Constructing second-order data dependent probability matching priors from first-order probability matching priors

By PAUL H GARTHWAITE

Department of Mathematics and Statistics, The Open University,

Milton Keynes, MK7 6AA, UK

e-mail: paul.garthwaite@open.ac.uk

FADLALLA G ELFADALY

Department of Mathematics and Statistics, The Open University,

Milton Keynes, MK7 6AA, UK, and

Department of Statistics, Cairo University, Egypt

e-mail: fadlalla.elfadaly@open.ac.uk

AND JOHN R CRAWFORD

School of Psychology, King's College, University of Aberdeen,

Aberdeen, AB24 2UB, UK

e-mail: j.crawford@abdn.ac.uk

1. Introduction

Probability matching priors give approximate frequentist validity to the posterior quantiles of an interest parameter, so that credible intervals approximately meet the definition of a confidence interval. Such priors have attracted significant attention (for example, Datta & Ghosh, 1995; Mukerjee & Ghosh, 1997; Wasserman, 2000; Mukerjee & Reid, 2001); a good review is given in Datta & Mukerjee (2004). A prior is called first- or second-order probability matching if posterior quantiles have frequency accuracy up to order $o(n^{-1/2})$ or $o(n^{-1})$, respectively, where n is the sample size. Finding a first-order matching prior is typically tricky, but finding a second-order matching prior is usually much harder. Moreover, a second-order probability matching prior does not exist for all problems.

The class of prior distributions can be broadened by allowing data to be used in their specification. This approach is explored in Wasserman (2000), Mukerjee (2008) and Ong & Mukerjee (2010), for example. It was also used by Sweeting (2005), in order to develop a general method for constructing first-order probability matching priors. Here we assume that a first-order probability matching prior has been obtained and address the task of using it to form a data dependent second-order matching prior.

We 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. We also suppose that the quantity of interest is a parametric function of θ and that $\pi(\theta)$ is a first-order probability matching prior. Under $\pi(\theta)$, the posterior distribution is

$$\pi(\theta|X) \propto \pi(\theta)L(\theta), \tag{1}$$

where $L(\theta) = \prod_{i=1}^{n} f(X_i;\theta)$ is the likelihood and $X = (X_1, \ldots, X_n)^T$. Following the approach of Wasserman (2000), we replace $L(\theta)$ with a pseudolikelihood $L^*(\theta)$ and base inferences on the pseudoposterior

$$\pi^*(\theta|X) \propto \pi(\theta)L^*(\theta). \tag{2}$$

The pseudoposterior is equivalent to a posterior based on the real likelihood, $L(\theta)$, and a data dependent prior, $\pi^*(\theta) \propto \pi(\theta)L^*(\theta)/L(\theta)$. Wasserman considered the case where $L^*(\theta) = L(\theta) - \inf\{L(\theta)\}$. Here we let $\widehat{\theta}$ denote the maximum likelihood estimates of θ from $L(\theta)$, and assume that $\widehat{\theta}$ also maximizes the pseudolikelihood $L^*(\theta)$. We further assume that $L^*(\theta)$ is close to $L(\theta)$, in the sense that derivatives of their logarithms differ by $O(n^{-1})$ when evaluated at $\widehat{\theta}$ or when their expectations are taken. We obtain a condition on $L^*(\theta)$ under which $\pi^*(\theta)$ is a second-order probability matching prior.

Special attention is given to the case where $L^*(\theta) = \{L(\theta)\}^{(n-\gamma)/n}$, with γ a scalar value determined from the data. This adjusts the data by altering the sample size while leaving the likelihood unchanged in other respects so that, for example, $\widehat{\theta}^* = \widehat{\theta}$. It leads to a mechanical method for constructing a second-order data dependent probability matching prior from a first-order probability matching prior. A benefit of the approach is that most users of statistics can judge whether a change in sample size is large or small, and hence judge whether use of the pseudolikelihood might be distorting results. Modifying the sample size is an obvious means of changing the width of credible intervals that do not have frequentist coverage properties.

