On priors providing frequentist validity for Bayesian inference

By GAURI SANKAR DATTA

Department of Statistics, University of Georgia, Athens, Georgia 30602-1952, U.S.A.

AND JAYANTA KUMAR GHOSH

Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India

Summary

We derive the differential equation that a prior must satisfy if the posterior probability of a one-sided credibility interval for a parametric function and its frequentist probability agree up to $O(n^{-1})$. This equation turns out to be identical with Stein's equation for a slightly different problem, for which also our method provides a rigorous justification. Our method is different in details from Stein's but similar in spirit to Dawid (1991) and Bickel & Ghosh (1990). Some examples are provided.

Some key words: Confidence set; Credible set; Noninformative prior; Posterior distribution; Probability-matching equation; Probability-matching prior.

1. Introduction

Suppose $X_1, \ldots, X_n$ are independently and identically distributed with density $f(x; \theta)$, where $\theta = (\theta_1, \ldots, \theta_p)^T$ is a p-dimensional parameter vector. Consider a prior density $\pi(\theta)$ for θ which has the following property of matching frequentist and posterior probability for a real-valued twice continuously differentiable parametric function $t(\theta)$:

$$P_\theta\left[\frac{\sqrt{n}\{t(\theta)-t(\hat{\theta})\}}{\sqrt{b}} \leqslant z\right] = P_\pi\left[\frac{\sqrt{n}\{t(\theta)-t(\hat{\theta})\}}{\sqrt{b}} \leqslant z \,\middle|\, X\right] + O_p(n^{-1}) \qquad (1)$$

for all z. In (1), $\hat{\theta}$ is the posterior mode or maximum likelihood estimator of θ and b is the asymptotic posterior variance of $\sqrt{n}\{t(\theta)-t(\hat{\theta})\}$ up to $O_p(n^{-\frac{1}{2}})$, $P_\theta(.)$ is the joint probability measure of $X = (X_1, \ldots, X_n)^T$ under θ, and $P_\pi(.|X)$ is the posterior probability measure of θ under π. Such a prior may be sought in an attempt to reconcile a frequentist and Bayesian approach (Peers, 1965), or to find or in some sense validate a noninformative prior (Berger & Bernardo, 1989; Ghosh & Mukerjee, 1991, 1992a, b; Nicolaou, 1993; Tibshirani, 1989), or to construct frequentist confidence sets (Stein, 1985). Another related paper is by DiCiccio & Martin (1993) where similar higher order frequentist confidence limits are obtained by using Bayesian asymptotic calculations.

One of our objects in this paper is to show that (1) holds if and only if

$$\sum \frac{\partial}{\partial\theta_\alpha}\{\eta_\alpha(\theta)\pi(\theta)\} = 0, \qquad (2)$$

where, for

$$\nabla_t(\theta) = \left(\frac{\partial}{\partial\theta_1}t(\theta), \ldots, \frac{\partial}{\partial\theta_p}t(\theta)\right)^T,$$

$\eta(\theta)=(\eta_1(\theta),\ldots,\eta_p(\theta))^{\mathrm{T}}$ is given by

$$\eta(\theta)=\frac{I^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\{\nabla_t^{\mathrm{T}}(\theta)I^{-1}(\theta)\nabla_t(\theta)\}}} \tag{3}$$

satisfying $\eta^{\mathrm{T}}(\theta)I(\theta)\eta(\theta)=1$ for all θ. In (2) and throughout this paper all the summations as well as the ranges of the subscripts α, β and γ extend over 1 to p, unless otherwise explicitly mentioned. Note that, in (3), $I^{-1}(\theta)$ is the inverse of $I(\theta)$, the information matrix of θ per unit observation.

Equation (2) is similar to equation (5·8) of Stein (1985) in the context of a somewhat different matching equation and we will refer to this as Stein's equation, and all priors satisfying (2) as probability-matching priors. It may be mentioned that, to achieve (1), (3) is the only choice for η. If the goal, following Stein (1985), is to get a multiparameter set of the form

$$S_\varepsilon(\hat\theta)=\{\theta:\eta^{\mathrm{T}}(\hat\theta)I(\hat\theta)\sqrt{n}(\theta-\hat\theta)\leqslant z_\varepsilon\}$$

for θ rather than a confidence interval for a real-valued $t(\theta)$ then other choice for η are possible. Stein (1985, p. 510) made a choice of $\eta(\theta)$ in a particular example involving the squared distance of the normal mean vector from the origin. In this example, his chosen η is same as ours. An intuitively attractive choice, at least for the construction of confidence sets for θ, is given by Tibshirani (1989, p. 605) though there is no guarantee that (1) will hold for this choice. Our equation (3) is in general different from Tibshirani's equation, but they agree when $t(\theta)=\theta_1$ and θ_1 is orthogonal to $(\theta_2,\ldots,\theta_p)$ in the sense of Cox & Reid (1987), the case mainly considered by Tibshirani.

