On priors that match posterior and frequentist distribution functions

J.K. GHOSH and Rahul MUKERJEE*

Purdue University and Indian Statistical Institute and *Indian Institute of Management Calcutta*

Key words and phrases: Cumulative distribution function, Jeffreys's prior, parametric orthogonality, reference prior.
AMS 1985 subject classifications: 62F20.

ABSTRACT

In the multiparameter case, this paper characterizes priors so as to match, up to $o(n^{-\frac{1}{2}})$, the posterior joint cumulative distribution function (c.d.f.) of a posterior standardized version of the parametric vector with the corresponding frequentist c.d.f.

RÉSUMÉ

Cet article caractérise les distributions a priori telles que la fonction de répartition conjointe a posteriori du vecteur paramétrique centré réduit a posteriori soit égale, à $o(n^{-\frac{1}{2}})$ près, à la fonction de répartition fréquentiste correspondante, dans le cas de plusieurs paramètres.

1. INTRODUCTION

In recent years, there has been a revival of interest in the problem of characterizing priors ensuring frequentist validity, up to $o(n^{-\frac{1}{2}})$, of posterior quantiles. Among the early authors, Welch and Peers (1963) studied this problem in the one-parameter case. In the multiparameter case, Peers (1965) derived conditions for the frequentist validity of the posterior quantiles of each parameter, and later Stein (1985) and Tibshirani (1989) considered the problem with reference to a single linear parametric function of interest. In particular, Tibshirani (1989) showed that if the parameter of interest be one-dimensional, then elegant simplifications are possible making use of parametric orthogonality (Cox and Reid 1987).

It appears that, in the multiparameter case, these authors treated the parameters individually rather than jointly and based their results on such "marginal" analyses. This is true also of Peers (1965), who first treated the parameters separately and then combined the resulting conditions. On the other hand, if interest lies in all the parameters in a multiparameter setup, then a joint treatment of the parameters seems to be intuitively desirable. An obvious difficulty in such a joint treatment is that in the multiparameter case posterior "joint quantiles" are not well defined. However, if instead of looking for "joint quantiles" one considers the joint posterior cumulative distribution function (c.d.f.) of the parameters, then this difficulty no longer remains. Based on this consideration, this paper attempts to characterize priors so as to match, up to $o(n^{-\frac{1}{2}})$, the posterior joint

*This author's work was supported by a grant from the centre for Management and Development Studies, Indian Institute of Management Calcutta.

c.d.f. of a posterior standardized version of the parametric vector with the corresponding frequentist c.d.f. In many models of importance, our approach yields the same result as obtained by Peers (1965), showing that for such models the conditions in Peers (1965) are appropriate not only under individual but also under a joint treatment of the parameters. That this is not generally so has also been illustrated. Incidentally, our approach for the frequentist computations is different from and possibly simpler than that in Peers (1965).

It may be noted that, as in the derivation of reference priors (see, e.g., Berger and Bernardo 1992), in the present approach we need an ordering of the parameters in order of importance. However, after that the reference priors are essentially determined by successive minimization of an information functional, whereas we determine our priors by matching the posterior and frequentist c.d.f.'s as detailed in the next section. As noted in Tibshirani (1989), studies like the present one may be helpful in defining noninformative priors which could be useful for comparative purposes in a Bayesian analysis; see Lee (1989), Berger and Bernardo (1992), and Ghosh and Mukerjee (1991, 1992), among others, for further references in this general area.

2. MAIN RESULT

Let $\{X_i\}$, $i \geq 1$, be a sequence of i.i.d., possibly vector-valued random variables with common density $f(x;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)^{\mathsf{T}} \in \mathbf{R}^p$ or some open subset thereof. We make the assumptions in Johnson (1970). Let $\boldsymbol{\theta}$ have a prior density $\pi(\cdot)$ which is positive and twice continuously differentiable. In case $\pi(\cdot)$ is not proper as assumed by Johnson (1970), we shall require that there exist an $n_0(>0)$ such that for all $X_1,\ldots,X_{n_0}$, the posterior of $\boldsymbol{\theta}$ given $X_1,\ldots,X_{N_0}$ is proper. Let P be the joint probability measure of $\boldsymbol{\theta}$ and $\mathbf{X} = (X_1,\ldots,X_n)^{\mathsf{T}}$, where n is the sample size. All formal expansions for the posterior, as used here, are valid for sample points in a set S, defined along the lines of Bickel and Ghosh (1990, Section 2), with $P_{\boldsymbol{\theta}}$-probability $1+O(n^{-1})$ uniformly on compact sets of $\boldsymbol{\theta}$.