When sampling is from a multivariate normal distribution, various objective prior distributions are functions of just the variance parameter and not the mean (Sun and Berger, 2007), suggesting that the variance is the critical parameter in the prior. Hence, in forming a data dependent prior, an attractive choice is to leave the sample size associated with the sample mean unchanged while adjusting the sample size associated with the sample-variance. Results that enable this to be done are given. If the first-order probability matching prior is a function of just the variance parameter, the same is true of the data dependent prior that we construct.

1. Background results

The results developed here lean heavily on impressive work by Mukerjee & Ghosh (1997) and Mukerjee & Reid (2001), also reported in Datta & Mukerjee (2004). Interest focuses on a smooth parametric function, $g(\theta)$. The key results are an approximation for the $(1-\alpha)$th quantile of the posterior distribution of $g(\theta)$ and the frequentist probability related to that quantile approximation. The frequentist probability is accurate to $o(n^{-1})$.

We make the same assumptions as Mukerjee & Reid, which are the assumptions of Johnson (1970, pp 852–853) and the Edgeworth assumptions of Bickel & Ghosh (1990, p 1078). They hold for a broad range of models belonging to the exponential family, as well as many other models that include Cauchy and Student-t. Broadly following Mukerjee and Reid (2001), for $1 \leq j, r \leq p$ let

$$\ell(\theta) = n^{-1} \log L(\theta), \quad D_j = \partial/\partial\theta_j, \quad c_{jr} = -\{D_j D_r \ell(\theta)\}_{\theta=\widehat{\theta}}.$$

Formal expansions for the posterior are assumed valid for sample points in a set S

with p_θ-probability $1+o(n^{-1})$ uniformly on compact θ-sets in the parameter space. An explicit description of S can be given along the lines of Bickel & Ghosh (1990, Section 3). The $p\times p$ matrix $C=(c_{ij})$ is positive-definite over S. Let $C^{-1}=(c^{jr})$.

We follow the summation convention with sums over all repeated superscripts or subscripts ranging from 1 to p. For $1\le\ j,r\ \le p$, let

$$g_j(\theta)=D_jg(\theta),\quad g_{jr}(\theta)=D_jD_rg(\theta),\quad \widehat{g}_j=g_j(\widehat{\theta}),$$

Assume that the gradient vector $\bigtriangledown g(\theta)=(g_1(\theta),\ldots,g_p(\theta))^T$ is nonnull for ever θ and let

$$\bigtriangledown\widehat{g}=(\widehat{g}_1,\ldots,\widehat{g}_p)^T,\quad b=\{\bigtriangledown\widehat{g})^TC^{-1}(\bigtriangledown\widehat{g})\}^{1/2}. \tag{3}$$

Also, let z_α be the $(1-\alpha)$th quantile of the standard univariate normal distribution and define

$$Q^{1-\alpha}(\pi,\,X)=g(\widehat{\theta})+n^{-1/2}b\{z_\alpha+n^{-1/2}K_1(\pi,\alpha)+n^{-1}K_2(\pi,\alpha)\}. \tag{4}$$

Both $K_1(\pi,\alpha)$ and $K_2(\pi,\alpha)$ are functions of $\pi(\theta)$, $\ell(\theta)$, $g(\theta)$ and their derivatives, evaluated at $\widehat{\theta}$. Their definitions, which require tedious notation, are given in Appendix A.

From Mukerjee & Reid (2001),

$$P^\pi\{g(\theta)\le Q^{1-\alpha}(\pi,X)|X\}=1-\alpha+o(n^{-1}) \tag{5}$$

where $P^\pi\{\,.\,|X\}$ is the posterior probability measure under the prior $\pi(.)$. Thus $Q^{1-\alpha}(\pi,X)$ is a second-order approximation to the $(1-\alpha)$th quantile of the posterior distribution of $g(\theta)$.