We also justify Stein's (1985) equation (5.8) in the context of his original probability matching problem. Our method of proof is more explicitly rigorous than Stein's; see, e.g., Tibshirani (1989). It is somewhat different in details from Stein's but similar in spirit to that of Dawid (1991) and Bickel & Ghosh (1990). Section 2 of the present paper contains the derivation of Stein's equation, the necessary assumptions and the related discussion. Section 3 contains a few illustrative examples.

2. THE EQUATION FOR PROBABILITY-MATCHING PRIORS

Let

$$l(\theta)=n^{-1}\sum_{i=1}^{n}\log f(X_i;\theta),\quad h=\sqrt{n}(\theta-\hat\theta),\quad a_{\alpha\beta}=\{D_\alpha D_\beta l(\theta)\}_{\theta=\hat\theta},$$

$$a_{\alpha\beta\gamma}=\{D_\alpha D_\beta D_\gamma l(\theta)\}_{\theta=\hat\theta},$$

$C=(-a_{\alpha\beta})$, $G=C^{-1}$, where $D_\alpha\equiv\partial/\partial\theta_\alpha$.

Following Ghosh & Mukerjee (1992b) we assume (Johnson, 1970) that θ has a prior density $\pi(\theta)$ which is positive and twice continuously differentiable for all θ. The prior $\pi(\theta)$ will be obtained by solving the probability-matching equation (2) for $t(\theta)$. If $\pi(\theta)$ is not proper we have to assume that there is a fixed positive integer n_0 such that for all $X_1,\ldots,X_{n_0}$ the posterior density of θ is proper. For a prior $\pi(\theta)$, let $P_\pi(.)$ denote the joint probability measure of θ and X. All formal expansions for the posterior, as used here, are valid for sample points in a set S which may be defined along the lines of Johnson (1970) or Bickel & Ghosh (1990, § 2) with $m=1$. The P_θ-probability of S is $1+O(n^{-1})$ uniformly on compact sets of θ. The matrix C is positive definite over S. We also make the Edgeworth assumptions following Bickel & Ghosh (1990, p. 1078). It may be noted that in addition to the Edgeworth assumptions we need the regularity conditions of Bickel & Ghosh (1990) or Ghosh, Sinha & Joshi (1982) to justify the limiting Bayesian arguments for frequentist calculations used later. The last two papers contain more details on these. For calculations up to $O(n^{-1})$ as needed here, the detailed rigorous justification of the limiting Bayesian argument is not as cumbrous as for $o(n^{-1})$ but it is still somewhat lengthy, though straightforward, and hence omitted. It should be mentioned that all the assumptions made about $f(x;\theta)$ will be satisfied for exponential family with θ a sufficiently smooth function of the natural parameter.

For real-valued twice differentiable function $f(\theta)$, we denote the gradient vector of f by $\nabla_f(\theta)=(D_1 f(\theta),\ldots,D_p f(\theta))^{\mathrm{T}}$ and the Hessian matrix of f by $H_f(\theta)=(D_{\alpha\beta}f(\theta))_{\alpha,\beta=1,\ldots,p}$. Then from (2.2) of Ghosh & Mukerjee (1991), the expansion of the posterior density of h is given by

$$\pi(h|X)=(2\pi)^{-p/2}|G|^{-\frac{1}{2}}\exp\left(-\frac{h^{\mathrm{T}}G^{-1}h}{2}\right)$$
$$\times\left\{1+\frac{1}{6\sqrt{n}}\sum_\alpha\sum_\beta\sum_\gamma a_{\alpha\beta\gamma}h_\alpha h_\beta h_\gamma+\frac{1}{\pi(\hat\theta)\sqrt{n}}h^{\mathrm{T}}\nabla_\pi(\hat\theta)+O_p(n^{-1})\right\}. \tag{4}$$