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

$$a_{ij} = \{D_iD_jl(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \qquad c_{ij} = -a_{ij}, \qquad a_{iju} = \{D_iD_jD_ul(\boldsymbol{\theta})\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

where D_i is the operator of partial differentiation with respect to θ_i, and $\hat{\boldsymbol{\theta}}$ is the maximum-likelihood estimator of $\boldsymbol{\theta}$ based on $X_1,\ldots,S_n$. The $p\times p$ matrix $\mathbf{C} = ((c_{ij}))$ is positive definite (p.d.) over S. For $1 \leq i \leq p-1$, let $\mathbf{c}_{(i)}$ denote the $(p-i)\times 1$ vector $(c_{i\,i+1},\ldots,c_{ip})^{\mathsf{T}}$, $\mathbf{C}_{(ii)}$ denote the $(p-i)\times(p-i)$ principal submatrix of $\mathbf{C}$ given by the last $p-i$ rows and columns of C, and $\bar{\mathbf{c}}_{(i)} = (\bar{c}_{i\,i+1},\ldots,\bar{c}_{ip})^{\mathsf{T}}$ denote the $(p-i)\times 1$ vector $\mathbf{C}_{(ii)}^{-1}\mathbf{c}_{(i)}$. Also, for $1 \leq i \leq p-1$, let $c^*_{ii} = c_{ii} - \mathbf{c}_{(i)}^{\mathsf{T}}\mathbf{C}_{(ii)}^{-1}\mathbf{c}_{(i)}$, and define $c^*_{pp} = c_{pp}$. Let $\bar{\mathbf{C}}$ be an upper triangular matrix with each diagonal element unity and (i,j)th element $-\bar{c}_{ij}$ $(1 \leq i < j \leq p)$. Then it is easy to see that $\mathbf{C} = \mathbf{C}^{*\mathsf{T}}\mathbf{C}^*$, where

$$\mathbf{C}^* = \operatorname{diag}\{(c^*_{11})^{\frac{1}{2}},\ldots,(c^*_{pp})^{\frac{1}{2}}\}\,(\bar{\mathbf{C}}^{\mathsf{T}})^{-1}. \tag{2.1}$$

We also define $\tilde{\mathbf{C}} = ((\tilde{c}_{ij})) = (\mathbf{C}^*)^{-1}$ and write $\mathbf{h} = (h_1,\ldots,h_p)^{\mathsf{T}} = n^{\frac{1}{2}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$.

We propose to characterize priors $\pi(\cdot)$ under which the posterior joint c.d.f. of the pivotal quantity

$$\mathbf{W} = \mathbf{W}(\boldsymbol{\theta},\mathbf{X}) = n^{\frac{1}{2}}\mathbf{C}^*(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \mathbf{C}^*\mathbf{h} \tag{2.2}$$

agrees, up to $o(n^{-\frac{1}{2}})$, with the frequentist joint c.d.f. In order to give an interpretation for $\mathbf{W}$, we note that (see, e.g., Ghosh and Mukerjee 1991) up to the first order of approximation the posterior mean vector and dispersion matrix of $\mathbf{h}$ equal $\mathbf{0}$ and $\mathbf{C}^{-1}$ respectively. Hence by (2.1), (2.2), up to the first order of approximation, the first element of $\mathbf{W}$ is the standardized version of h_1, and for $2 \le i \le p$, the ith element of $\mathbf{W}$ is the standardized regression residual of h_i on $h_1, \ldots, h_{i-1}$ in the posterior setup. If we are interested in each of $\theta_1, \ldots, \theta_p$ but the θ_i's are of decreasing (in i) importance, then the present formulation appears to be natural (cf. Berger and Bernardo 1992). This formulation depends on the ordering of the elements of $\boldsymbol{\theta}$, but, as seen later, in many situations of importance it can lead to priors which work for each such ordering.