To relate $Q^{1-\alpha}(\pi,X)$ to a frequentist probability, define I, the per observation Fisher information matrix, from

$$I_{jr}=-E_\theta\{D_jD_r\log f(X_1;\theta)\},\quad I=(I_{jr}). \tag{6}$$

Also, put

$$I^{-1}=(I^{jr}),\quad \eta_j=I^{jr}g_r(\theta),\quad L_{jrs}=E_\theta\{D_jD_rD_s\log f(X_1;\theta)\}, \tag{7}$$

$$\lambda = \{(\bigtriangledown g(\theta))^T I^{-1} (\bigtriangledown g(\theta))\}^{1/2}.$$

Denote the probability density function and $(1-\alpha)$th quantile of the standard univariate normal distribution by $\phi(\,.\,)$ and z_α, respectively. From Mukerjee & Reid (2001), the frequentist coverage probability is:

$$\begin{aligned} P_\theta\{g(\theta) \le Q^{1-\alpha}(\pi, X)\} &= 1 - \alpha + n^{-1/2}\frac{\phi(z_\alpha)}{\pi(\theta)}\Delta_1(\pi,\theta) \\ &+ n^{-1}\frac{z_\alpha\phi(z_\alpha)}{\pi(\theta)}\Delta_2(\pi,\theta) + o(n^{-1}), \end{aligned} \tag{8}$$

where

$$\Delta_1(\pi,\theta) = D_j\{\lambda^{-1}\eta_j\pi(\theta)\}, \tag{9}$$

$$\begin{aligned} \Delta_2(\pi,\theta) &= D_u[\{\lambda^{-4}\eta_r(\tfrac{1}{3}L_{jrs}\eta_r\eta_s + g_{jr}(\theta)\eta_j\eta_r)) - \lambda^{-2}I^{jr}(\tfrac{1}{2}L_{jsu}\eta_s\eta_u \\ &+ g_{js}(\theta)\eta_s)\}\pi(\theta)] + \tfrac{1}{2}D_jD_r\{\lambda^{-2}\eta_j\eta_r\pi(\theta)\}. \end{aligned} \tag{10}$$

This gives the standard results for testing whether a prior is probability matching. A prior is first-order matching if $\Delta_1(\pi,\theta) = 0$ and second-order matching if $\Delta_2(\pi,\theta) = 0$.

Remark 1. Much research has considered the case where θ_1, the first component of θ, is the quantity whose quantiles are of interest (Tibshirani, 1989; Mukerjee & Ghosh, 1997). When $g(\theta) = \theta_1$, the formula for $\Delta_2(\pi,\theta)$ in equation (11) simplifies to

$$\Delta_2^{\#}(\pi,\theta) = \tfrac{1}{6}D_u\{\pi(\theta)\tau^{jr}L_{jrs}(3I^{su} - 2\tau^{su})\} - \tfrac{1}{2}D_jD_r\{\pi(\theta)\tau^{jr}\}, \tag{11}$$

where $\tau^{jr} = I^{j1}I^{r1}/I^{11}$.

2. Data dependent priors

By assumption $\pi(\theta)$ is a first-order probability matching prior, so $\Delta_1(\pi,\theta) = 0$. The aim is to form a data dependent prior, $\pi^*(\theta)$, such that $\Delta_1(\pi^*,\theta) = \Delta_1(\pi,\theta) + o(n^{-1})$ and, in addition,

$$P_\theta\{g(\theta) \le Q^{(1-\alpha)}(\pi^*, X)\} =$$

$$P_\theta\{g(\theta) \le Q^{(1-\alpha)}(\pi, X)\} - \ n^{-1}\frac{z_\alpha \phi(z_\alpha)}{\pi(\theta)}\Delta_2(\pi,\theta) \ + \ o(n^{-1}). \tag{12}$$

As noted in the introduction, the posterior distribution from the data dependent prior is identical to the posterior distribution obtained by combining $\pi(\theta)$ with the pseudolikelihood, $L^*(\theta)$. Here, rather than changing the prior distribution from $\pi(\theta)$ to $\pi^*(\theta)$, we instead treat $L^*(\theta)$ as the likelihood.

