ON A PROPERTY OF PROBABILITY MATCHING PRIORS: MATCHING THE ALTERNATIVE COVERAGE PROBABILITIES

Rahul Mukerjee
Indian Institute of Management
Post Box No 16757, Calcutta 700 027, India
rmuk1@hotmail.com

N. Reid
Department of Statistics
University of Toronto
Toronto, Canada M5S 3G3
reid@utstat.utoronto.ca

Summary

In frequentist inference based on confidence sets, both the true coverage and the probability for a confidence set to include an alternative value of the parameter of interest are important. Thus if probability matching priors also match such alternative coverage probabilities there is perhaps a stronger justification for calling them noninformative. Considering contiguous alternatives, we obtain the relevant necessary and sufficient conditions. In particular, with n denoting the sample size, it is seen that probability matching priors up to $o(n^{-1/2})$ also match the alternative coverage probabilities up to that order while this is not necessarily the case with probability matching priors up to $o(n^{-1})$.

Some key words: Contiguous alternative; First order; Jeffreys' prior; Noninformative prior; parametric orthogonality; Second order.

1. Introduction

In recent years, there has been considerable interest in the characterisation of priors ensuring, under suitable regularity conditions, approximate frequentist validity of posterior credible sets; see Reid (1995) and Kass & Wasserman (1996) for reviews. In particular, priors $\pi(\cdot)$ for which the relation

$$P_\theta\{\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, X)\} = 1 - \alpha + o(n^{-r/2}) \ , \tag{1.1}$$

holds for $r = 1$ or 2 and for each $\alpha(0 < \alpha < 1)$, have received much attention. Here n is the sample size, $\theta = (\theta_1, \ldots, \theta_p)'$ is an unknown parameter vector, θ_1 is the one-dimensional parameter of interest, P_θ is the frequentist probability measure under θ, and $\theta_1^{(1-\alpha)}(\pi, X)$ is the $(1-\alpha)$th posterior quantile of θ_1 given the data X. Priors satisfying (1.1) for $r = 1$ or 2 are called first- or second-order probability matching priors respectively. For $p = 1$, i.e. in the absence of nuisance parameters, Welch & Peers (1963) characterised Jeffreys' prior as first-order probability matching and explored model conditions under which it is also second-order probability matching. The corresponding problems for $p \geq 2$ have been investigated by, among others, Peers (1965), Tibshirani (1989), Nicolaou (1993) and Mukerjee & Dey (1993). As noted by Tibshirani (1989), studies of this kind help in getting accurate frequentist confidence sets and defining noninformative priors.

From the frequentist point of view, however, the probability for a confidence set to include an alternative value of the interest parameter is as important as that of true coverage (Lehmann, 1986, Ch.3). Such an alternative coverage probability indicates how selective a confidence set is, and its complement is linked with the power of the associated test. We consider in this paper how far a prior satisfying (1.1) also matches $P_\theta\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)\}$ with the corresponding posterior probability, up to the same order of approximation and for each δ and α, where I^{11} is the $(1,1)$ element of the inverse of the per observation expected information matrix at θ and the scalar δ is free from n, θ and X. If a prior satisfying (1.1) also matches the alternative coverage probabilities in the above sense, then there is a stronger justification for calling it noninformative in so far as agreement with a frequentist is concerned.

After presenting the preliminaries in §2, we obtain necessary and sufficient conditions for matching true and alternative coverage probabilities, up to the first and second orders of approximation, in §3. It is seen that first-order probability matching priors also match alternative coverage probabilities up to that order while that is not necessarily the case at the second order of approximation. Finally, some illustrative examples, including one where the present study helps in narrowing down the class of second-order probability matching priors, are presented in §4.

2. Notation and preliminary 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 the parameter vector $\theta = (\theta_1, \ldots, \theta_p)'$ belongs to $\mathcal{R}^p$ or some open subset thereof and θ_1 is the parameter of interest. Along the lines of Mukerjee & Dey (1993), we work essentially under the assumptions of Johnson (1970) and also need the Edgeworth assumptions of Bickel & Ghosh (1990, p. 1078). All formal expansions for the posterior, as used here, are valid for sample points in a set S with $P_\theta-$probability $1 + o(n^{-1})$ uniformly over compact sets of θ. The set S may be defined following Bickel & Ghosh (1990, Section 2).