Let $\mathbf{I} = \mathbf{I}(\boldsymbol{\theta}) = ((I_{ij}(\boldsymbol{\theta})))$ be the per-observation information matrix at $\boldsymbol{\theta}$, which is p.d. at each $\boldsymbol{\theta}$. With reference to $\mathbf{I}$, define $\mathbf{I}_{(i)}$, $\mathbf{I}_{(ii)}$, $\bar{\mathbf{I}}_{(i)}$, I^*_{ii} $(1 \le i \le p-1)$, I^*_{pp}, $\bar{\mathbf{I}}$, $\mathbf{I}^*$, $\tilde{\mathbf{I}} = ((\tilde{I}_{ij}))$ exactly as $\mathbf{c}_{(i)}$, $\mathbf{C}_{(ii)}$, $\bar{\mathbf{c}}_{(i)}$, c^*_{ii} $(1 \le i \le p-1)$, c^*_{pp}, $\bar{\mathbf{C}}$, $\mathbf{C}^*$, $\tilde{\mathbf{C}} = ((\tilde{c}_{ij}))$ were defined with reference to $\mathbf{C}$. Also, for $1 \le i, j, u \le p$, let $L_{iju} = \mathcal{E}_{\boldsymbol{\theta}}\{D_i D_j D_u \log f(X_1, \boldsymbol{\theta})\}$. We are now in a position to present our main result.

THEOREM 1. *Let the assumptions stated in the beginning of this section hold, and let $P^{\pi}(\cdot|\mathbf{X})$ denote the posterior probability measure of $\boldsymbol{\theta}$ given $\mathbf{X}$ under the prior $\pi(\cdot)$. Then under $\boldsymbol{\theta}$, the relation $P^{\pi}(\mathbf{W} \le \mathbf{t}|\mathbf{X}) = P_{\boldsymbol{\theta}}(\mathbf{W} \le \mathbf{t}) + o(n^{-\frac{1}{2}})$ holds for each $\boldsymbol{\theta}$ and each $\mathbf{t} = (t_1, \ldots, t_p)^{\mathsf{T}}$ $(\in \mathbb{R}^p)$ if and only if*

$$\sum_{i=1}^{p} D_i\{\tilde{I}_{ir}\pi(\boldsymbol{\theta})\} = 0 \qquad \text{for all } \boldsymbol{\theta} \qquad (1 \le r \le p). \tag{2.3}$$

Proof. Consider any fixed $\mathbf{t} = (t_1, \ldots, t_p)^{\mathsf{T}}$ $(\in \mathbb{R}^p)$ which is free from n and $X_1, \ldots, X_n$. Then by (2.2), the definition of $\tilde{\mathbf{C}}$, and Equation (2.2) in Ghosh and Mukerjee (1991) giving an expansion for the posterior density of h, we have

$$P^{\pi}(\mathbf{W} \le \mathbf{t}|X) = \Psi(\mathbf{t}) + n^{-\frac{1}{2}} G_{\pi}(\mathbf{X}, \mathbf{t}) + o(n^{-\frac{1}{2}}), \tag{2.4}$$

where

$$\Psi(\mathbf{t}) = \int \cdots \prod_{i=1}^{p} \phi(w_i)\, dw_p \cdots dw_1, \tag{2.5a}$$

$$G_{\pi}(\mathbf{X}, \mathbf{t}) = \sum_i \sum_r \tilde{c}_{ir}\hat{\pi}_i\hat{\pi}^{-1}\Psi_{1r}(t) + \tfrac{1}{6}\sum_i \sum_j \sum_u \sum_r \sum_s \sum_v \tilde{c}_{ir}\tilde{c}_{js}\tilde{c}_{uv}a_{iju}\Psi_{3rsv}(t), \tag{2.5b}$$

$$\Psi_{1r}(\mathbf{t}) = \int \cdots \int w_r \prod_{i=1}^{p} \phi(w_i)\, dw_p \cdots dw_1 \qquad (1 \le r \le p), \tag{2.5c}$$

$$\Psi_{3rsv}(\mathbf{t}) = \int \cdots \int w_r w_s w_v \prod_{i=1}^{p} \phi(w_i)\, dw_p \cdots dw_1 \qquad (1 \le r, s, v \le p), \tag{2.5d}$$

$\phi(\cdot)$ denotes the standard univariate normal density, $\hat{\pi} = \pi(\hat{\boldsymbol{\theta}})$, $\hat{\pi}_i = \pi_i(\hat{\boldsymbol{\theta}})$ with $\pi_i(\boldsymbol{\theta}) = D_i\pi(\boldsymbol{\theta})$ $(1 \le i \le p)$, each summation is over the range 1 to p, and the integrals in (2.5a, c, d) are over the region $\{(w_1, \ldots, w_p) : w_i \le t_i, 1 \le i \le p\}$.