We will now derive from (4) a formal expansion of the posterior characteristic function of $U=\sqrt{n}\{t(\theta)-t(\hat\theta)\}$ up to $O_p(n^{-1})$ by expanding $t(\theta)-t(\hat\theta)$ around $\hat\theta$ and retaining the first two terms. After considerable algebraic simplification we obtain

$$E\{\exp(iqU)|X\}=\exp\left\{\frac{(iq)^2b}{2}\right\}\left\{1+\frac{1}{\sqrt{n}}\pi_1(iq)+O_p(n^{-1})\right\}, \tag{5}$$

where

$$\begin{gathered}\pi_1(y)=\frac{1}{2}y^3g+\frac{1}{2}y\,\mathrm{tr}\,\{GH_t(\hat\theta)\}+\frac{1}{6}ye_1(\hat\theta)+\frac{1}{6}y^3e_2(\hat\theta)+\frac{y}{\pi(\hat\theta)}\tau^{\mathrm{T}}\nabla_\pi(\hat\theta),\\ b=\nabla_t^{\mathrm{T}}(\hat\theta)G\nabla_t(\hat\theta),\quad \tau=(\tau_1,\ldots,\tau_p)^{\mathrm{T}}=G\nabla_t(\hat\theta),\quad g=\tau^{\mathrm{T}}H_t(\hat\theta)\tau,\\ e_1(\hat\theta)=\sum_\alpha\sum_\beta\sum_\gamma a_{\alpha\beta\gamma}(\tau_\alpha g_{\beta\gamma}+\tau_\beta g_{\alpha\gamma}+\tau_\gamma g_{\alpha\beta}),\quad e_2(\hat\theta)=\sum_\alpha\sum_\beta\sum_\gamma a_{\alpha\beta\gamma}\tau_\alpha\tau_\beta\tau_\gamma.\end{gathered} \tag{6}$$

Let $\phi(u|0,b)$ denote a normal density with mean 0 and variance b. Using repeated integration by parts and the normal characteristic function we obtain

$$E\{\exp(iqU)|X\}=\int_{-\infty}^{\infty}\exp(iqu)\left\{1+\frac{1}{\sqrt{n}}\pi_1\left(-\frac{d}{du}\right)\right\}\phi(u|0,b)\,du+O_p(n^{-1}), \tag{7}$$

where $\pi_1(-d/du)\phi(u|0,b)$ is the result obtained by operating $\pi_1(-d/du)$ on $\phi(u|0,b)$. Following Bhattacharya & Ghosh (1978, Lemma), we get from (7) that, for fixed z, the posterior probability on the right-hand side of (1) is given by

$$P_\pi(U\leqslant z\sqrt{b}|X)=\Phi(z)+\frac{1}{\sqrt{n}}\int_{-\infty}^{z}\pi_1\left(-\frac{1}{\sqrt{b}}\frac{d}{dv}\right)\phi(v)\,dv+O_p(n^{-1}), \tag{8}$$

where $\Phi(z)$ and $\phi(z)$ are respectively the standard normal distribution function and density function.

Let $p\lim_\theta$ denote the probability limit under θ. Then define

$$\xi_1(\hat\theta, z) = \frac{1}{6\sqrt{b}}\left\{e_1(\hat\theta) + \frac{z^2-1}{b}e_2(\hat\theta)\right\}, \quad \xi_2(\hat\theta, z) = \frac{1}{2\sqrt{b}}\left[\frac{g(z^2-1)}{b} + \operatorname{tr}\{GH_t(\hat\theta)\}\right],$$

$$d_2^*(\hat\theta, \pi, z) = \frac{\nabla_t^T(\hat\theta)G\nabla_\pi(\hat\theta)}{\pi(\hat\theta)\sqrt{b}} + \sum_{k=1}^{2}\xi_k(\hat\theta, z), \quad \zeta_k(\theta, z) = p\lim_\theta \xi_k(\hat\theta, z) \quad (k=1,2),$$

$$\nabla_2^*(\theta, \pi, z) = p\lim_\theta d_2^*(\hat\theta, \pi, z) = \frac{\eta^T(\theta)\nabla_\pi(\theta)}{\pi(\theta)} + \sum_{k=1}^{2}\zeta_k(\theta, z).$$

Then using the above notation and standard results on Hermite polynomials, we have, under $\theta = \theta_0$,

$$P_\pi(U \leqslant z\sqrt{b}\,|X) = \Phi(z) - \frac{1}{\sqrt{n}}\phi(z)d_2^*(\hat\theta, \pi, z) + O_p(n^{-1})$$

$$= \Phi(z) - \frac{1}{\sqrt{n}}\phi(z)\Delta_2^*(\theta_0, \pi, z) + O_p(n^{-1}). \tag{9}$$