Let $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)'$ be the maximum likelihood estimator of θ based on $X = (X_1, \ldots, X_n)'$, $\ell(\theta) = n^{-1} \sum_{i=1}^{n} \log f(X_i; \theta)$ and, with $D_j \equiv \partial/\partial\theta_j$, let

$$a_{jr} = \{D_j D_r \ell(\theta)\}_{\theta=\hat{\theta}} \ , \ a_{jrs} = \{D_j D_r D_s \ell(\theta)\}_{\theta=\hat{\theta}} \ , \ c_{jr} = -a_{jr} \ ,$$

$$V_j = D_j \log f(X_1; \theta) \ , \ V_{jr} = D_j D_r \log f(X_1; \theta) \ , \ V_{jrs} = D_j D_r D_s \log f(X_1; \theta) \ ,$$

$$I_{jr} = E_\theta(V_j V_r) \ , \ L_{j,r,s} = E_\theta(V_j V_r V_s) \ , \ L_{j,rs} = E_\theta(V_j V_{rs}) \ , \ L_{jrs} = E_\theta(V_{jrs}) \ .$$

The matrix $C = (c_{jr})$ is positive definite over S. Let $C^{-1} = (c^{jr})$ and $k^{jr} = c^{jr} - (c^{j1}c^{r1}/c^{11})$. Similarly, let $I = (I_{jr})$ be the per observation Fisher information matrix at θ, $I^{-1} = (I^{jr})$, $\tau^{jr} = I^{j1}I^{r1}/I^{11}$ and $\sigma^{jr} = I^{jr} - \tau^{jr}$. The quantities I_{jr}, I^{jr}, τ^{jr}, σ^{jr}, $L_{j,r,s}$, etc. are all functions of θ.

Let θ have a prior density $\pi(\cdot)$ which is positive and thrice continuously differentiable for all θ. We define $\pi_j(\theta) = D_j\pi(\theta)$, $\pi_{jr}(\theta) = D_j D_r \pi(\theta)$, $\hat{\pi} = \pi(\hat{\theta})$, $\hat{\pi}_j = \pi_j(\hat{\theta})$, $\hat{\pi}_{jr} = \pi_{jr}(\hat{\theta})$. Similarly, let $\beta(\theta) = (I^{11})^{1/2}$, $\beta_j(\theta) = D_j\beta(\theta)$, $\beta_{jr}(\theta) = D_j D_r \beta(\theta)$, $\hat{\beta} = \beta(\hat{\theta})$, $\hat{\beta}_j = \beta_j(\hat{\theta})$, $\hat{\beta}_{jr} = \beta_{jr}(\hat{\theta})$.

As indicated in §1, the quantity $\theta_1 + \delta(I^{11}/n)^{1/2}$, where δ does not involve n, θ or X, will play a crucial role in the present work. To facilitate the presentation, we consider a related quantity

$$\lambda = (c^{11})^{-1/2}[n^{1/2}\{\theta_1 + \delta(I^{11}/n)^{1/2} - \hat{\theta}_1\} - \delta\hat{\beta}] \ , \tag{2.1}$$

and, recalling the definition of $\beta(\theta)$, note that

$$\lambda = (c^{11})^{-1/2}\{h_1 + \delta(n^{-1/2}h_j\hat{\beta}_j + \frac{1}{2}n^{-1}h_j h_r \hat{\beta}_{jr})\} + o(n^{-1}), \tag{2.2}$$

where $h = (h_1, \ldots, h_p)' = n^{1/2}(\theta - \hat{\theta})$. In (2.2) and elsewhere, unless otherwise specified, we follow the summation convention with sums ranging from 1 to p. Ghosh & Mukerjee

(1991) gave an expansion, up to $o(n^{-1})$, for the posterior density of h under the prior $\pi(\cdot)$. From this, using (2.2), one can consider the joint posterior density of $(\lambda, h_2, \ldots, h_p)'$ and then integrate out $h_2, \ldots, h_p$ to obtain

$$\begin{aligned}\tilde{\pi}(\lambda|X) = \phi(\lambda)\big[1 &+ n^{-1/2}\{A_1(\pi, X)\lambda + \delta A_2(X)(\lambda^2 - 1) + A_3(X)\lambda^3\} \\ &+ n^{-1}\{\delta A_4(\pi, X)\lambda + A_5(\pi, \delta, X)(\lambda^2 - 1) + \delta A_6(\pi, X)\lambda^3 \\ &\quad + A_7(\pi, \delta, X)(\lambda^4 - 3) + \delta A_8(X)\lambda^5 + A_9(X)(\lambda^6 - 15)\}\big] \\ &+ o(n^{-1}) \ , \end{aligned} \tag{2.3}$$