From (2.5b), note that under $\boldsymbol{\theta}$,

$$G_\pi(\mathbf{X}, \mathbf{t}) = \sum_i \sum_r \tilde{I}_{ir}\pi_i(\boldsymbol{\theta})\{\pi(\boldsymbol{\theta})\}^{-1}\Psi_{1r}(\mathbf{t})$$
$$+ \tfrac{1}{6}\sum_i \sum_j \sum_u \sum_r \sum_s \sum_v \tilde{I}_{ir}\tilde{I}_{js}\tilde{I}_{uv}L_{iju}\Psi_{3rsv}(\mathbf{t}) + o(1). \qquad (2.6)$$

In consideration of (2.4), (2.6), proceeding as in Section 2 of Ghosh and Mukerjee (1991), one gets the frequentist probability

$$P_{\boldsymbol{\theta}}(\mathbf{W} \leq \mathbf{t}) = \Psi(\mathbf{t}) + n^{-\frac{1}{2}}\Big\{ -\sum_i \sum_r (D_i\tilde{I}_{ir})\Psi_{1r}(\mathbf{t})$$
$$+ \tfrac{1}{6}\sum_i \sum_j \sum_u \sum_r \sum_s \sum_v \tilde{I}_{ir}\tilde{I}_{js}\tilde{I}_{uv}L_{iju}\Psi_{3rsv}(\mathbf{t})\Big\} + o(n^{-\frac{1}{2}}).$$

Hence by (2.4), (2.6), the result follows. Q.E.D.

Remark 1. In particular, suppose global parametric orthogonality holds (see Cox and Reid 1987), i.e., let $I_{ij} = 0$, identically in $\boldsymbol{\theta}$, for each $i \neq j$. Then $\bar{\mathbf{I}}$ equals the $p \times p$ identity matrix and $\tilde{\mathbf{I}} = \mathrm{diag}(I_{11}^{-\frac{1}{2}}, \ldots, I_{pp}^{-\frac{1}{2}})$. Hence (2.3) reduces to

$$D_r\{I_{rr}^{-\frac{1}{2}}\pi(\boldsymbol{\theta})\} = 0 \qquad \text{for all } \boldsymbol{\theta} \qquad (1 \leq r \leq p). \qquad (2.7)$$

By (2.7), under global parametric orthogonality, a solution to (2.3), if existent, will work for all possible orderings of the θ_i's. When satisfied, (2.7) also leads to the frequentist validity, up to $o(n^{-\frac{1}{2}})$, of the posterior quantiles of each θ_i — *vide* Tibshirani (1989).

Remark 2. More generally, when global parametric orthogonality may not hold, $\bar{\mathbf{I}}$ will be a triangular matrix with elements dependent on the ordering of the θ_i's. Hence, in such situations, (2.3) itself may depend on this ordering. Incidentally, note that neither (2.3) nor (2.7) involves $\mathbf{t}$.

Remark 3. For $p = 1$, (2.3) leads to Jeffreys's prior; this is in agreement with Welch and Peers (1963). In the multiparameter case as well, there are important situations (see Examples 1, 2 below) where (2.3) leads to Jeffreys's prior, but this does not happen always (see Example 3). In fact, in the multiparameter case, depending upon the model, the system of partial differential equations (2.3) may not have any solution at all (see Example 4). Following Section 4 of Peers (1965), we note in this connection that (2.3) can be uniquely solved, as a system of linear equations, for $H_i(\boldsymbol{\theta}) = D_i \log \pi(\boldsymbol{\theta})$ $(1 \leq i \leq p)$, and that (2.3) will admit a solution if and only if $D_jH_i(\boldsymbol{\theta}) = D_iH_j(\boldsymbol{\theta})$ for each i, j $(i \neq j)$.

3. SOME EXAMPLES

In Examples 1 and 3 below, which are of somewhat general nature, we assume, as usual, the existence and p.d.-ness of $\mathbf{I}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.

Example 1. If $f(\mathbf{x}; \boldsymbol{\theta})$ represents the location or the scale model, as detailed in Examples 3.1 and 3.2 respectively of Ghosh and Mukerjee (1991), then calculations similar to theirs reveal that the respective Jeffreys priors solve (2.3).