The last equality follows since $\hat\theta - \theta_0 = O_p(n^{-\frac{1}{2}})$ implies

$$d_2^*(\hat\theta, \pi, z) - \Delta_2^*(\theta_0, \pi, z) = O_p(n^{-\frac{1}{2}}).$$

Now to find the expansion of the frequentist probability $P_{\theta_0}(U \leqslant z\sqrt{b})$ under $\theta = \theta_0$, we proceed following Ghosh & Mukerjee (1991). See also Ghosh (1994, Ch. 8) for a detailed argument. Since the difference between the posterior mode and the maximum likelihood estimator of θ is $O_p(n^{-1})$, for the following calculations we assume that $\hat\theta$ is the maximum likelihood estimator. Note that $P_{\theta_0}(U \leqslant z\sqrt{b})$ is obtained by integrating $\Phi(z) - n^{-\frac{1}{2}}\phi(z)\Delta_2^*(\theta, \pi, z)$ with respect to a prior $\pi(\theta)$ which vanishes at the boundary of a rectangle containing θ_0 and satisfies the assumptions of Bickel & Ghosh (1990) or Ghosh, Sinha & Joshi (1982) and then allowing this prior to converge weakly to the measure degenerate at θ_0. To illustrate the limiting process we denote this prior by $\pi_\delta(\theta)$, where δ is the length of each side of the rectangle. Now by integrating by parts the first integral on the right-hand side of (10)

$$\int \Delta_2^*(\theta, \pi_\delta, z)\pi_\delta(\theta)\,d\theta = \sum \int \eta_\alpha(\theta)\frac{\partial \pi_\delta}{\partial\theta_\alpha}\,d\theta + \sum_{k=1}^{2}\int \zeta_k(\theta, z)\pi_\delta(\theta)\,d\theta$$

$$= \int\left\{-\sum\frac{\partial\eta_\alpha}{\partial\theta_\alpha} + \sum_{k=1}^{2}\zeta_k(\theta, z)\right\}\pi_\delta(\theta)\,d\theta \tag{10}$$

and since, for any continuous function $a(\theta)$,

$$\lim_{\delta\downarrow 0}\int a(\theta)\pi_\delta(\theta)d\theta = a(\theta_0),$$

we have

$$\lim_{\delta\downarrow 0}\int \Delta_2^*(\theta, \pi_\delta, z)\pi_\delta(\theta)\,d\theta = -\sum\frac{\partial\eta_\alpha}{\partial\theta_\alpha}\bigg\|_{\theta=\theta_0} + \sum_{k=1}^{2}\zeta_k(\theta_0, z).$$

Using arguments similar to those of Bickel & Ghosh (1990), we derive from the preceding discussion that

$$P_{\theta_0}(U \leqslant z\sqrt{b}) = \Phi(z) - \frac{1}{\sqrt{n}}\phi(z)\left\{-\sum \frac{\partial \eta_\alpha}{\partial \theta_\alpha}\bigg|_{\theta=\theta_0} + \sum_{k=1}^{2} \zeta_k(\theta_0, z)\right\} + O(n^{-1}). \qquad (11)$$

We now determine the matching prior π by equating the coefficients of $n^{-\frac{1}{2}}$ on the right-hand sides of (9) and (11) for all θ_0, that is by solving the differential equation

$$\frac{1}{\pi(\theta)}\eta^{\mathrm{T}}(\theta)\nabla_\pi(\theta) = -\sum \frac{\partial \eta_\alpha}{\partial \theta_\alpha};$$

that is

$$\sum \frac{\partial}{\partial \theta_\alpha}\{\eta_\alpha(\theta)\pi(\theta)\} = 0. \qquad (12)$$

Remark 1. Using $a\,\mathrm{cov}_\theta$ to denote asymptotic covariance under θ, note that

$$\eta_\alpha(\theta) = \frac{a\,\mathrm{cov}_\theta\{\sqrt{n}t(\hat{\theta}), \sqrt{n}\hat{\theta}_\alpha\}}{\sqrt{[a\,\mathrm{cov}_\theta\{\sqrt{n}t(\hat{\theta}), \sqrt{n}t(\hat{\theta})\}]}},$$

up to $O(n^{-\frac{1}{2}})$, where $\hat{\theta}$ is the maximum likelihood estimator of θ.

