PROBABILITY MATCHING EQUATION FOR A PARAMETRIC FUNCTION

by

Gauri Sankar Datta and Jayanta Kumar Ghosh
University of Georgia
Purdue University and Indian Statistical Institute

Technical Report #94-23C

Department of Statistics
Purdue University

September 1994

PROBABILITY MATCHING EQUATION FOR A PARAMETRIC FUNCTION

GAURI SANKAR DATTA*
Department of Statistics
University of Georgia
Athens, Georgia 30602-1952, USA

JAYANTA KUMAR GHOSH†
Department of Statistics
Purdue University
West Lafayette, Indiana
and Indian Statistical Institute
Calcutta, 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 and Ghosh (1990). Some examples are provided.
AMS 1990 Subject Classification:62F15
Key words and phrases: Confidence set, Credible set, Noninformative prior, Posterior distribution, Probability matching equation, Probability matching prior.

* Research partially supported by NSF Grant DMS 89-230-71 of Professor Jim

Berger of Purdue University. All correspondence should be addressed to the first author.

† Research partially supported by NSF Grant Number DMS 93-077-27.

1. INTRODUCTION

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

$$P_\theta\Big\{\frac{\sqrt{n}(\theta_1 - \hat\theta_1)}{\sqrt{b}} \le z\Big\} = P_\pi\Big\{\frac{\sqrt{n}(\theta_1 - \hat\theta_1)}{\sqrt{b}} \le z | X_1, \ldots, X_n\Big\} + O_p(n^{-1}) \tag{1}$$

for all z where $\hat\theta$ is the posterior mode or maximum likelihood estimator of θ and b is the asymptotic posterior variance of $\sqrt{n}(\theta_1 - \hat\theta_1)$ up to $O_p(n^{-1})$, $P_\theta(\cdot)$ is the joint probability measure of $X_1, \ldots, X_n$ under θ and $P_\pi(\cdot|X_1, \ldots, X_n)$ 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 as in Peers (1965), or to find or in some sense validate a noninformative prior as in Berger and Bernardo (1989), Ghosh and Mukerjee (1991, 1992a,b) and Tibshirani (1989), or to construct frequentist confidence sets as in Stein (1985).

Suppose we wish to generalize (1) by considering a twice continuously differentiable function $t(\theta)$ instead of θ_1 and require

$$P_\theta\Big[\frac{\sqrt{n}\{t(\theta) - t(\hat\theta)\}}{\sqrt{b}} \le z\Big] = P_\pi\Big[\frac{\sqrt{n}\{t(\theta) - t(\hat\theta)\}}{\sqrt{b}} \le z | X\Big] + O_p(n^{-1}) \tag{2}$$

where henceforth b denotes the asymptotic posterior variance of $\sqrt{n}(t(\theta) - t(\hat\theta))$ up to $O_p(n^{-1})$ and $X = (X_1, \ldots, X_n)^T$. One of our objects in this article is to show that (2) holds if and only if

$$\sum_{\alpha=1}^{p} \frac{\partial}{\partial\theta\alpha}\left\{\frac{\rho_\alpha^T I^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta) I^{-1}(\theta)\nabla_t(\theta)}}\pi(\theta)\right\} = 0 \tag{3}$$

which may be rewritten as Stein's (1985) equation (5.8)

$$\sum_{\alpha=1}^{p} \frac{\partial}{\partial\theta_\alpha}\left\{\eta_\alpha(\theta)\pi(\theta)\right\} = 0 \tag{4}$$

where $\nabla_t(\theta) = (\frac{\partial}{\partial\theta_1}t(\theta), \dots, \frac{\partial}{\partial\theta_p}t(\theta))^T, \rho_\alpha$ is the αth unit column p-vector and

$$\eta(\theta) = \frac{I^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta)I^{-1}(\theta)\nabla_t(\theta)}} \tag{5}$$