2.1 *General result*

By assumption, $\widehat{\theta}$ is the value of θ that maximizes both $L(\theta)$ and $L^*(\theta)$. Put

$$\tilde{\ell}(\theta) = n^{-1}\log L^*(\theta). \tag{13}$$

We add a tilde to the quantities defined in Section 2.1 so as to denote the corresponding quantities derived from $\tilde{\ell}(\theta)$. Thus, for example, $\tilde{c}_{jr} = -\{D_jD_r\tilde{\ell}(\theta)\}_{\theta=\widehat{\theta}}$, $\tilde{C} = (\tilde{c}_{ij})$, $\tilde{C}^{-1} = (\tilde{c}^{jr})$, $\tilde{b} = \{(\bigtriangledown\widehat{g})^T\tilde{C}^{-1}(\bigtriangledown\widehat{g})\}^{1/2}$, and

$$\tilde{Q}^{1-\alpha}(\pi,\, X) = g(\widehat{\theta}) + n^{-1/2}\tilde{b}\{z_\alpha + n^{-1/2}\tilde{K}_1(\pi,\alpha) + n^{-1}\tilde{K}_2(\pi,\alpha)\}. \tag{14}$$

From its definition, $Q^{(1-\alpha)}(\pi^*, X) = \tilde{Q}^{(1-\alpha)}(\pi, X)$.

We make the assumption that $L(\theta)$ and $L^*(\theta)$ are close. Specifically we assume that for $1 \le j, r, s, u \le p$,

$$\tilde{c}_{jr} \simeq c_{jr} + O(n^{-1}), \quad \{D_jD_rD_s\tilde{\ell}(\theta)\}_{\theta=\widehat{\theta}} \simeq \{D_jD_rD_s\ell(\theta)\}_{\theta=\widehat{\theta}} + O(n^{-1}), \tag{15}$$

$$\{D_jD_rD_sD_u\tilde{\ell}(\theta)\}_{\theta=\widehat{\theta}} \simeq \{D_jD_rD_sD_u\ell(\theta)\}_{\theta=\widehat{\theta}} + O(n^{-1}). \tag{16}$$

As C is a positive-definite matrix, it follows that $\tilde{c}^{jr} = c^{jr} + O(n^{-1})$ for $1 \le j, r \le p$, and $\tilde{\lambda}^{-1} = \lambda^{-1} + O(n^{-1})$. Now for any $\tilde{a}, a, \tilde{b}, b$, if $\tilde{a} = a + O(n^{-1})$ and $\tilde{b} = b + O(n^{-1})$ then (i) $\tilde{a} + \tilde{b} = a + b + O(n^{-1})$ and (ii) $\tilde{a}.\tilde{b} = a.b + O(n^{-1})$. From the definitions of $K_1(\pi,\alpha)$ and $K_2(\pi,\alpha)$, making repeated use of (i) and (ii) yields the following lemma.

Lemma 1. Suppose the conditions in equations (15) and (16) hold. Then $\tilde{K}_1(\pi,\alpha) = K_1(\pi,\alpha) + O(n^{-1})$ and $\tilde{K}_2(\pi,\alpha) = K_2(\pi,\alpha) + O(n^{-1})$.

Our first goal is to obtain a useful formula for the posterior $(1-\alpha)$th quantile of $g(\theta)$ when the posterior distribution is derived from the pseudolikelihood, $L^*(\theta)$. Theorem 1 relates this quantile, $Q^{(1-\alpha)}(\pi^*, X)$, to that obtained from the true likelihood. Theorem 2 then gives the frequentist probability associated with the quantile.

Theorem 1. Suppose the conditions in equations (15) and (16) hold. Then

$$Q^{(1-\alpha)}(\pi^*, X) = Q^{(1-\alpha)}(\pi, X) + n^{-1/2}(\tilde{b} - b)z_\alpha + O(n^{-2}).$$