where $\phi(\cdot)$ is the standard univariate normal density,

$$A_1(\pi, X) = A_{11}(\pi, X) + A_{12}(X) \ ,$$

$$A_{11}(\pi, X) = (c^{11})^{-1/2}(\hat{\pi}_i/\hat{\pi})c^{i1} \ , \ A_{12}(X) = \frac{1}{2}(c^{11})^{-1/2}a_{ijr}k^{ij}c^{r1} \ , \tag{2.4}$$

$$A_2(X) = (c^{11})^{-1}\hat{\beta}_i c^{i1} \ , \ A_3(X) = \frac{1}{6}(c^{11})^{-3/2}a_{ijr}c^{i1}c^{j1}c^{r1} \ , \tag{2.5}$$

$$A_4(\pi, X) = A_{41}(\pi, X) + A_{42}(X) \ , \ A_5(\pi, \delta, X) = A_{51}(\pi, X) + A_{52}(X) + \delta^2 A_{53}(X) \ ,$$

$$A_6(\pi, X) = A_{61}(\pi, X) + A_{62}(X) \ , \ A_7(\pi, \delta, X) = A_{71}(\pi, X) + A_{72}(X) + \delta^2 A_{73}(X) \ ,$$

$$A_{41}(\pi, X) = (c^{11})^{-1/2}\hat{\beta}_i(\hat{\pi}_r/\hat{\pi})c^{ir} - 3A_{11}(\pi, X)A_2(X) \ , \tag{2.6}$$

$$A_{51}(\pi, X) = \frac{1}{2}(\hat{\pi}c^{11})^{-1}c^{i1}c^{j1}(\hat{\pi}_{ij} + \hat{\pi}_s a_{ijr}k^{rs}) + A_{11}(\pi, X)A_{12}(X) \ , \tag{2.7}$$

$$A_{61}(\pi, X) = A_{11}(\pi, X)A_2(X), \ , \ A_{71}(\pi, X) = A_{11}(\pi, X)A_3(X) \ ,$$

and $A_{42}(X)$, $A_{52}(X)$, $A_{53}(X)$, $A_{62}(X)$, $A_{72}(X)$, $A_{73}(X)$, $A_8(X)$, $A_9(X)$, like $A_{12}(X)$, $A_2(X)$ and $A_3(X)$, are at most of order $O(1)$ and do not involve $\pi(\cdot)$ or δ; the detailed expressions for these quantities will not be needed in the subsequent development. The derivation of (2.3) involves considerable algebra which is omitted here to save space. Incidentally, we verified that inversion of the approximate posterior characteristic function of λ also produces the same result as in (2.3).

Let $P^{\pi}\{\cdot|X\}$ be the posterior probability measure under $\pi(\cdot)$ given X, and

$$\theta_1^{(1-\alpha)}(\pi, X) = \hat{\theta}_1 + (c^{11}/n)^{1/2}(z + n^{-1/2}u_1 + n^{-1}u_2) \ , \tag{2.8}$$

where z is the $(1-\alpha)$th quantile of a standard normal variate and

$$u_1 = u_1(\alpha, \pi, X) = A_1(\pi, X) + (z^2 + 2)A_3(X), \tag{2.9}$$

$$\begin{aligned} u_2 = u_2(\alpha, \pi, X) &= 2zA_3(X)u_1 - \frac{1}{2}zu_1^2 + zA_5(\pi, 0, X) \\ &+ (z^3 + 3z)A_7(\pi, 0, X) + (z^5 + 5z^3 + 15z)A_9(X) \ . \end{aligned} \tag{2.10}$$

Note that $\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, X)$ if and only if $y \leq z + n^{-1/2}u_1 + n^{-1}u_2$, where $y = (n/c^{11})^{1/2}$ $(\theta_1 - \hat{\theta}_1)$. Since, by (2.1), λ reduces to y when $\delta = 0$, it now follows from (2.3), with $\delta = 0$, and (2.9), (2.10) that

$$P^{\pi}\{\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, X)|X\} = 1 - \alpha + o(n^{-1}) \ , \tag{2.11}$$

so that $\theta_1^{(1-\alpha)}(\pi, X)$ represents the $(1-\alpha)$th posterior quantile of θ_1, under $\pi(\cdot)$, up to the order of approximation $o(n^{-1})$; compare Mukerjee & Dey (1993).

