

# Reference priors with partial information

BY D. SUN

*Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, USA*

e-mail: dsun@stat.missouri.edu

AND J. O. BERGER

*Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708, USA*

e-mail: berger@stat.duke.edu

## SUMMARY

In this paper, reference priors are derived for three cases where partial information is available. If a subjective conditional prior is given, two reasonable methods are proposed for finding the marginal reference prior. If, instead, a subjective marginal prior is available, a method for defining the conditional reference prior is proposed. A sufficient condition is then given under which this conditional reference prior agrees with the conditional reference prior derived in the first stage of the reference prior algorithm of Berger and Bernardo. Finally, under the assumption of independence, a method for finding marginal reference priors is also proposed. Various examples are given to illustrate the methods.

*Some key words:* Kullback-Leibler divergence, noninformative prior, normal distribution, gamma distribution, beta distribution, Neyman-Scott Problem.

## 1. INTRODUCTION

Bayesian analysis using noninformative or default priors has received considerable attention in recent years. A common noninformative prior is the Jeffreys prior (Jeffreys, 1961), which is proportional to the square root of the determinant of the Fisher information matrix. The Jeffreys prior is quite useful for a single parameter, but can be seriously deficient in multiparameter problems (cf. Berger and Bernardo, 1992). For a recent review of various approaches to the development of noninformative priors, see Kass and Wasserman (1996). Here, we will concentrate on the reference prior approach, as developed in Bernardo (1979) and Berger and Bernardo (1992).

In many practical problems, one has partial prior information for some of the parameters. For example, in a  $N(\mu, \sigma^2)$  population, one might possess reasonably strong prior information about  $\mu$ , while the prior information for  $\sigma$  is vague. As another example, Lavine *et al.* (1991) considered robust Bayesian inference with specified prior marginals. As a third example, the prior knowledge could be of independence of the parameters, such as in the ECMO clinical trial example studied by Ware (1989). Another common type of partial information is constraints on the parameter space. This is typically easily handled, however, in that reference priors for a constrained space are almost always just the unconstrained reference prior times the indicator function on the constrained space.

The paper is arranged as follows. In Section 2, we will develop the reference priors when two types of partial information are available. We first consider the case when a conditional prior is known and it is desired to find the marginal reference prior. Two options are given. One is similar to the final stage of Berger and Bernardo's reference prior algorithm; the other is more intuitive and is based on deriving the marginal model.

Next, the conditional reference prior is derived when a marginal prior is known. A sufficient condition is found under which the conditional noninformative prior often agrees with the conditional reference prior from the first stage of Berger and Bernardo's algorithm. Some examples are given in Section 3, illustrating various aspects of these results. In Section 4, an algorithm is proposed for determining marginal reference priors when the two parameters are known to be independent. Some sufficient conditions are given under which the answers can be written in closed form. Formal examples and the ECMO clinical trial example are used for illustration. Finally, Section 5 contains some discussion.

## 2. KNOWING A MARGINAL OR CONDITIONAL PRIOR

### 2.1. INTRODUCTION

Let  $X_n = (x_1, \dots, x_n)$  be a random sample from the density  $p(x; \theta_1, \theta_2)$ , where the parameters  $\theta_1$  and  $\theta_2$  are vectors of dimensions  $d_1$  and  $d_2$ , respectively. Let  $\pi(\theta_1, \theta_2)$  denote the prior density of  $(\theta_1, \theta_2)$ . The following questions are of interest.

1. Suppose there is available a subjective conditional prior density  $\pi^s(\theta_2|\theta_1)$  for  $\theta_2$  given  $\theta_1$ . How can we find the marginal noninformative prior  $\pi^r(\theta_1)$  for  $\theta_1$ ?
2. Suppose there is available a subjective marginal prior density  $\pi^s(\theta_1)$  for  $\theta_1$ . How can we find the conditional noninformative prior  $\pi^r(\theta_2|\theta_1)$  for  $\theta_2$  given  $\theta_1$ .