EXAMPLE 2. Consider a version of the exponential regression model with

$$f(\mathbf{x};\, \boldsymbol{\theta}) = \prod_{i=1}^{\tau} [\theta_2^{-1} e^{-\theta_1 z_i} \exp\{-x^{(i)}\theta_2^{-1} e^{-\theta_1 z_i}\}], \qquad x^{(1)}, \ldots, x^{(\tau)} > 0,$$

with $\mathbf{x} = (x^{(1)}, \ldots, x^{(\tau)})^{\mathsf{T}}$, and $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathsf{T}}$, $-\infty < \theta_1 < \infty$, $\theta_2 > 0$, τ (≥ 2) being a fixed positive integer, and $z_1, \ldots, z_\tau$ being known constants, not all zeros, satisfying $z_1 + \cdots + z_\tau = 0$. Then, as noted in Cox and Reid (1987), global parametric orthogonality holds and $\mathbf{I}(\boldsymbol{\theta}) = \mathrm{diag}(\sum_{i=1}^{\tau} z_i^2, \theta_2^{-2}\tau)$. Hence (2.3) reduces to (2.7) and is satisfied by Jeffreys's prior, namely $\pi_0(\theta) \propto \theta_2^{-1}$.

EXAMPLE 3. Consider the location-scale family with $f(x;\, \boldsymbol{\theta})$ of the form $f(x;\, \boldsymbol{\theta}) = \theta_2^{-1} g(\theta_2^{-1}(x - \theta_1))$, where $-\infty < \theta_1 < \infty$, $\theta_2 > 0$, and $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathsf{T}}$. Then $\mathbf{I}(\boldsymbol{\theta}) = \theta_2^{-2}\mathbf{Q}$, where $\mathbf{Q}$ is a 2×2 p.d. matrix with elements free from $\boldsymbol{\theta}$. If one writes $\mathbf{Q} = ((q_{ij}))$ and evaluates $\tilde{\mathbf{I}}$ explicitly, then (2.3) becomes

$$D_1\{\theta_2 q_{11.2}^{-\frac{1}{2}} \pi(\boldsymbol{\theta})\} - D_2\{\theta_2 q_{12} q_{22}^{-1} q_{11.2}^{\frac{1}{2}} \pi(\boldsymbol{\theta})\} = 0, \qquad D_2\{\theta_2 q_{22}^{-\frac{1}{2}} \pi(\boldsymbol{\theta})\} = 0,$$

where $q_{11.2} = q_{11} - q_{12}^2 q_{22}^{-1}$. A solution to the above is given by $\pi^*(\boldsymbol{\theta}) \propto \theta_2^{-1}$. Also note that Jeffreys's prior, given by $\pi_0(\boldsymbol{\theta}) \propto \theta_2^{-2}$, is not a solution.

It is interesting to note that the solutions to (2.3) indicated in Examples 1–3 work under all possible orderings of the parameters. We now illustrate a situation where (2.3) has no solution.

EXAMPLE 4. Consider the model given by

$$f(\mathbf{x};\, \boldsymbol{\theta}) = \theta_2^{-\frac{1}{2}} \phi(\theta_2^{-\frac{1}{2}}(x^{(1)} - \theta_1)) \exp\{-(\theta_1^3 + x^{(2)} e^{-\theta_1^3})\},$$

where $\mathbf{x} = (x^{(1)}, x^{(2)})^{\mathsf{T}}$, $-\infty < x^{(1)} < \infty$, $x^{(2)} > 0$, $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathsf{T}}$, $-\infty < \theta_1 < \infty$, $\theta_2 > 0$, and, as before, $\phi(\cdot)$ is the standard univariate normal density. Here global parametric orthogonality holds and $\mathbf{I}(\boldsymbol{\theta}) = \mathrm{diag}(\theta_2^{-1} + 9\theta_1^4, \frac{1}{2}\theta_2^{-2})$. Therefore, by Remark 1, (2.3) has a solution if and only if (2.7) has a solution, which happens only if $(I_{11}/I_{22})^{\frac{1}{2}}$ is of the form $\{d_1(\theta_1)/d_2(\theta_2)\}$, where $d_1(\theta_1)$ [$d_2(\theta_2)$] may involve θ_1 [θ_2] but not θ_2 [θ_1]. Since this is not the case here, the nonexistence of a solution to (2.3) follows.

EXAMPLE 5. Consider the three-class multinomial model given by

$$f(\mathbf{x};\, \boldsymbol{\theta}) = \theta_1^{x^{(1)}} \theta_2^{x^{(2)}} (1 - \theta_1 - \theta_2)^{1 - x^{(1)} - x^{(2)}}, \qquad x^{(1)}, x^{(2)} = 0 \text{ or } 1, \quad x^{(1)} + x^{(2)} \leq 1,$$

where $\mathbf{x} = (x^{(1)}, x^{(2)})^{\mathsf{T}}$, $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathsf{T}}$, $\theta_1 > 0$, $\theta_2 > 0$, and $\theta_1 + \theta_2 < 1$. Explicit calculations show that the proper prior $\pi^*(\boldsymbol{\theta}) \propto \{\theta_1\theta_2(1 - \theta_1)(1 - \theta_1 - \theta_2)\}^{-\frac{1}{2}}$ solves (2.3), while Jeffreys's prior given by $\pi_0(\boldsymbol{\theta}) \propto \{\theta_1\theta_2(1 - \theta_1 - \theta_2)\}^{-\frac{1}{2}}$ fails to do so. It may, however, be remarked that the findings in this example may not have a rigorous implication, as the model is not represented by a density.