3. Main results

We now proceed to match $P_\theta\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)\}$ with its posterior counterpart $P^{\pi}\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)|X\}$. The latter is stochastic in a frequentist set-up and its expectation, under θ, is comparable with the former. Thus we characterise priors $\pi(\cdot)$ which match true as well as alternative coverage probabilities in the sense that

$$
\begin{aligned}
&P_\theta\{\theta_1 + \delta(I^{11}/n)^{1/2} \le \theta_1^{(1-\alpha)}(\pi, X)\} \\
&- E_\theta\big[P^\pi\{\theta_1 + \delta(I^{11}/n)^{1/2} \le \theta_1^{(1-\alpha)}(\pi, X)|X\}\big] = o(n^{-r/2}) \ , \qquad (3.1)
\end{aligned}
$$

for each δ and α, where $r = 1$ or 2. This is done through Steps I-III below which eventually yield the necessary and sufficient conditions (3.15) and (3.16) representing the main result of this paper. Consideration of an auxiliary prior $\overline{\pi}(\cdot)$, as in Step I, facilitates the frequentist computation. In particular, it keeps the algebra relatively simple by avoiding an explicit Edgeworth expansion.

Step I. Consider an auxiliary prior $\overline{\pi}(\cdot)$ and note that the posterior density of λ, under $\overline{\pi}(\cdot)$, is as in (2.3) with $\pi(\cdot)$ replaced by $\overline{\pi}(\cdot)$. Hence, if we write $g = \hat{\beta}(c^{11})^{-1/2}$, it follows from (2.1) and (2.8)-(2.10), after some simplification, that

$$
\begin{aligned}
&P^{\overline{\pi}}\{\theta_1 + \delta(I^{11}/n)^{1/2} \le \theta_1^{(1-\alpha)}(\pi, X)|X\} \\
&= P^{\overline{\pi}}\{\lambda \le z - \delta g + n^{-1/2}u_1(\alpha, \pi, X) + n^{-1}u_2(\alpha, \pi, X)|X\} \\
&= \Phi(z-\delta g) + n^{-1/2}\phi(z-\delta g)\{A_{11}(\pi, X) - A_{11}(\overline{\pi}, X) + W_1(z, \delta, X)\} \\
&\quad + n^{-1}\phi(z-\delta g)\big(\{A_{11}(\pi, X) - A_{11}(\overline{\pi}, X)\}[\delta\{(z-\delta g)^2 - 1\}A_2(X) - (z-\delta g)A_{11}(\pi, X) \\
&\qquad + \{z - 2(z^2-1)\delta g + 3z\delta^2 g^2 - \delta^3 g^3\}A_3(X)] \\
&\qquad + \delta g[A_{50}(\overline{\pi}, X) - \frac{1}{2}\{A_{11}(\pi, X)\}^2 + 3A_{11}(\overline{\pi}, X)A_3(X)] \\
&\qquad + z\{A_{50}(\pi, X) - A_{50}(\overline{\pi}, X)\} - \delta A_{40}(\overline{\pi}, X) + W_2(z, \delta, X)\big) + o(n^{-1}) \ , \qquad (3.2)
\end{aligned}
$$

where Φ is the standard normal cumulative distribution function,

$$A_{40}(\pi, X) = A_{41}(\pi, X) + 3A_{11}(\pi, X)A_2(X) \ ,$$

$$A_{50}(\pi, X) = A_{51}(\pi, X) - A_{11}(\pi, X)A_{12}(X) \ , \tag{3.3}$$

$$W_1(z, \delta, X) = \delta g(2z - \delta g)A_3(X) - \delta(z - \delta g)A_2(X)$$

and like, $W_1(z, \delta, X)$, the quantity $W_2(z, \delta, X)$ is at most of order $O(1)$, does not involve $\pi(\cdot)$ or $\overline{\pi}(\cdot)$, and vanishes if $\delta = 0$.