Solutions to these two questions will be discussed in §§2.2 and 2.3, respectively. We will use  $\pi(\theta_1, \theta_2|X_n)$  to denote the joint posterior density of  $\theta_1$  and  $\theta_2$  and  $\pi(\theta_1|X_n)$  the marginal posterior density of  $\theta_1$ .

### 2.2. FINDING THE MARGINAL PRIOR

When the conditional density of  $\theta_2$  given  $\theta_1$  is available, there are two reasonable options for finding a marginal reference prior  $\pi^r(\theta_1)$ .

*Option 1:* Following Bernardo (1979), define the expected Kullback-Leibler divergence between the marginal posterior density of  $\theta_1$  given  $X_n$  and the marginal prior of  $\theta_1$  by

$$\mathcal{I}\{X_n; \pi^r(\cdot)\} = E\left[\int \pi(\theta_1|X_n) \log\left\{\frac{\pi(\theta_1|X_n)}{\pi^r(\theta_1)}\right\} d\theta_1\right], \quad (1)$$

where the expectation is with respect to the marginal density  $m(X_n) = \int p(X_n|\theta_1)\pi^r(\theta_1)d\theta_1$ .

We seek that prior,  $\pi^r(\theta_1)$ , which maximizes (1) asymptotically, since maximizing the distance between the prior and posterior is a plausible way to define a prior which has minimal influence on the analysis. It follows from Ghosh and Mukerjee (1992) that, under some regularity conditions and as  $n \rightarrow \infty$ ,

$$\mathcal{I}\{X_n; \pi^r(\cdot)\} = \frac{d_1}{2} \log\left(\frac{n}{2\pi e}\right) + \int \pi^r(\theta_1) \log\left\{\frac{\eta(\theta_1)}{\pi^r(\theta_1)}\right\} d\theta_1 + o(1), \quad (2)$$

where

$$\eta(\theta_1) = \exp\left\{\frac{1}{2} \int \pi^s(\theta_2|\theta_1) \log\left(\frac{|\Sigma|}{|\Sigma_{22}|}\right) d\theta_2\right\}. \quad (3)$$

Here  $\Sigma = \Sigma(\theta_1, \theta_2)$  is the per observation Fisher information matrix for  $(\theta_1, \theta_2)$ ,  $\Sigma_{22} = \Sigma_{22}(\theta_1, \theta_2)$  is the per observation Fisher information for  $\theta_2$ , given that  $\theta_1$  is held fixed, and  $|\Sigma|$  is the determinant of  $\Sigma$ .

The reference prior strategy suggests choosing  $\pi^r$  to maximize (1) or (2) asymptotically on compact sets; this can easily be seen to lead to

$$\pi_1^r(\theta_1) \propto \eta(\theta_1). \quad (4)$$

In fact, this is essentially the solution used in Berger and Bernardo (1989, 1992). The following theorem gives an important special case.

THEOREM 1. (a) If  $|\Sigma|/|\Sigma_{22}|$  does not depend on  $\theta_2$ , then, for any subjective conditional density  $\pi^s(\theta_2|\theta_1)$ , the marginal reference prior from Option 1 has the form

$$\pi_1^r(\theta_1) \propto (|\Sigma|/|\Sigma_{22}|)^{\frac{1}{2}}. \quad (5)$$

(b) If  $|\Sigma|/|\Sigma_{22}|$  depends only on  $\theta_2$  and  $\pi^s(\theta_2|\theta_1)$  does not depend on  $\theta_1$ , then  $\pi_1^r(\theta_1) \propto 1$ .