satisfying $\eta^T(\theta)I(\theta)\eta(\theta) = 1$ for all θ and $I^{-1}(\theta)$ is the inverse of $I(\theta)$, the per unit observation information matrix of θ. Equation (4) is due to Stein (1985) in the context of a somewhat different matching equation and we will call this equation as Stein's equation and all priors satisfying (4) as probability matching priors. It may be mentioned that (5) is not the only choice for η. Another intuitively attractive choice, at least for the construction of confidence sets for θ, is given in p. 605 of Tibshirani (1989). Our equation (5) is in general different from Tibshirani's equation, but both agree when $t(\theta) = \theta_1$ and θ_1 is orthogonal to $(\theta_2, \dots, \theta_p)$ in the sense of Cox and 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 quite rigorous unlike 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 and Ghosh (1990). Section 2 of this article 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 $l(\theta) = n^{-1}\sum_{i=1}^{n} \log f(X_i;\theta)$, $h = \sqrt{n}(\theta - \hat{\theta})$, $a_{\alpha\beta} = \{D_\alpha D_\beta l(\theta)\}_{\theta=\hat{\theta}}$, $a_{\alpha\beta\gamma} = \{D_\alpha D_\beta D_\gamma l(\theta)\}_{\theta=\hat{\theta}}$, $C = ((-a_{\alpha\beta}))$, $G = C^{-1}$ where $D_\alpha \equiv \frac{\partial}{\partial\theta_\alpha}$.

Following Ghosh and Mukerjee (1992b) we assume, as in 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 (3) for a real-valued parametric function $t(\theta)$. If $\pi(\theta)$ is not proper we have to assume that there is a fixed positive integer n_0 such that for all $X_1, \dots, X_{n_0}$ the posterior pdf of θ is proper. For a prior pdf $\pi(\theta)$, let $P_\pi(\cdot)$ 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 Section 2 of Bickel and Ghosh (1990) with $m = 1$ with P_θ-probability of S equals to $1+O(n^{-1})$ uniformly on compact sets of θ. The matrix C is positive definite over S. We also make the Edgeworth assumptions as in Bickel and Ghosh (1990, p 1078). It may be noted that in addition to the Edgeworth assumptions as mentioned above we need the regularity conditions in Bickel and Ghosh (1990) or Ghosh, Sinha and Joshi (1982) to justify the limiting Bayesian arguments for frequentist calculations used later. The last two articles 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 messy 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 a real-valued twice differentiable function $f(\theta)$, we denote the gradient vector of f by $\nabla_f(\theta) = (D_1 f(\theta), \ldots, D_p f(\theta))^T$ 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 and Mukerjee (1991), the expansion of the posterior pdf of h is given by

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

Let $U = \sqrt{n}(t(\theta) - t(\hat{\theta}))$. Now we will derive from (6) a formal expansion of the posterior characteristic function of U up to $O_p(n^{-1})$ by expanding $t(\theta) - t(\hat{\theta}))$ around $\hat{\theta}$ by Taylor's expansion retaining the first two terms. After considerable algebraic simplifications we obtain

$$E\left\{\exp(iqU)|X\right\} = \exp\left\{\frac{(iq)^2 b}{2}\right\}\left[1 + \frac{1}{\sqrt{n}}\,\pi_1(iq) + O_p(n^{-1})\right] \tag{7}$$

where $b = \nabla_t^T(\hat\theta)G\nabla_t(\hat\theta)$ and for $\tau = (\tau_1, \ldots, \tau_p)^T = G\nabla_t(\hat\theta)$, $g = \tau^T H_t(\hat\theta)\tau$, $e_1(\hat\theta) = \sum\limits_\alpha \sum\limits_\beta \sum\limits_\gamma a_{\alpha\beta\gamma}(\tau_\alpha g_{\beta\gamma} + \tau_\beta g_{\alpha\gamma} + \tau_\gamma g_{\alpha\beta})$ and $e_2(\hat\theta) = \sum\limits_\alpha \sum\limits_\beta \sum\limits_\gamma a_{\alpha\beta\gamma}\tau_\alpha\tau_\beta\tau_\gamma$,

$$\pi_1(y) = \frac{1}{2}y^3 g + \frac{1}{2}\, y\, tr\{GH_t(\hat\theta)\} + \frac{1}{6}\, y e_1(\hat\theta) + \frac{1}{6}\, y^3 e_2(\hat\theta) + \frac{y}{\pi(\hat\theta)}\tau^T\nabla_\pi(\hat\theta). \quad (8)$$