Step II. Let

$$H_1(\pi, \theta) = (I^{11})^{-1/2}\{\pi_i(\theta)/\pi(\theta)\}I^{i1} \ , \ H_2(\theta) = (I^{11})^{-1}\beta_i(\theta)I^{i1} \ , \tag{3.4}$$

$$H_3(\theta) = \frac{1}{6}(I^{11})^{-3/2}L_{ijr}I^{i1}I^{j1}I^{r1} \ , \tag{3.5}$$

$$H_4(\pi, \theta) = (I^{11})^{-1/2}\beta_i(\theta)\{\pi_r(\theta)/\pi(\theta)\}I^{ir} \ , \tag{3.6}$$

$$H_5(\pi, \theta) = \frac{1}{2}\{\pi(\theta)I^{11}\}^{-1}I^{i1}I^{j1}\{\pi_{ij}(\theta) + \pi_s(\theta)L_{ijr}\sigma^{rs}\} \ . \tag{3.7}$$

Since $E_\theta(g) = 1 + o(n^{-1/2})$, by (2.4)-(2.7) and (3.2)-(3.7),

$$\begin{aligned}
&E_\theta\big[P^{\overline{\pi}}\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)|X\}\big] \\
&= E_\theta\{\Phi(z - \delta g)\} + n^{-1/2}\phi(z - \delta)\{H_1(\pi, \theta) - H_1(\overline{\pi}, \theta) + W^{(1)}(z, \delta, \theta)\} \\
&\quad + n^{-1}\phi(z - \delta)\big(\{H_1(\pi, \theta) - H_1(\overline{\pi}, \theta)\}[\delta\{(z - \delta)^2 - 1\}H_2(\theta) - (z - \delta)H_1(\pi, \theta) \\
&\qquad\qquad + \{z - 2(z^2 - 1)\delta + 3z\delta^2 - \delta^3\}H_3(\theta)] \\
&\qquad + \delta[H_5(\overline{\pi}, \theta) - \frac{1}{2}\{H_1(\pi, \theta)\}^2 + 3H_1(\overline{\pi}, \theta)H_3(\theta) - H_4(\overline{\pi}, \theta)] \\
&\qquad + z\{H_5(\pi, \theta) - H_5(\overline{\pi}, \theta)\} + W^{(2)}(z, \delta, \theta)\big) + o(n^{-1}) \ ,
\end{aligned} \tag{3.8}$$

where $W^{(i)}(z, \delta, \theta)(i = 1, 2)$ are at most of order $O(1)$ and do not involve $\pi(\cdot)$ or $\overline{\pi}(\cdot)$. In particular, choosing $\overline{\pi}(\cdot)$ as the same as $\pi(\cdot)$ in (3.8) we have

$$E_\theta\big[P^{\pi}\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)|X\}\big]$$

$$= E_\theta\{\Phi(z-\delta g)\} + n^{-1/2}\phi(z-\delta)W^{(1)}(z,\delta,\theta)$$

$$+ n^{-1}\phi(z-\delta)\big(\delta[H_5(\pi,\theta) - \frac{1}{2}\{H_1(\pi,\theta)\}^2 + 3H_1(\pi,\theta)H_3(\theta) - H_4(\pi,\theta)]$$

$$+ W^{(2)}(z,\delta,\theta)\big) + o(n^{-1}) \ . \tag{3.9}$$

Step III. Equation (3.8) enables us to use the approach of Ghosh & Mukerjee (1991), which is reminiscent of Dawid (1991), to calculate the frequentist probability appearing in the left-hand side of (3.1). To that end, let the prior $\overline{\pi}(\cdot)$ in (3.8) satisfy the conditions in Bickel & Ghosh (1990) and suppose $\overline{\pi}(\cdot)$ and its first-order partial derivatives vanish on the boundaries of a rectangle containing θ as an interior point. Now $P_\theta\{\theta_1 + \delta(I^{11}/n)^{1/2} \leq \theta_1^{(1-\alpha)}(\pi, X)\}$ is obtained, up to $o(n^{-1})$, by integrating (3.8) by parts with respect to such a $\overline{\pi}(\cdot)$ and then allowing $\overline{\pi}(\cdot)$ to converge weakly to the degenerate measure at θ. Further simplification, using (3.4)-(3.7) and (3.9), shows that the left-hand side of (3.1) is equal to

$$n^{-1/2}\phi(z-\delta)\{\pi(\theta)\}^{-1}\Delta_1(\pi,\theta)$$

$$+n^{-1}\phi(z-\delta)\{\pi(\theta)\}^{-1}\big[-\frac{1}{2}(z-\delta)\Delta_2(\pi,\theta) + \frac{1}{6}(z-\delta)(1-2z\delta+\delta^2)\Delta_3(\pi,\theta)$$