Proof. The results follow from the definition of  $\pi_1^r(\theta_1)$  and the assumption.  $\square$

Option 2: Find the marginal model

$$p(X_n|\theta_1) = \int p(X_n|\theta_1, \theta_2)\pi^s(\theta_2|\theta_1)d\theta_2. \quad (6)$$

Let  $\Sigma^*(\theta_1)$  be the Fisher information matrix for  $\theta_1$  based on this marginal model. Then the reference prior for  $\theta_1$  in this model, again maximizing, asymptotically, the expected Kullback-Leibler divergence between the marginal posterior and the marginal prior, is

$$\pi_2^r(\theta_1) \propto \{|\Sigma^*(\theta_1)|\}^{\frac{1}{2}}. \quad (7)$$

Option 2 more closely mirrors the underlying motivation for reference priors in that, with  $\pi^s(\theta_2|\theta_1)$  given, the information in the data about  $\theta_1$  resides in  $p(X_n|\theta_1)$ . Hence the marginal reference prior for  $\theta_1$  should, ideally, be computed with respect to this mixture model. Unfortunately, the Fisher information matrix for such mixture models is often difficult to compute, so that implementation of Option 2 is often difficult. This same difficulty also motivated the use of the analogue of Option 1 in Berger and Bernardo (1989, 1992), in place of the more natural Option 2.

### 2.3. FINDING THE CONDITIONAL PRIOR

For the case when the marginal prior density  $\pi^s(\theta_1)$  is known, consider the expected Kullback-Leibler divergence between the conditional posterior density of  $\theta_2$ , given  $\theta_1$  and  $X_n$ , and the conditional prior of  $\theta_2$ , given  $\theta_1$ ,

$$\begin{aligned} \mathcal{I}\{X_n; \pi^r(\cdot|\theta_1)\} &= E\left[\int_{\Theta_1} \pi(\theta_1|X_n) \int_{\Theta_2} \pi(\theta_2|\theta_1, X_n) \log\left\{\frac{\pi(\theta_2|\theta_1, X_n)}{\pi^r(\theta_2|\theta_1)}\right\} d\theta_2 d\theta_1\right] \\ &= E\left[\int_{\Theta_1} \int_{\Theta_2} \pi(\theta_1, \theta_2|X_n) \log\left\{\frac{\pi(\theta_1, \theta_2|X_n)}{\pi(\theta_1, \theta_2)}\right\} d\theta_2 d\theta_1\right] \\ &\quad - E\left[\int_{\Theta_1} \pi(\theta_1|X_n) \log\left\{\frac{\pi(\theta_1|X_n)}{\pi(\theta_1)}\right\} d\theta_1\right]. \end{aligned} \quad (8)$$

From an asymptotic expansion of the first term, compare formula (1.1) of Ghosh and Mukerjee (1992, p. 192) and (2), we have that

$$\mathcal{I}\{X_n; \pi^r(\cdot|\theta_1)\} = \frac{d_2}{2} \log\left(\frac{n}{2\pi e}\right) + \int_{\Theta_1} \pi^s(\theta_1) \left[\int_{\Theta_2} \pi^r(\theta_2|\theta_1) \log\left\{\frac{|\Sigma_{22}|^{\frac{1}{2}}}{\pi^r(\theta_2|\theta_1)}\right\} d\theta_2\right] d\theta_1 + o(1). \quad (9)$$

Choosing  $\pi^r(\theta_2|\theta_1)$  to maximize (9) asymptotically thus suggests defining the conditional reference prior as

$$\pi^r(\theta_2|\theta_1) \propto |\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}}. \quad (10)$$

When this conditional reference prior is proper, matters are straightforward. In practice, it may be improper and will then encounter normalization concerns. The compact support argument that is typically used in the reference prior approach (Berger and Bernardo, 1992) may then be applied here. Choose a nested sequence  $\Lambda_1 \subset \Lambda_2 \subset \dots$  of compact subsets of the parameter space  $\Lambda$  for  $(\theta_1, \theta_2)$ , such that  $\cup_i \Lambda_i = \Lambda$  and  $\pi^r(\theta_2|\theta_1)$  has finite mass on  $\Omega_i = \{\theta_2 : (\theta_1, \theta_2) \in \Lambda_i\}$ . Let  $1_A$  be the indicator function on  $A$  and

$$K_i(\theta_1) = \int_{\Omega_i} |\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}} d\theta_2.$$

The conditional reference prior of  $\theta_2$  on  $\Omega_i$  is

$$\pi_i^r(\theta_2|\theta_1) = \frac{|\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}}}{K_i(\theta_1)} 1_{\Omega_i}(\theta_2).$$

