2)$.

Since the second term in the left-hand side of (2.23) is rather involved, it is difficult to present a general solution, if any, to (2.23) which holds for all parametric models satisfying the assumptions indicated earlier. This continues to be true even under global parametric orthogonality when the off-diagonal elements of $\mathcal{I}$ vanish identically in θ and (2.23) reduces to

$$\sum_i D_i(\mathcal{I}^{ii}\pi_i) + \sum_i \sum_j D_i(\pi \mathcal{I}^{ii}\mathcal{I}^{jj} L_{ij,j}) = 0.$$

However, as seen above, for many models of practical interest, like the location model, scale model, location-scale model, exponential model and so on, solutions to (2.23) are readily available.

Acknowledgements

The authors are thankful to the referees for very constructive suggestions.

Appendix

Derivation of (2.12). Writing $\xi = (-1)^{1/2}t$, the (approximate) posterior characteristic function of $W(\pi, X, h)$ under the prior $\pi(\cdot)$ is given by

$$\begin{aligned}
&\int \tilde{\pi}(h \mid X) \exp\{\xi W(\pi, X, h)\} dh \\
&= \int \Big[1 + n^{-1/2}(1-2\xi) \Big\{ T_{11}(\pi, h) + \frac{1}{6} T_{12}(h) \Big\} \\
&\quad + n^{-1} \Big\{ \xi G_5(\pi) - \xi(1-2\xi) T_{11}^2(\pi, h) \\
&\quad + \frac{1}{72}(1-2\xi)^2 T_{12}^2(h) - \frac{1}{72} G_4 + \frac{1}{6}(1-2\xi)^2 T_{11}(\pi, h) T_{12}(h) - \frac{1}{6} G_3(\pi) \\
&\quad + \frac{1}{24}(1-2\xi) T_{22}(h) - \frac{1}{24} G_2 + \frac{1}{2}(1-2\xi) T_{21}(\pi, h) - \frac{1}{2} G_1(\pi) \Big\} \Big] \\
&\quad \times e^{\xi h' C h} \phi(h; C^{-1}) dh + o(n^{-1}) \\
&= (1-2\xi)^{-p/2} \left[1 + n^{-1} \left(\frac{1}{1-2\xi} - 1 \right) \left(\frac{1}{24} G_2 + \frac{1}{72} G_4 \right) \right] + o(n^{-1}),
\end{aligned}$$

after a considerable algebra using (2.4)–(2.6), (2.9) and (2.10). Inverting the above, which can be justified as in Chandra ((1980), pp. 58–60) (see also Chandra and Ghosh ((1979), p. 32), Bickel and Ghosh ((1990), p. 1078)),

$$\begin{aligned}
\text{(A.1)}\quad & P^{\pi}(W(\pi, X, h(\theta)) \le \lambda(X) \mid X) \\
&= Q_p(\lambda(X)) + \frac{1}{24} n^{-1} \left(G_2 + \frac{1}{3} G_4 \right) \{ Q_{p+2}(\lambda(X)) - Q_p(\lambda(X)) \} \\
&\quad + o(n^{-1}),
\end{aligned}$$

where $Q_\nu(\cdot)$ is the cumulative distribution function of a central chi-square variate with ν degrees of freedom. By (2.11),

$$Q_\nu(\lambda(X)) = Q_\nu(z^2) + \frac{1}{12}(np)^{-1} z^2 \left(G_2 + \frac{1}{3} G_4 \right) q_\nu(z^2) + o(n^{-1}),$$

whence by (A.1), the relation (2.12) follows.

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. (1980). Asymptotic expansions and deficiency, Ph.D. Thesis, Indian Statistical Institute, Calcutta.

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.

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. (1983). Invariant prior distributions, *Encyclopedia of Statistical Sciences*, Vol. 4 (eds. S. Kotz and N. L. Johnson), 228–236, Wiley, New York.

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. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests, *J. Roy. Statist. Soc. Ser. B*, **54**, 867–875.

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, Indiana.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Part 2, Cambridge University Press, Cambridge.

Loh, W. L. (1988). On the coverage probability of some confidence sets for a multivariate normal mean, Tech. Report 88-62, Department of Statistics, Purdue University, Indiana.

Peers, H. W. (1965). On confidence sets 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.

Stein, C. (1985). On coverage probability of confidence sets based on a prior distribution, *Sequential Methods in Statistics*, Vol. 16, 485–514, Banach Center Publications, Warsaw.

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

Welch, B. 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.