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

$$E\Big\{\exp(iqU)|X\Big\} = \text{characteristic function of } \Big[1 + \frac{1}{\sqrt{n}}\,\pi_1\big(-\frac{d}{du}\big)\Big]\ \phi(u|0,b) + O_p(n^{-1}) \quad (9)$$

where $\pi_1(-\frac{d}{du})\ \phi(u|0,b)$ is the result obtained by operating $\pi_1(-\frac{d}{du})$ on $\phi(u|0,b)$. Proceeding as in Bhattacharya and Ghosh(1978, vide Lemma), we get from (9) for fixed z, the posterior probability on the right hand side of (2) is given by

$$P_\pi\{U \le z\sqrt{b}|X\} = \Phi(z) + \frac{1}{\sqrt{n}}\int_{-\infty}^{z} \pi_1\Big(-\frac{1}{\sqrt{b}}\frac{d}{dv}\Big)\phi(v)dv + O_p(n^{-1}) \quad (10)$$

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

Let $H_2(z) = z^2 - 1$, $\xi_1(\hat\theta, z) = \frac{1}{6\sqrt{b}}\{e_1(\hat\theta) + \frac{H_2(z)}{b}e_2(\hat\theta)\}$, $\xi_2(\hat\theta, z) = \frac{1}{2\sqrt{b}}\Big[\frac{gH_2(z)}{b} + tr\{GH_t(\hat\theta)\}\Big]$, $d_2^\star(\hat\theta,\pi,z) = \frac{\nabla_t^T(\hat\theta)G\nabla_\pi(\hat\theta)}{\pi(\hat\theta)\sqrt{b}} + \sum\limits_{k=1}^{2}\xi_k(\hat\theta, z)$, $\zeta_k(\theta, z) = p\lim\limits_{\theta}\ \xi_k(\hat\theta, z)$, $k = 1, 2$, $\Delta_2^\star(\theta,\pi,z) = p\lim\limits_{\theta}\ d_2^\star(\hat\theta,\pi,z) = \frac{\eta^T(\theta)\nabla_\pi(\theta)}{\pi(\theta)} + \sum\limits_{k=1}^{2}\zeta_k(\theta,\ z)$

where $p\lim\limits_\theta$ is the convergence in probability limit under θ. Then using the above notations and standard results on Hermite polynomials, we have under $\theta = \theta_0$,

$$\begin{aligned} P_\pi\{U \le z\sqrt{b}|X\} &= \Phi(z) - \frac{1}{\sqrt{n}}\,\phi(z)d_2^\star(\hat\theta,\pi,z) + O_p(n^{-1}) \\ &= \Phi(z) - \frac{1}{\sqrt{n}}\,\phi(z)\Delta_2^\star(\theta_0,\pi,z) + O_p(n^{-1}). \end{aligned} \quad (11)$$

The last equality follows since $\hat{\theta}-\theta_0 = O_p(n^{-1/2})$ implies $d_2^\star(\hat{\theta},\pi,z)-\Delta_2^\star(\theta_0,\pi,z) = O_p(n^{-1/2})$. Now to find the expansion of the frequentist probability $P_{\theta_0}\{U \le z\sqrt{b}\}$ under $\theta=\theta_0$, we proceed as in Ghosh and Mukerjee (1991). See also Ghosh (1993, Ch. 8) for a detailed argument. Since the difference between the posterior mode and the maximum likelihood estimator of θ is $O_p(n^{-1/2})$, for the following calculations we assume that $\hat{\theta}$ is the maximum likelihood estimator. Note that $P_{\theta_0}\{U \le z\sqrt{b}\}$ is obtained by integrating $\Phi(z) - \frac{1}{\sqrt{n}}\,\phi(z)\Delta_2^\star(\theta,\pi,z)$ with respect to a prior $\pi(\theta)$ such that it vanishes at the boundary of a rectangle containing θ_0 (and satisfying the assumptions of Bickel and Ghosh (1990) or Ghosh, Sinha and 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 the each side of the rectangle. Now by integrating by parts the first integral on the right hand side of (12)

$$\begin{aligned}\int \Delta_2^\star(\theta,\pi_\delta,z)\ \pi_\delta(\theta)d\theta &= \sum\int \eta_\alpha(\theta)\frac{\partial\pi_\delta}{\partial\theta_\alpha}d\theta\ +\ \sum_{k=1}^{2}\int \zeta_k(\theta,z)\pi_\delta(\theta)d\theta \\ &= \int\left\{-\sum\frac{\partial\eta_\alpha}{\partial\theta_\alpha}\ +\ \sum_{k=1}^{2}\zeta_k(\theta,z)\right\}\pi_\delta(\theta)d\theta\end{aligned} \tag{12}$$