Remark 2. We will now derive the probability-matching equation for Stein's approximate $(1-\varepsilon)$ confidence set, his (5.3), given in our notation by $S_\varepsilon(\hat{\theta}) = \{\theta : \eta^{\mathrm{T}}(\hat{\theta})I(\hat{\theta})h \leqslant z_\varepsilon\}$, where $\eta(\theta)$ is an arbitrary differentiable vector satisfying $\eta^{\mathrm{T}}(\theta)I(\theta)\eta(\theta) = 1$, and z_ε is the 100ε upper percentile of the standard normal distribution.

To find the posterior and the frequentist probabilities of the set $S_\varepsilon(\hat{\theta})$, we first express the expansion in (4) by using $I(\hat{\theta})$ in place of G^{-1}. Since $G^{-1} - I(\hat{\theta}) = O_p(n^{-\frac{1}{2}})$ under θ, we can rewrite the right-hand side of (4) after some simplification as

$$\pi(h|X) = (2\pi)^{-p/2}|I(\hat{\theta})|^{-\frac{1}{2}}\exp\left\{-\frac{h^{\mathrm{T}}I(\hat{\theta})h}{2}\right\} \times \left\{1 + \frac{1}{\sqrt{n}}P_3(h) + \frac{h^{\mathrm{T}}\nabla_\pi(\hat{\theta})}{\sqrt{n}\pi(\hat{\theta})} + O_p(n^{-1})\right\}, \qquad (13)$$

where $P_3(h)$ is a third degree polynomial in h not involving the prior π. Now we use a linear transformation $W = BI^{\frac{1}{2}}(\hat{\theta})h$, where $I^{\frac{1}{2}}(\hat{\theta})$ is the symmetric positive definite square root of $I(\hat{\theta})$ and $B^{\mathrm{T}} = (b_1, \ldots, b_p)$ is a $p \times p$ orthogonal matrix with $b_1 = I^{\frac{1}{2}}(\hat{\theta})\eta(\hat{\theta})$. Note that

$$W_1 = b_1^{\mathrm{T}}I^{\frac{1}{2}}(\hat{\theta})h = \eta^{\mathrm{T}}(\hat{\theta})I(\hat{\theta})h, \quad h^{\mathrm{T}}\nabla_\pi(\hat{\theta}) = W_1\eta^{\mathrm{T}}(\hat{\theta})\nabla_\pi(\hat{\theta}) + \sum_{\alpha=2}^{p} W_\alpha b_\alpha^{\mathrm{T}}I^{-\frac{1}{2}}(\hat{\theta})\nabla_\pi(\hat{\theta}).$$

By this transformation and integrating out $W_2, \ldots, W_p$, we get from (13) that the expansion of the posterior density of W_1 is given by

$$\pi(w_1|X) = (2\pi)^{-\frac{1}{2}}\exp\left(-\frac{w_1^2}{2}\right)\left\{1 + \frac{1}{\sqrt{n}}Q_3(w_1, \hat{\theta}) + w_1\frac{\eta^{\mathrm{T}}(\hat{\theta})\nabla_\pi(\hat{\theta})}{\sqrt{n}\pi(\hat{\theta})} + O_p(n^{-1})\right\},$$

where $Q_3(w_1, \hat{\theta})$ is a third degree polynomial in w_1 depending on $\hat{\theta}$ but not on the prior π. Consequently, the posterior coverage probability of $S_\varepsilon(\hat{\theta})$ is given by

$$\begin{aligned} P_\pi(W_1 \leqslant z_\varepsilon | X) &= 1 - \varepsilon - \frac{1}{\sqrt{n}} \phi(z_\varepsilon) d_s^*(\hat{\theta}, \pi, z_\varepsilon) + O_p(n^{-1}) \\ &= 1 - \varepsilon - \frac{1}{\sqrt{n}} \phi(z_\varepsilon) \Delta_s^*(\theta_0, \pi, z_\varepsilon) + O_p(n^{-1}), \end{aligned} \tag{14}$$

where

$$d_s^*(\hat{\theta}, \pi, z_\varepsilon) = \xi_s(\hat{\theta}, z_\varepsilon) + \frac{\eta^{\mathrm{T}}(\hat{\theta})\nabla_\pi(\hat{\theta})}{\pi(\hat{\theta})}, \quad \Delta_s^*(\theta_0, \pi, z_\varepsilon) = \zeta_s(\theta_0, z_\varepsilon) + \frac{\eta^{\mathrm{T}}(\theta_0)\nabla_\pi(\theta_0)}{\pi(\theta_0)},$$

$$\xi_s(\hat{\theta}, z_\varepsilon) = -\phi^{-1}(z_\varepsilon) \int_{-\infty}^{z_\varepsilon} Q_3(w_1, \hat{\theta})\phi(w_1)\, dw_1, \quad \zeta_s(\theta_0, z_\varepsilon) = p \lim_{\theta_0} \xi_s(\hat{\theta}, z_\varepsilon).$$