Proof. From Lemma 1, $n^{-1/2}\tilde{K}_1(\pi,\alpha)+n^{-1}\tilde{K}_2(\pi,\alpha) = n^{-1/2}K_1(\pi,\alpha)+n^{-1}K_2(\pi,\alpha)+O(n^{-3/2})$. Also, $\tilde{b} = b+O(n^{-1})$, so $\tilde{b}\{n^{-1/2}K_1(\pi,\alpha)+n^{-1}K_2(\pi,\alpha)\} = b\{n^{-1/2}K_1(\pi,\alpha)+n^{-1}K_2(\pi,\alpha)\} + O(n^{-3/2})$. Hence, from equation (14), $Q^{1-\alpha}(\pi^*, X) = g(\widehat{\theta}) + n^{-1/2}[b\{z_\alpha+n^{-1/2}K_1(\pi,\alpha)+n^{-1}K_2(\pi,\alpha)\}+(\tilde{b}-b)z_\alpha+O(n^{-3/2})] = Q^{1-\alpha}(\pi, X)+ n^{-1/2}(\tilde{b} - b)z_\alpha + O(n^{-2})$. □

Theorem 2. Suppose the conditions in equations (15) and (16) hold and $\pi(\theta)$ is a first-order probability matching prior. Then

$$P_\theta\{g(\theta) \leq Q^{(1-\alpha)}(\pi^*, X)\}$$

$$= 1 - \alpha + n^{-1}\frac{z_\alpha\phi(z_\alpha)}{\pi(\theta)}\Delta_2(\pi,\theta) - \frac{(\tilde{b}-b)z_\alpha}{b}\phi(z_\alpha) + o(n^{-1}) \tag{17}$$

and $\pi^*(\theta)$ is a second-order probability matching prior if

$$(\tilde{b} - b)/b = n^{-1}[\Delta_2(\pi,\theta)/\{\pi(\theta)\}]_{\theta=\widehat{\theta}}. \tag{18}$$

Proof. From equation (4), $\{Q^{(1-\alpha)}(\pi, X) - g(\widehat{\theta})\}/(n^{-1/2}b) = z_\alpha + O(n^{-1/2})$. Also, $(\tilde{b} - b)$ is $O(n^{-1})$ and $\{g(\theta) - g(\widehat{\theta})\}/(n^{-1/2}b)$ asymptotically follows a standard normal distribution. Hence,

$$P_\theta\left\{\frac{g(\theta) - g(\widehat{\theta})}{n^{-1/2}b} \leq \frac{Q^{(1-\alpha)}(\pi, X) - g(\widehat{\theta})}{n^{-1/2}b} + \frac{(\tilde{b}-b)z_\alpha}{b} + O(n^{-3/2})\right\}$$

$$= P_\theta\left\{\frac{g(\theta) - g(\widehat{\theta})}{n^{-1/2}b} \leq \frac{Q^{(1-\alpha)}(\pi, X) - g(\widehat{\theta})}{n^{-1/2}b}\right\} + \frac{(\tilde{b}-b)z_\alpha}{b}\phi(z_\alpha) + o(n^{-1}).$$

As $\Delta_1(\pi,\theta)=0$, equation (17) follows from equation (8) and Theorem 1.

The regularity conditions imposed on $\pi(\theta)$ imply that $\{\pi(\widehat{\theta})\}^{-1}\to\{\pi(\theta)\}^{-1}$ as $\widehat{\theta}\to\theta$. Also $\Delta_2(\pi,\widehat{\theta})\to\Delta_2(\pi,\theta)$ as $\widehat{\theta}\to\theta$. Thus equation (17) still holds when $\Delta_2(\pi,\theta)/\pi(\theta)$ is replaced by $\{\Delta_2(\pi,\theta)/\pi(\theta)\}_{\theta=\widehat{\theta}}$. It follows that $\pi^*(\theta)$ is a second-order probability matching prior if (18) holds. □

Remark 2. To obtain a second-order probability matching prior choose $L^*(\theta)$ such that equations (15) and (16) hold and equation (18) is satisfied. Then take $\pi^*(\theta)=\pi(\theta)L^*(\theta)/L(\theta)$ as the prior.

Remark 3. Our approach to forming second-order probability matching priors requires the assumption that $\tilde{b}=b+O(n^{-1})$. Hence the approach might have fallen over when equation (18) was reached. Fortunately, equation (18) implies that $\tilde{b}=b+O(n^{-1})$. Also, on reaching equation (18) we might have found that $\tilde{b}$ depended on z_α, when only one specified quantile at a time could be matched.