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

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

Using similar arguments as in Bickel and Ghosh (1990), we have from above

$$P_{\theta_0}\{U \le z\sqrt{b}\} = \Phi(z) - \frac{1}{\sqrt{n}}\phi(z)\left[-\sum\frac{\partial\eta_\alpha}{\partial\theta_\alpha}\|_{\theta=\theta_0} + \sum_{R=1}^{2}\zeta_k(\theta_0,z)\right] + O(n^{-1}). \tag{13}$$

We now determine the probability matching prior π by matching the coefficients of $n^{-\frac{1}{2}}$ on the right hand sides of (11) and (13) for all θ_0, i.e., by solving the differential equation

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

i.e.,

$$\sum \frac{\partial}{\partial \theta_\alpha}\left\{\eta_\alpha(\theta)\pi(\theta)\right\} = 0. \tag{14}$$

Remark 1. Note that, up to $O(n^{-1})$, $\eta_\alpha(\theta) = \frac{a\ \mathrm{cov}_\theta(\sqrt{n}t(\hat\theta),\sqrt{n}\hat\theta_\alpha)}{\sqrt{a\ \mathrm{cov}_\theta(\sqrt{n}t(\hat\theta),\sqrt{n}t(\hat\theta))}}$ where $\hat\theta$ is the maximum likelihood estimator of θ and $a\ \mathrm{cov}_\theta$ is the asymptotic covariance under θ.

Remark 2. We will now derive the probability matching equation for Stein's confidence set given by his (5.3) (in our notations) $S_\alpha(\hat\theta) = \left\{\theta\colon\ \eta^T(\hat\theta)I(\hat\theta)h \le z_\alpha\right\}$ where $\eta(\theta)$ is an arbitrary differentiable vector satisfying $\eta^T(\theta)I(\theta)\eta(\theta) = 1$.

To find the posterior and the frequentist probabilities of the set $S_\alpha(\hat\theta)$, we first express the posterior expansion (6) 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 (6) after some simplifications as

$$\pi(h|X) = (2\pi)^{-\frac{p}{2}}|I(\hat\theta)|^{-\frac{1}{2}}\exp\left\{-\frac{h^TI(\hat\theta)h}{2}\right\}\times$$
$$\left[1 + \frac{1}{\sqrt{n}}P_3(h) + \frac{h^T\nabla_\pi(\hat\theta)}{\sqrt{n}\pi(\hat\theta)} + O_p(n^{-1})\right] \tag{15}$$

where $P_3(h)$ is a third degree polynomial of h not involving the prior π. Now we use 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 $B^T = (b_1,\dots,b_p)$ is orthogonal with $b_1 = I^{\frac{1}{2}}(\hat\theta)\eta(\hat\theta)$. Note that $W_1 = b_1^TI^{\frac{1}{2}}(\hat\theta)h = \eta^T(\hat\theta)I(\hat\theta)h$ and $h^T\nabla_\pi(\hat\theta) = W_1\eta^T(\hat\theta)\nabla_\pi(\hat\theta) + \sum_{\alpha=2}^{p} W_\alpha b_\alpha^T\times$ $I^{-\frac{1}{2}}(\hat\theta)\nabla_\pi(\hat\theta)$. By this transformation and integrating out $W_2,\dots,W_p$, we get from (15) that the expansion of the posterior pdf of W_1 is given by

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

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

$$P_\pi\left\{W_1 \le z_\alpha|X\right\} = 1 - \alpha - \frac{1}{\sqrt{n}}\phi(z_\alpha)d_s^*(\hat\theta,\pi,z_\alpha) + O_p(n^{-1})$$

$$\overset{\text{under } \theta=\theta_0}{=} 1-\alpha-\frac{1}{\sqrt{n}}\phi(z_\alpha)\times$$
$$\left\{\zeta_s(\theta_0,z_\alpha)+\frac{\eta^T(\theta_0)\nabla_\pi(\theta_0)}{\pi(\theta_0)}\right\}+O_p(n^{-1}) \tag{16}$$

where $d_s^*(\hat\theta,\pi,z_\alpha)=\xi_s(\hat\theta,z_\alpha)+\frac{\eta^T(\hat\theta)\nabla_\pi(\hat\theta)}{\pi(\hat\theta)}$,