$$+ \delta\Delta_4(\pi,\theta) + \delta\{(z-\delta)^2 - 1\}\Delta_5(\pi,\theta)\big] + o(n^{-1}),$$

where

$$\Delta_1(\pi,\theta) = D_i\{(I^{11})^{-1/2}I^{i1}\pi(\theta)\}, \tag{3.10}$$

$$\Delta_2(\pi,\theta) = D_iD_j\{\tau^{ij}\pi(\theta)\} - D_s\{L_{ijr}\tau^{ij}\sigma^{rs}\pi(\theta)\}, \tag{3.11}$$

$$\Delta_3(\pi,\theta) = D_i\{L_{jrs}\tau^{ij}\tau^{rs}\pi(\theta)\}, \tag{3.12}$$

$$\Delta_4(\pi,\theta) = D_i\{(I^{11})^{-1/2}\beta_j(\theta)I^{ij}\pi(\theta)\}, \tag{3.13}$$

$$\Delta_5(\pi,\theta) = D_i\{(I^{11})^{-1/2}\beta_j(\theta)\tau^{ij}\pi(\theta)\} \ . \tag{3.14}$$

Thus (3.1) holds with $r = 1$ for each δ and α if and only if

$$\Delta_1(\pi,\theta) = 0 \ . \tag{3.15}$$

Furthermore, (3.1) holds with $r = 2$ for each δ and α if and only if in addition

$$\Delta_2(\pi, \theta) = \Delta_3(\pi, \theta) = \Delta_4(\pi, \theta) = \Delta_5(\pi, \theta) = 0 \ . \tag{3.16}$$

Remark 1. From Peers (1965), it is known that the necessary and sufficient condition for a prior to be first-order probability matching is also given by (3.15). Thus, we reach the satisfying conclusion that first-order probability matching priors also match the alternative coverage probabilities up to the same order of approximation. In particular, the Jeffreys' prior enjoys this property in the absence of nuisance parameters.

Remark 2. However, the same conclusion does not necessarily hold with second-order probability matching priors. As noted in Mukerjee & Ghosh (1997), a prior is second-order probability matching if and only if

$$\Delta_1(\pi, \theta) = 0 \quad \text{and} \quad \Delta_2(\pi, \theta) - \frac{1}{3}\Delta_3(\pi, \theta) = 0. \tag{3.17}$$

Since the conditions (3.15), (3.16) are more stringent than (3.17), a second-order probability matching prior may or may not also match the alternative coverage probabilities up to the same order of approximation. In §4 we shall present illustrative examples; see also Remark 4 below.

Remark 3. Considerable simplification of our results is possible under orthogonal parameterisation. Then $I_{1j} = 0$, $2 \leq j \leq p$, and by (3.10) a prior satisfies (3.15) if and only if it is of the form $\pi(\theta) = d(\theta_2, \ldots, \theta_p) I_{11}^{1/2}$, see Tibshirani (1989), where $d(\cdot)(> 0)$ is any smooth function. By (3.11)-(3.14), such a prior satisfies (3.16) as well if and only if

$$\sum_{i=2}^{p} \sum_{j=2}^{p} D_i \{ L_{11j} I^{ij} I_{11}^{-1/2} d(\theta_2, \ldots, \theta_p) \} = 0 \ , \tag{3.18}$$

$$\sum_{i=2}^{p} \sum_{j=2}^{p} D_i\{L_{j,11} I^{ij} I_{11}^{-1/2} d(\theta_2, \ldots, \theta_p)\} = 0 \ , \tag{3.19}$$

$$D_1(I_{11}^{-3/2} L_{111}) = 0 \ , \quad D_1(I_{11}^{-3/2} L_{1,11}) = 0 \ . \tag{3.20}$$

Note that the two conditions in (3.20) are model conditions.

Remark 4. In particular, for $p = 1$, i.e. under the absence of nuisance parameters, conditions (3.18), (3.19) do not arise and Jeffreys' prior matches the alternative coverage probabilities at the second order if and only if the model conditions (3.20) hold. Incidentally, as noted in Welch Peers (1963), for $p = 1$ Jeffreys' prior is second-order probability matching if and only if $D_1(I_{11}^{-3/2} L_{1,1,1}) = 0$, a condition which is implied by (3.20).