Remark 4. It has been assumed that $\widehat{\theta}$ maximises both $L^*(\theta)$ and $L(\theta)$. This requirement could be relaxed, so that the the values of θ that maximise $L^*(\theta)$ and $L(\theta)$ differed by up to $O(n^{-1})$. A similar result to Theorem 2 would hold if appropriate regularity conditions were imposed on $\pi(\theta)$, $g(\theta)$ and their derivatives.

2.2 *Change of apparent sample size*

One simple choice for $L^*(\theta)$ is to set it equal to $\{L(\theta)\}^{(n-\gamma)/n}$. As noted in the introduction, this alters the apparent sample size but leaves $\widehat{\theta}^*=\widehat{\theta}$. Also, $\tilde{C}=C(n-\gamma)/n$ so, from equation (3), $\tilde{b}=bn/(n-\gamma)$. Define

$$\Omega=[\Delta_2(\pi,\theta)/\{\pi(\theta)\}]_{\theta=\widehat{\theta}} \tag{19}$$

and put

$$\gamma=\Omega/(1+n^{-1}\Omega). \tag{20}$$

It is readily checked that equation (18) is satisfied. Also, γ is $O(1)$ so $\tilde{c}_{jr}=c_{jr}(n-\gamma)/n=c_{jr}+)(n^{-1}$ and $\tilde{\ell}(\theta)=\ell(\theta)(n-\gamma)/n=\ell(\theta)-(\gamma/n)\ell(\theta)$. Thus the equations in (15) and (16) are also satisfied, so Theorem 2 holds. From Remark 2, it follows that

$$\pi^*(\theta) = \pi(\theta)\{L(\theta)\}^{-\gamma/n}$$

is a second-order probability matching prior if γ is given by equation (19) and (20).

2.3 *Multivariate normal distribution*

Suppose $X_1, \ldots, X_n$ are independent observations for a p-dimensional multivariate normal (MVN) distribution with mean μ and variance matrix Σ. Let $\overline{X}$ and S be the maximum likelihood estimates of μ and Σ. Then $\overline{X}$ and nS are independently distributed as $\mathrm{MVN}(\mu, \Sigma/n)$ and $\mathrm{Wishart}(\Sigma, n-1)$.

To form a data dependent prior, the apparent sample sizes that gave $\overline{X}$ and S can be changed, and flexibility can be increased by allowing them to change to differing values. We shall consider the likelihood that results when $\overline{X}$ and n_2S are independently distributed as $\mathrm{MVN}(\mu, \Sigma/n_1)$ and $\mathrm{Wishart}(\Sigma, n_2-1)$. This likelihood is

$$\begin{aligned} L^*(\mu,\Sigma) = \; & h\,.\,|\Sigma|^{-1/2}\exp\{-(n_1/2)(\overline{X}-\mu)^T\Sigma^{-1}(\overline{X}-\mu)\} \\ & .\,\exp\{-(n_2/2)\,\mathrm{tr}(\Sigma^{-1}S)\}\,|\Sigma|^{-(n_2-1)/2}. \end{aligned} \tag{21}$$

As with the true likelihood, this likelihood is maximised by setting μ and Σ equal to $\overline{X}$ and S. From equation (21),

$$\left[\frac{\partial^2 \log L^*(\mu,\Sigma)}{\partial\mu\partial\mu^T}\right]_{(\mu,\Sigma)=(\overline{X},S)} = -n_1\Sigma^{-1}, \quad \left[\frac{\partial^2 \log L^*(\mu,\Sigma)}{\partial\mu\partial\Sigma}\right]_{(\mu,\Sigma)=(\overline{X},S)} = 0, \tag{22}$$

$$\left[\frac{\partial^2 \log L^*(\mu,\Sigma)}{\partial\Sigma\partial\Sigma}\right]_{(\mu,\Sigma)=(\overline{X},S)} = -\frac{n_2}{2}\left[\frac{\partial^2}{\partial\Sigma\partial\Sigma}\{\mathrm{tr}(\Sigma^{-1}S) - \log|\Sigma|\}\right]_{(\mu,\Sigma)=(\overline{X},S)} \tag{23}$$

Setting $n_1 = n_2 = n$ in (22) and (23) gives the corresponding quantities for the true likelihood.