The expression given by the last approximation of (14) is valid under $\theta = \theta_0$. Now as in (10) and (11), it follows from (14) that the frequentist coverage probability of $S_\varepsilon(\hat{\theta})$ under θ_0 is given by

$$P_{\theta_0}(W_1 \leqslant z_\varepsilon) = 1 - \varepsilon - \frac{\phi(z_\varepsilon)}{\sqrt{n}} \left\{ \zeta_s(\theta_0, z_\varepsilon) - \sum \frac{\partial \eta_\beta(\theta)}{\partial \theta_\beta} \Bigg\|_{\theta = \theta_0} \right\} + O(n^{-1}). \tag{15}$$

Equating the coefficients of $n^{-\frac{1}{2}}$ on the right-hand sides of (14) and (15), we can match $P_\theta(W_1 \leqslant z_\varepsilon)$ and $P_\pi(W_1 \leqslant z_\varepsilon | X)$ for all θ up to $O_p(n^{-1})$ if π satisfies

$$\sum \frac{\partial}{\partial \theta_\beta} \{\eta_\beta(\theta)\pi(\theta)\} = 0,$$

which is Stein's (1985) equation (5.8).

Remark 3. Note that, from (14), the Bayesian coverage probability of $S_\varepsilon(\hat{\theta})$ under an arbitrary prior π_a is given by

$$P_{\pi_a}\{\theta \in S_\varepsilon(\hat{\theta})\} = 1 - \varepsilon - \frac{\phi(z_\varepsilon)}{\sqrt{n}} \int \{\zeta_s(\theta, z_\varepsilon)\pi(\theta) + \eta^{\mathrm{T}}(\theta)\nabla_\pi(\theta)\} \frac{\pi_a(\theta)}{\pi(\theta)}\, d\theta + O(n^{-1}),$$

which is not equal to $1 - \varepsilon$ up to $O(n^{-1})$ as suggested in (5.5) of Stein (1985). However a simple modification of $S_\varepsilon(\hat{\theta})$ will have the desired accuracy. Define $S_\varepsilon'(\hat{\theta}, \pi)$ by

$$S_\varepsilon'(\hat{\theta}, \pi) = \left\{ \theta : \eta^{\mathrm{T}}(\hat{\theta}) I(\hat{\theta}) h - \frac{1}{\sqrt{n}} d_s^*(\hat{\theta}, \pi, z_\varepsilon) \leqslant z_\varepsilon \right\}.$$

Note that $S_\varepsilon'(\hat{\theta}, \pi)$ depends on π. Since the expansions given in (14) and (15) are locally uniform, it follows that

$$P_\pi\{\theta \in S_\varepsilon'(\hat{\theta}, \pi) | X\} = 1 - \varepsilon + O_p(n^{-1})$$

and consequently the frequentist coverage probability of $S_\varepsilon'(\hat{\theta}, \pi)$ is equal to $1 - \varepsilon$, up to $O(n^{-1})$.

Remark 4. From (9) it follows that the credible set $A_\varepsilon(\hat{\theta}) = (-\infty, t(\hat{\theta}) + \sqrt{(b/n)} z_\varepsilon]$ for $t(\theta)$ has posterior coverage probability $1 - \varepsilon$ accurate only up to $O_p(n^{-\frac{1}{2}})$. However, modifying $A_\varepsilon(\hat{\theta})$ as in Remark 3 to

$$A'_\varepsilon(\hat{\theta}, \pi) = \left(-\infty,\ t(\hat{\theta}) + \surd(b/n) \left\{ z_\varepsilon + \frac{d_2^*(\hat{\theta}, \pi, z_\varepsilon)}{\surd n} \right\} \right],$$

one has the posterior coverage probability of $A'_\varepsilon(\hat{\theta}, \pi)$ and hence the Bayes and the frequentist coverage probability equal to $1-\varepsilon$, up to $O(n^{-1})$.