Remark 5. We are working here under the assumptions of Johnson (1970), with $K = 2$ in his notation. From these, together with appropriate Edgeworth assumptions, it can be seen that the margins of error in (2.11) and also in the expression for the left-hand side of (3.1), as given just before (3.10), are actually at most of order $O(n^{-3/2})$.

4. Examples

As indicated in Remark 2, in any given context if the class of second-order probability matching priors is nonempty then either (i) all or (ii) none or (iii) only some but not all of them may match second-order alternative coverage probabilities as well. Examples corresponding to these three possibilities are presented below. Note that under situation (iii), as illustrated in Example 3, consideration of alternative coverage probabilities helps in narrowing down the class of second-order probability matching priors.

Example 1. Consider the ratio of exponential means model

$$f(x;\theta) = \{\mu_1(\theta)\mu_2(\theta)\}^{-1} \exp\left[- \{x^{(1)}/\mu_1(\theta) + x^{(2)}/\mu_2(\theta)\}\right] \ , \quad x^{(1)} \ , \ x^{(2)} > 0 \ ,$$

where $x = (x^{(1)}, x^{(2)})'$, $\mu_1(\theta) = \theta_2\theta_1^{1/2}$, $\mu_2(\theta) = \theta_2\theta_1^{-1/2}$; θ_1, $\theta_2 > 0$. Note that $\theta_1 = \mu_1(\theta)/\mu_2(\theta)$. Under this parameterisation, parametric orthogonality holds and $I_{11} = \frac{1}{2}\theta_1^{-2}$, $I_{22} = 2\theta_2^{-2}$, $L_{111} = -3L_{1,11} = \frac{3}{2}\theta_1^{-3}$, $L_{112} = -L_{2,11} = \frac{1}{2}(\theta_1^2\theta_2)^{-1}$. Hence, by (3.18)-(3.20), it follows that the unique second-order probability matching prior $\pi(\theta) \propto (\theta_1\theta_2)^{-1}$, see Mukerjee & Dey (1993), also matches alternative coverage probabilities up to the second order.

Example 2. Consider the ratio of normal means model (Cox & Reid, 1987)

$$f(x;\theta) = \phi(x^{(1)} - \mu_1(\theta))\phi(x^{(2)} - \mu_2(\theta)) \ ,$$

where $x = (x^{(1)}, x^{(2)})'$, $\mu_1(\theta) = \theta_1\theta_2/(\theta_1^2+1)^{1/2}$, $\mu_2(\theta) = \theta_2/(\theta_1^2+1)^{1/2}$; θ_1, $\theta_2 > 0$. Note that $\theta_1 = \mu_1(\theta)/\mu_2(\theta)$. Under this parameterisation, parametric orthogonality holds and $I_{11} = \theta_2^2/(\theta_1^2+1)^2$, $L_{111} = -3L_{1,11} = 6\theta_1\theta_2^2/(\theta_1^2+1)^3$, so that the model conditions (3.20) do not hold. Hence it is impossible to match alternative coverage probabilities up to the second order although, as noted in Mukerjee & Dey (1993), a second-order probability matching prior is available for this problem.

Example 3. Consider the bivariate normal model with means μ_1, μ_2, variances ξ_1^2, ξ_2^2 and correlation coefficient ρ. The parameters are all unknown and interest lies in the regression coefficient $\rho\xi_2/\xi_1$; compare with Datta & Ghosh (1995). Reparameterise as $\theta_1 = \rho\xi_2/\xi_1$, $\theta_2 = \xi_2^2(1-\rho^2)$, $\theta_3 = \xi_1^2$, $\theta_4 = \mu_1$, $\theta_5 = \mu_2$ ($-\infty < \theta_1, \theta_4, \theta_5 < \infty$; θ_2, $\theta_3 > 0$). Then θ_1 becomes the parameter of interest, parametric orthogonality holds,

$$I_{11} = \theta_3/\theta_2 \ , \ I_{22} = \frac{1}{2}\theta_2^{-2} \ , \ I_{33} = \frac{1}{2}\theta_3^{-2} \ , \ I_{23} = I_{24} = I_{25} = I_{34} = I_{35} = 0 \ ,$$

$$L_{111} = L_{113} = L_{114} = L_{115} = L_{1,11} = L_{2,11} = L_{4,11} = L_{5,11} = 0 \ ,$$

$$L_{112} = \theta_3/\theta_2^2 \ , \ L_{3,11} = -\theta_2^{-1}$$

and as noted in Mukerjee & Ghosh (1997), a prior $\pi(\cdot)$ is second-order probability matching if and only if