To return to earlier notation, suppose μ is a $p^*\times 1$ vector and put $(\theta_1, \ldots, \theta_{p^*}) = \mu$. Also, Σ has $p^{\#} = p^*(p^*+1)/2$ distinct parameters parameters which we denote as $\theta_{p^*+1}, \ldots, \theta_p$, where $p^* + p^{\#} = p$. From equations (22) and (23), C

and $\tilde{C}$ are block-diagonal matrices with blocks of size $p^* \times p^*$ and $p^\# \times p^\#$. Put $C = \text{block}(C_{(1)}, C_{(2)})$ and $\tilde{C} = \text{block}(\tilde{C}_{(1)}, \tilde{C}_{(2)})$. Also, the equations in (15) and (16) hold.

Let $\widehat{g}_{(1)} = (\widehat{g}_1, \ldots, \widehat{g}_{p^*})^T$ and $\widehat{g}_{(2)} = (\widehat{g}_{p^*+1}, \ldots, \widehat{g}_p)^T$. Put $T_1 = \widehat{g}_{(1)}^T C_{(1)}^{-1} \widehat{g}_{(1)}$ and $T_1 = \widehat{g}_{(2)}^T C_{(2)}^{-2} \widehat{g}_{(2)}$. then,

$$b = T_1 + T_2, \qquad \tilde{b} = (n/n_1)T_1 + (n/n_2)T_2. \tag{24}$$

The values of n_1 and n_2 should be chosen such that equation (18) is satisfied. Then the second-order probability matching prior is $\pi^*(\theta) \propto \pi(\theta)L^*(\theta)/L(\theta)$. More explicitly,

$$\pi^*(\theta) \propto \pi(\theta) \exp\left\{-\frac{n_1 - n}{2}((\overline{X} - \mu)^T \Sigma^{-1} (\overline{X} - \mu)\right\}$$
$$. \, |\Sigma|^{(n-n_2)/2} \exp\left\{\frac{n_2 - n}{2} \operatorname{tr}(\Sigma^{-1} S)\right\}. \tag{25}$$

Remark 5. Probability matching priors for a multivariate normal distribution are commonly functions of just Σ, and not μ. Two examples are given in Geisser and Cornfield (1963) and further examples may be found in Sun and Berger (2007). For this to be true of $\pi^*(\theta)$, n_1 should be set equal to n. Then (c.f. equation (20)),

$$n - n_2 = \frac{\Omega(T_1 + T_2)}{T_2 + n^{-1}\Omega(T_1 + T_2)}$$

gives n_2.

3. Examples

Acknowledgement

We are grateful to Rahul Mukerjee for helpful advice. The work reported here was funded by a grant from the Medical Research Council, UK.

Appendix

Definition of $K_1(\pi, \alpha)$ *and* $K_2(\pi, \alpha)$. The following additional notation is required. For $1 \le\ j, r, s\ \le p$, let

$$a_{jrs} = \{D_j D_r D_s \ell(\theta)\}_{\theta = \widehat{\theta}}, \quad a_{jrsu} = \{D_j D_r D_s D_u \ell(\theta)\}_{\theta = \widehat{\theta}},$$

$$\pi_j(\theta) = D_j\pi(\theta),\ \pi_{jr}(\theta) = D_jD_r\pi(\theta),\ \widehat{\pi} = \pi(\widehat{\theta}),\ \widehat{\pi}_j = \pi_j(\widehat{\theta}),\ \widehat{\pi}_{jr} = \pi_{jr}(\widehat{\theta}),$$

$$g_{jrs}(\theta) = D_jD_rD_sg(\theta), \quad \widehat{g}_{jr} = g_{jr}(\widehat{\theta}), \quad \widehat{g}_{jrs} = g_{jrs}(\widehat{\theta}), \quad m_j = c^{jr}\widehat{g}_r,$$

$$\widehat{g}_{j0} = \widehat{g}_{jr}m_r, \quad v_{jr} = \widehat{g}_{jr} + a_{jrs}m_s, \quad w_j = \widehat{g}_{j0} + a_{jrs}m_rm_s.$$