Remark 5. We notice that the matching equation (12) to match up to $O(n^{-1})$ the posterior and the frequentist distribution functions of $\surd n\{t(\theta) - t(\hat{\theta})\}/\surd b$ for a prior π does not depend on the Hessian matrix $H_t(\theta)$ of $t(\theta)$. From this one may correctly guess that it is possible to approximate up to $O_p(n^{-1})$ the distribution function $\surd n\{t(\theta) - t(\hat{\theta})\}/\surd b$ at some z by the distribution function at some z' of only the first term of Taylor's expansion of $\surd n\{t(\theta) - t(\hat{\theta})\}/\surd b$, that is by that of $\nabla_t^{\mathrm{T}}(\hat{\theta})h/\surd b = U'$, say. Since U' is only a linear function of h, as in Remark 2 we get from (4) directly by linear transformation of variables without all the involved algebra

$$P_\pi(U' \leqslant z \mid X) = \Phi(z) - \frac{\phi(z)}{\surd n}\left\{ m(\hat{\theta}, z) + \frac{s^{\mathrm{T}} \nabla_\pi(\hat{\theta})}{\pi(\hat{\theta})} \right\} + O_p(n^{-1}), \tag{16}$$

where $s = \tau/\surd b$ and, $m(\hat{\theta}, z)$ is a function of $\hat{\theta}$ and z, and does not depend on π. In fact it can be seen through indirect or complicated algebraic arguments that $m(\hat{\theta}, z) = \xi_1(\hat{\theta}, z)$. Since the last expansion is locally uniform in z, we have

$$\begin{aligned} P_\pi\left\{ U' \leqslant z - \frac{\xi_2(\hat{\theta}, z)}{\surd n} \,\middle|\, X \right\} &= \Phi(z) - \frac{\phi(z)}{\surd n} d_2^*(\hat{\theta}, \pi, z) + O_p(n^{-1}) \\ &= P_\pi(U \leqslant z\surd b \mid X) + O_p(n^{-1}). \end{aligned}$$

Finally, one would get the same matching equation (12) by matching the posterior and the frequentist distribution functions of U' up to $O_p(n^{-1})$.

We conclude this section by referring to the more accurate probability-matching results of Mukerjee & Dey (1993). They have determined a prior by matching the posterior and the frequentist distribution functions of scalar θ_1 up to $o_p(n^{-1})$ when there is a single nuisance parameter θ_2 orthogonal to θ_1.

3. Examples

Example 1. Let $X_i = (X_{1i}, X_{2i})^{\mathrm{T}}$ $(i = 1, \ldots, n)$ be independently and identically distributed as $N_2(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_2)^{\mathrm{T}}$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \rho_2^2 \end{pmatrix}.$$

Here $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^{\mathrm{T}}$. We suppose the parametric function of interest is $t(\theta) = \rho\sigma_2/\sigma_1 = \beta_{2|1}$, say, the regression coefficient of X_{21} on X_{11}. The inverse of the information matrix $I(\theta)$ is given by $I^{-1}(\theta) = \text{block diagonal } (\Sigma, D)$, where

$$D = \begin{Bmatrix} \frac{1}{2}\sigma_1^2 & \frac{1}{2}\rho^2\sigma_1\sigma_2 & \frac{1}{2}\sigma_1\rho(1-\rho^2) \\ \frac{1}{2}\rho^2\sigma_1\sigma_2 & \frac{1}{2}\sigma_2^2 & \frac{1}{2}\sigma_2\rho(1-\rho^2) \\ \frac{1}{2}\sigma_1\rho(1-\rho^2) & \frac{1}{2}\sigma_2\rho(1-\rho^2) & (1-\rho^2)^2 \end{Bmatrix}.$$

The probability-matching equation simplifies to

$$\frac{\partial}{\partial\sigma_2}\{(1-\rho^2)^{\frac{1}{2}}\sigma_2\rho\pi(\theta)\}+\frac{\partial}{\partial\rho}\{(1-\rho^2)^{3/2}\pi(\theta)\}=0,$$

which has a solution given by $\pi(\theta)=\sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-3/2}$. This prior has been proposed by Geisser (1965) for inference for ρ and is shown to avoid the marginalisation paradox. Since σ_1 and σ_2 have symmetric roles in $\pi(\theta)$ above, this is also the probability-matching prior for $\rho\sigma_1/\sigma_2=\beta_{1\cdot 2}$ (say), the regression coefficient of X_{11} on X_{21}.