$$\pi(\theta) = \{\tilde{d}(\theta_3, \theta_4, \theta_5)/\theta_2^{1/2}\} I_{11}^{1/2} \ , \tag{4.1}$$

where $\tilde{d}(\cdot)(>0)$ is any smooth function. By (3.18)-(3.20), such a prior also matches alternative coverage probabilities up to the second order if and only if $\tilde{d}(\theta_3, \theta_4, \theta_5) = d^*(\theta_4, \theta_5)/\theta_3^{3/2}$, i.e. if and only if

$$\pi(\theta) = d^*(\theta_4, \theta_5)/(\theta_2\theta_3) \ , \tag{4.2}$$

where $d^*(\cdot)(>0)$ is any smooth function.

For further illustration, we consider the class of priors of the form

$$\pi(\theta) = d^*(\theta_4, \theta_5)\theta_3^q/\theta_2 \ , \tag{4.3}$$

where q is any real number and $d^*(\cdot)(>0)$ is any smooth function. It is easy to see that every such prior is also of the form (4.1) and hence second-order probability matching. With $\pi(\theta)$ as in (4.3), one can check that the expression for the left-hand side of (3.1), as shown just before (3.10), reduces to

$$-n^{-1}\phi(z-\delta)\delta(q+1) + o(n^{-1}) \ .$$

The above is of order $o(n^{-1})$, identically in α and δ, if and only if $q=-1$, in which case (4.3) reduces to (4.2) and matches alternative coverage probabilities up to the second order. For $q \neq -1$, obviously such matching of alternative coverage probabilities no longer takes place. The departure from matching then depends on the absolute value of $\phi(z-\delta)\delta(q+1)$ and hence, for any fixed α and $\delta(\neq 0)$, is directly proportional to $|q+1|$.

Acknowledgement

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

References

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

Bickel, P.J. & Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction – a Bayesian argument. *Ann. Statist.* **18**, 1070-90.

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

Datta, G.S. & Ghosh, J.K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika* **82**, 37-45.

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

Ghosh, J.K. & Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *J. Mult. Anal.* **36**, 385-93.

Johnson, R.A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851-64.

Kass, R.E & Wasserman, L. (1996). Formal rules for selecting prior distributions: a review and annotated bibliography. *J. Am. Statist. Assoc.* **91**, 1343–70.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd ed. New York: Wiley.

Mukerjee, R. & Dey, D.K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika* **80**, 499-505.

Mukerjee, R. & Ghosh, M. (1997). Second order probability matching priors. *Biometrika* **84**, 970-5.

Nicolaou, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. R. Statist. Soc.* B **55**, 377-90.

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

Reid, N. (1995). Likelihood and Bayesian approximation methods (with Discussion). In *Bayesian Statistics V.* Ed. J.M. Bernardo et al.), pp. 351-68. Oxford: Oxford Univ. Press.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-8.

Welch, B. & Peers, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Statist. Soc.* B **25**, 318-29.

List of symbols used in the draft paper

A: in posterior density (2.3)

C: observed information matrix

D: derivative operator

E: expectation

H: as in (3.4)-(3.7)

I: information matrix

L: moments involving derivatives of log-density

O: order

P: probability

S: set on which posterior expansions are valid

V: derivatives of log-density

W: quantities which do not involve priors (in (3.2), (3.8))

X: random variable

a: derivatives of log likelihood

c: elements of observed information matrix

d: function of nuisance parameters

f: density

g: as in (3.2)

h: as in (2.2)

i: dummy subscript/superscript

j: dummy subscript/superscript

k: related to observed information matrix (see p.3 of draft paper)

n: sample size

o: order

p: dimension of parametric vector

r: dummy subscript/superscript; also, order of approximation

s: dummy subscript/superscript

u: as in (2.8)

x: as in $f(x;\theta)$

y: as defined just below (2.10)

z: normal quantile

$\mathcal{R}$: real space

α: in coverage probability

β: as in page 3 of draft paper

δ: contiguous alternative

π: prior

θ: parameter

τ: related to Fisher information matrix (see p.3 of draft paper)

σ: related to Fisher information matrix (see p.3 of draft paper)

λ: as defined in (2.1)

ϕ: standard normal density

μ: mean

ξ: standard deviation

ρ: correlation coefficient

Φ: standard normal cdf

Δ: as in (3.10)-(3.14)