4. CONNECTION WITH PEERS'S RESULTS AND REFERENCE PRIORS

As indicated in Remark 3, one can express (2.3) as

$$\mathbf{H}(\boldsymbol{\theta}) = -(\tilde{\mathbf{I}}^{\mathsf{T}})^{-1}\mathbf{b}_1(\boldsymbol{\theta}), \tag{4.1}$$

Considering two different choices of the sequence of subsets of the parameter space in their algorithm, Berger and Bernardo (1989) obtained the reference priors $\pi_s(\boldsymbol{\theta}) = \theta_1^{\frac{1}{2}}(1+\theta_2^{-4})^{\frac{1}{2}}$ and $\pi_r(\theta) = \theta_1^{-\frac{1}{2}}(1+\theta_2^{-4})^{\frac{1}{2}}$. If one writes $\tilde{\mathbf{I}}$ and hence (2.3) explicitly from the expression for I, then it can be seen that $\pi_s(\boldsymbol{\theta})$ satisfies (2.3) while $\pi_r(\boldsymbol{\theta})$ does not. Note that the reference prior $\pi_s(\boldsymbol{\theta})$ was recommended by Berger and Bernardo (1989), who matched posterior and frequentist coverage probabilities numerically, and also by Tibshirani (1989), who checked the frequentist validity of the posterior quantiles of θ_1 via an orthogonal parametrization. This example shows that $\pi_s(\boldsymbol{\theta})$ works also when interest lies in matching, up to $o(n^{-\frac{1}{2}})$, the posterior and frequentist joint c.d.f.'s of a posterior standardized version of $\boldsymbol{\theta}$.

Example 8. Consider the balanced-variance-components model treated in Berger and Bernardo (1990) and given by $Y_{ij} = \theta_1 + A_i + e_{ij}$ $(1 \le i \le k, 1 \le j \le \nu)$, where k and ν are fixed positive integers (≥ 2), θ_1 is the general mean, the A_i's are each normal with mean zero and variance θ_2, and the e_{ij}'s are each normal with mean zero and variance θ_3, the A_i's and the e_{ij}'s being all independent. Considering asymptotics based on independent replications of the above setup, it can be checked after lengthy algebra, which we omit here to save space, that for each of the orderings $(\theta_1, \theta_3, \theta_2)$, $(\theta_3, \theta_2, \theta_1)$, $(\theta_3, \theta_1, \theta_2)$ of the parameters, the reference prior reported in Berger and Bernardo (1990), namely $\pi^*(\boldsymbol{\theta}) \propto \{\theta_3(\nu\theta_2+\theta_3)\}^{-1}$, satisfies (2.3), and that this is not the case under other orderings of the parameters.

In the above examples, we checked whether specific reference priors satisfy (2.3) or not. One might well wish to investigate the reverse problem, namely, whether or not a solution to (2.3), if existent, will always be a reference prior for some choice of the sequence of subsets of the parameter space in the Berger-Bernardo algorithm (Berger and Bernardo 1989, 1992). This appears to be an interesting but hard question to which we do not know an answer at this stage.

ACKNOWLEDGEMENT

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

REFERENCES

Berger, J.O., and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, 84, 200–207.

Berger, J.O., and Bernardo, J.M. (1990). Reference priors in a variance components problem. *Proc. Indo-USA Workshop on Bayesian Analysis in Statistics and Econometrics* (P. Goel, ed.), to appear.

Berger, J.O., and Bernardo, J.M. (1992). On the development of the reference prior method. *Bayesian Statistics 4*. Oxford Univ. Press, to appear.

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

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

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*, to appear.

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.

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.

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

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.

Received 10 September 1991
Revised 5 May 1992
Accepted 11 August 1992

Department of Statistics
Purdue University
West Lafayette, Indiana
U.S.A. 47907

Indian Institute of Management
Joka, Diamond Harbour Road
P.O. Box 16757, Alipore Post Office
Calcutta 700 027
India