Example 2. Let $X_1, \ldots, X_n$ be independently and identically distributed as $N_p(\mu, \sigma^2 I_p)$, where $\theta=(\mu_1, \ldots, \mu_p, \sigma)^{\mathrm{T}}$. Suppose the parameter of interest is $t(\theta)=\mu^{\mathrm{T}}\mu/\sigma^2$. The information matrix is $I(\theta)=\sigma^{-2}\operatorname{diag}(1, \ldots, 1, 2p)$. The probability-matching equation is given by

$$\sum_{i=1}^{p}\frac{\partial}{\partial\mu_i}\left[\frac{\mu_i\pi(\theta)}{\surd\{2p(\mu^{\mathrm{T}}\mu/\sigma^2)+(\mu^{\mathrm{T}}\mu/\sigma^2)^2\}}\right]=\frac{\partial}{\partial\sigma}\left[\frac{\mu^{\mathrm{T}}\mu\pi(\theta)}{2p\sigma\surd\{2p(\mu^{\mathrm{T}}\mu/\sigma^2)+(\mu^{\mathrm{T}}\mu/\sigma^2)^2\}}\right],$$

which has a solution given by

$$\pi(\theta)=\sigma^{-1}(\mu^{\mathrm{T}}\mu+2p\sigma^2)^{-\frac{1}{2}}(\mu^{\mathrm{T}}\mu)^{-(p-1)/2}.$$

It can be checked that this prior will result in a proper posterior, and for $p=1$ this reduces to the reference prior for μ/σ, proposed by Bernardo (1979).

Example 3. Let $X_1, \ldots, X_n$ be independently and identically distributed as log-normal with parameter $\theta=(\mu, \sigma)^{\mathrm{T}}$. Suppose the parameter of interest is $t(\theta)=\exp(\mu+\frac{1}{2}\sigma^2)$, the mean of X_1. The information matrix is $I(\theta)=\sigma^{-2}\operatorname{diag}(1, 2)$. The probability-matching equation is given by

$$\frac{\partial}{\partial\mu}\left\{\frac{\sigma}{\surd(1+\frac{1}{2}\sigma^2)}\pi(\mu, \sigma)\right\}+\frac{\partial}{\partial\sigma}\left\{\frac{\sigma^2}{2\surd(1+\frac{1}{2}\sigma^2)}\pi(\mu, \sigma)\right\}=0,$$

which has a general solution given by

$$\pi(\mu, \sigma)=\sigma^{-2}(1+\tfrac{1}{2}\sigma^2)^{\frac{1}{2}}f(\sigma^2e^{-\mu})$$

for any nonnegative function f.

ACKNOWLEDGEMENT

G. S. Datta is indebted to Professor Jim Berger for financial support through his National Science Foundation grant. He thanks Professor S. S. Gupta and other members of his department for their warm hospitality. J. K. Ghosh's research was partially supported by a National Science Foundation grant. The authors also thank Professor Berger, Professor Malay Ghosh and an associate editor for many valuable comments.

REFERENCES

BERGER, J. O. & BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Statist. Assoc.* **84**, 200–7.

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. R. Statist. Soc.* B **41**, 113–47.

BHATTACHARYA, R. N. & GHOSH, J. K. (1978). On the validity of formal Edgeworth expansion. *Ann. Statist.* **6**, 434–51.

BICKEL, P. & 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.

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

DICICCIO, T. & MARTIN, M. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc.* B **55**, 305–16.

GEISSER, S. (1965). Bayesian estimation in multivariate analysis. *Ann. Math. Statist.* **36**, 150–9.

GHOSH, J. K. (1994). *Higher Order Asymptotics*. Hayward, CA: Institute of Mathematical Statistics and American Statistical Association.

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.* **38**, 385–93.

GHOSH, J. K. & MUKERJEE, R. (1992a). Non-informative priors (with discussion). In *Bayesian Statistics 4*, Ed. J. M. Bernardo, J. O. Berger, D. V. Lindley and A. F. M. Smith, pp. 195–210. London: Oxford University Press.

GHOSH, J. K. & MUKERJEE, R. (1992b). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. R. Statist. Soc.* B **54**, 867–75.

GHOSH, J. K., SINHA, B. K. & JOSHI, S. N. (1982). Expansions for posterior probability and integrated Bayes risk. In *Proc. 3rd Purdue Symp. on Statistical Decision Theory and Related Topics*, **1**, Ed. S. S. Gupta and J. O. Berger, pp. 403–56. New York: Academic Press.

JOHNSON, R. A. (1970). Asymptotic expansion associated with posterior distributions. *Ann. Math. Statist.* **41**, 851–64.

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

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 points and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc.* B **27**, 9–16.

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

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