$$\xi_s(\hat\theta,z_\alpha)=-\phi^{-1}(z_\alpha)\int_{-\infty}^{z_\alpha}Q_3(w_1,\hat\theta)\phi(w_1)dw_1,$$

and

$$\zeta_s(\theta_0,z_\alpha)=p\lim_{\theta_0}\xi_s(\hat\theta,z_\alpha).$$

Now as in (12) and (13), we get from (16), the frequentist coverage probability of $S_\alpha(\hat\theta)$ under θ_0 is given by

$$P_{\theta_0}\{W_1\le z_\alpha\}=1-\alpha-\frac{\phi(z_\alpha)}{\sqrt{n}}\left\{\zeta_s(\theta_0,z_\alpha)-\sum\frac{\partial\eta_\beta(\theta)}{\partial\theta_\beta}\|_{\theta=\theta_0}\right\}+O(n^{-1}) \tag{17}$$

Equating the coefficients of $n^{-\frac{1}{2}}$ on the right hand sides of (16) and (17), we can match $P_\theta\{W_1\le Z_\alpha\}$ and $P_\pi\{W_1\le z_\alpha|X\}$ for all θ up to $O_p(n^{-1})$ if π satisfies

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

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

Remark 3. Note that from (16), the Bayesian coverage probability of $S_\alpha(\hat\theta)$ under an arbitrary prior π_a is given by

$$P_{\pi_a}\{\theta\in S_\alpha(\hat\theta)\}=1-\alpha-\frac{\phi(z_\alpha)}{\sqrt{n}}\int\left[\zeta_s(\theta,z_\alpha)\pi(\theta)+\eta^T(\theta)\nabla_\pi(\theta)\right]\times\frac{\pi_a(\theta)}{\pi(\theta)}d\theta+O(n^{-1})$$

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

$$S'_\alpha(\hat\theta,\pi)=\left\{\theta\colon\ \eta^T(\hat\theta)I(\hat\theta)h-\frac{1}{\sqrt{n}}d_s^*(\hat\theta,\pi,z_\alpha)\le z_\alpha\right\}.$$

Since the expansions given in (16) and (17) are locally uniform, it follows that

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

and consequently for an arbitrary prior π_a

$$P_{\pi_a}\{\theta \in S'_\alpha(\hat{\theta}, \pi)\} = 1 - \alpha + O(n^{-1}),$$

$$P_\theta\{\theta \in S'_\alpha(\hat{\theta}, \pi)\} = 1 - \alpha + O(n^{-1}).$$

Remark 4. From (11) it follows that the credible set $A_\alpha(\hat{\theta}) = \left\{t(\theta) \le t(\hat{\theta}) + \sqrt{\frac{b}{n}} z_\alpha\right\}$ of $t(\theta)$ has posterior coverage probability $1-\alpha$ accurate only up to $O_p(n^{-\frac{1}{2}})$. However, modifying $A_\alpha(\hat{\theta})$ as in Remark 3 by

$$A'_\alpha(\hat{\theta}, \pi) = \left\{t(\theta) \le t(\hat{\theta}) + \sqrt{\frac{b}{n}}\left(z_\alpha + \frac{d_2^*(\hat{\theta}, \pi, z_\alpha)}{\sqrt{n}}\right)\right\}$$

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

Remark 5. We notice that the matching equation (14) to match up to $O(n^{-1})$ the posterior and the frequentist distribution functions of $\frac{\sqrt{n}\{t(\theta)-t(\hat{\theta})\}}{\sqrt{b}}$ for a prior π does not depend on the Hessian matrix $H_t(\theta)$ of $t(\theta)$. From this one may guess, restrospectively but rightly, that it is possible to approximate up to $O_p(n^{-1})$ the distribution function of $\frac{\sqrt{n}\{t(\theta)-t(\hat{\theta})\}}{\sqrt{b}}$ at some z by the distribution function (at some z') of only the first term of Taylor's expansion of $\frac{\sqrt{n}\{t(\theta)-t(\hat{\theta})\}}{\sqrt{b}}$, i.e., by that of $\frac{\nabla_t^T(\hat{\theta})h}{\sqrt{b}} = U'$ (say). Since U' is only a linear function of h, as in Remark 2 we get from (6) directly by linear transformation of variables without all the involved algebra