Define

$$G_1(\pi) = \left(\frac{\widehat{\pi}_j}{\widehat{\pi}}\right)m_j + \tfrac{1}{2}\,v_{jr}c^{jr},$$

$$G_2(\pi) = \tfrac{1}{2}\,\{G_1(\pi)\}^2 - \tfrac{1}{2}\,\left\{\left(\frac{\widehat{\pi}_j}{\widehat{\pi}}\right)m_j\right\}^2 + \tfrac{1}{2}\,\left(\frac{\widehat{\pi}_{jr}}{\widehat{\pi}}\right)m_jm_r + c^{jr}w_j\left(\frac{\widehat{\pi}_r}{\widehat{\pi}}\right)$$
$$+c^{jr}\left(\tfrac{1}{2}\,\widehat{g}_{jrs}m_s + \tfrac{1}{4}\,a_{jrsu}m_sm_u\right) + c^{js}c^{ru}(\tfrac{1}{4}\,v_{jr}v_{su} + \tfrac{1}{2}\,w_ja_{rsu}),$$

$$G_3 = \tfrac{1}{6}\,a_{jrs}m_jm_rm_s + \tfrac{1}{2}\,\widehat{g}_{jr}m_jm_r, \quad G_6 = \tfrac{1}{2}\,G_3^2,$$

$$G_4(\pi) = G_1(\pi)G_3 + \{\tfrac{1}{24}\,a_{jrsu}m_jm_rm_sm_u + \tfrac{1}{6}\,\widehat{g}_{jrs}m_jm_rm_s + \tfrac{1}{2}\,c^{jr}w_jw_r.$$

Then $K_1(\pi,\alpha)$ and $K_2(\pi,\alpha)$ are given by,

$$K_1(\pi,\alpha) = b^{-1}G_1(\pi) + b^{-3}G_3J_2(z_\alpha),$$

$$\begin{aligned} K_2(\pi,\alpha) = \;& b^{-2}G_2(\pi)J_1(z_\alpha) + b^{-4}G_4(\pi)J_3(z_\alpha) + b^{-6}G_6J_5(z_\alpha) \\ & +2b^{-3}G_3z_\alpha K_1(\pi,\alpha) - \tfrac{1}{2}\,z_\alpha\{K_1(\pi,\alpha)\}^2, \end{aligned}$$

where $J_i(.)$ is the ith degree Hermite polynomial.

References

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.

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

Datta, G. S. & Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics.* New York: Springer.

Geisser, S. & Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *J. R. Statist, Soc.* **B 25** 368–376.

Johnson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41**, 851–864.

Mukerjee, R. (2008). Data-dependent probability matching priors for empirical and related likeihoods. *IMS Collections* **3**, 60–70.

Mukerjee, R. & Ghosh, J. K. (1997). Second-order probability matching priors. *Biometrika* **84**, 970–975.

Mukerjee, R. & Reid, N. (2001). Second-order probability matching priors for a parametric function with application to Bayesian tolerance limits. *Biometrika* **88**, 587–592.

Ong, S. H. & Mukerjee, R. (2010). Data-dependent probability matching priors of the second order. *Statistics* **44**, 291–302.

Sun, D. & Berger, J. O. (2007). Objective Bayesian analysis for the multivariate normal model. In *Bayesian Statistics 8*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. David, D. Heckerman, A, F, M. Smith and M. West (Eds.). Oxford: University Press, 1–38.

Sweeting, T. J. (2005). On the implementation of local probability matching priors for interest parameters. *Biometrika* **92** 47–57.

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

Wasserman, L. (2000). Asymptotic inference for mixture models using data dependent priors. *J. R. Statist, Soc.* **B 62** 159–180.