$$P_\pi\{U' \le z|X\} = \Phi(z) - \frac{\phi(z)}{\sqrt{n}}\left[m(\hat{\theta}, z) + \frac{s^T \nabla_\pi(\hat{\theta})}{\pi(\hat{\theta})}\right] + O_p(n^{-1}) \qquad (18)$$

where $m(\hat{\theta}, z)$ 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

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

Finally, one would get the same matching equation (14) 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 and Dey (1993). They have determined 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. $X_i = (X_{1i}, X_{2i})^T$, $i = 1, \ldots, n$ are i.i.d. $N_2(\mu, \Sigma)$ where $\mu = (\mu_1, \mu_2)^T$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. Here $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)^T$ and 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 $I^{-1}(\theta) =$ Block diagonal (Σ, 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

$$\frac{\partial}{\partial\sigma_2}\left[(1-\rho^2)^{\frac{1}{2}}\sigma_2\rho\pi(\theta)\right] + \frac{\partial}{\partial\rho}\left[(1-\rho^2)^{\frac{3}{2}}\pi(\theta)\right] = 0$$

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

Example 2. $X_1, \ldots, X_n$ are i.i.d. $N_p(\mu, \ \sigma^2 I_p)$ where $\theta = (\mu_1, \ldots, \mu_p, \sigma)^T$. Suppose the parameter of interest is $t(\theta) = \frac{\mu^T\mu}{\sigma^2}$. The information matrix is $I(\theta) = \sigma^{-2}$ Diag $(1, \ldots, 1, 2p)$. The probability matching equation is given by

$$\sum_{i=1}^{p} \frac{\partial}{\partial \mu_i}\left(\frac{\mu_i \pi(\theta)}{\sqrt{2p\frac{\mu^T\mu}{\sigma^2} + \left(\frac{\mu^T\mu}{\sigma^2}\right)^2}}\right) = \frac{\partial}{\partial \sigma}\left(\frac{\mu^T\mu\pi(\theta)}{2p\sigma\sqrt{2p\frac{\mu^T\mu}{\sigma^2} + \left(\frac{\mu^T\mu}{\sigma^2}\right)^2}}\right)$$

which has a solution given by $\pi(\theta) = \sigma^{-1}(\mu^T\mu + 2p\sigma^2)^{-\frac{1}{2}} \ (\mu^T\mu)^{-\frac{p-1}{2}}$. 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. $X_1, \ldots, X_n$ are i.i.d. log-normal with parameter $\theta = (\mu, \sigma)^T$. Suppose the parameter of interest is $t(\theta) = \exp\{\mu + \frac{1}{2}\sigma^2\}$, the mean of X_1. The information matrix is $I(\theta) = \sigma^{-2}$ Diag (1, 2). The probability matching equation is given by

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

which has a general solution given by

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

for any nonnegative function f.

ACKNOWLEDGEMENT

This work was started while the first author was visiting the Department of Statistics at Purdue University in the Fall of 1993. He would like to acknowledge to

Professor Jim Berger for many valuable discussions, and for the financial support. He is thankful to Professor S.S. Gupta and other members of the department for their warm hospitality. Finally, the authors are also thankful to Professor Malay Ghosh for his comments on a preliminary draft.

REFERENCES

Berger, J.O. and 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. Roy. Statist. Soc. Ser B* **41**, 113-47.

Bhattacharya, R.N. and Ghosh, J.K. (1978). Validity of formal Edgeworth expansion. *Ann. Statist.* **6**, 434-51.

Bickel, P. and 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. and Reid, N.(1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser B* **49**, 1-39.

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

Ghosh, J.K. (1994). Higher Order Asymptotics. *Lecture Notes* (to appear).

Geisser, S. (1965). Bayesian Estimation in multivariate analysis. *Ann. Math. Statist.* **36**, 150-9.

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**, 385-93.

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

Ghosh, J.K. and Mukerjee, R. (1992b). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. Roy. Statist. Soc. Ser B* **54**, 867-75.

Ghosh, J.K., Sinha, B.K. and Joshi, S.N. (1982). Expansions for posterior probability and integrated Bayes risk. *Proc. 3rd Purdue Symp. on Statistical Decision Theory and Related Topics* **1**, S.S. Gupta and J.O. Berger (eds.). Academic Press, New York, 403-56.

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

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

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

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