Now define the conditional reference prior of  $\theta_2$  by

$$\pi^r(\theta_2|\theta_1) = \lim_{i \rightarrow \infty} \frac{\pi_i^r(\theta_2|\theta_1)}{\pi_i^r(\theta_{20}|\theta_{10})},$$

when the limit exists; here  $(\theta_{10}, \theta_{20})$  is any fixed point. It is easy to see that

$$\pi^r(\theta_2|\theta_1) \propto \lim_{i \rightarrow \infty} \frac{K_i(\theta_{10})}{K_i(\theta_1)} |\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}}. \quad (11)$$

The following theorem gives a sufficient condition under which this limit is proportional to  $|\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}}$ .

**THEOREM 2.** *Assume*

$$|\Sigma_{22}(\theta_1, \theta_2)| = g_1(\theta_1)g_2(\theta_2), \quad (12)$$

*for some functions  $g_1$  and  $g_2$ . Suppose  $\Lambda = \Theta_1 \times \Theta_2$  and the compact sets are chosen to be of the form  $\Lambda_i = \Theta_{1i} \times \Theta_{2i}$ . Then the conditional reference prior of  $\theta_2$  satisfies*

$$\pi^r(\theta_2|\theta_1) \propto |\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}} \propto \{g_2(\theta_2)\}^{\frac{1}{2}}.$$

**Proof.** Clearly, the normalizing constants  $K_i$  are independent of  $\theta_1$ . The result follows immediately. □

Note that the conditional reference prior,  $\pi^r(\theta_2|\theta_1)$ , never depends on the specified marginal prior  $\pi^s(\theta_1)$ .

### 3. EXAMPLES

#### 3.1. MULTINOMIAL DISTRIBUTION

Consider the multinomial density for 3 cells

$$p(y_1, y_2|p_1, p_2) = \frac{k!}{y_1!y_2!(k - y_1 - y_2)!} p_1^{y_1} p_2^{y_2} (1 - p_1 - p_2)^{k - y_1 - y_2}, \quad (13)$$

where  $k$  is a positive integer,  $y_i$  is the observed frequency in cell  $i$ ,  $p_i$  is the probability of cell  $i$ , and  $0 \leq p_1 + p_2 \leq 1$ . Without loss of generality, we will suppress the cell count and probability for the third cell. The Fisher information matrix for  $(p_1, p_2)$  is

$$\Sigma(p_1, p_2) = \frac{1}{k} \left\{ \begin{pmatrix} p_1^{-1} & 0 \\ 0 & p_2^{-1} \end{pmatrix} + \frac{1}{1 - p_1 - p_2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}. \quad (14)$$

The Marginal Reference Prior. Given  $p_1$ , assume that  $p_2$  has a conditional density  $\pi^s(p_2|p_1)$  on  $(0, 1 - p_1)$ . We want to find the marginal reference prior for  $p_1$ .

*Option 1.* Since  $|\Sigma| = 1/\{p_1 p_2 (1 - p_1 - p_2)\}$  and  $\Sigma_{22} = p_2^{-1} + (1 - p_1 - p_2)^{-1}$ ,  $|\Sigma|/\Sigma_{22} = 1/\{p_1(1 - p_1)\}$ . From Theorem 1, the marginal reference prior for  $p_1$  is

$$\pi_1^r(p_1) \propto \frac{1}{\{p_1(1 - p_1)\}^{\frac{1}{2}}}. \quad (15)$$

*Option 2.* Assume that  $n$  observations  $x_i = (y_{1i}, y_{2i})$ ,  $i = 1, \dots, n$ , are obtained. Note that, in reference prior developments, one considers replications of the full experiment in creating the asymptotics. A considerable simplification results if

$$\pi^s(p_2|p_1) = (1 - p_1)^{-1} g\{p_2/(1 - p_1)\} \quad (16)$$

for some density  $g$  on  $(0, 1)$ . This says that the random variable  $p_2/(1 - p_1)$  has the same distribution for any  $p_1$ . Then the marginal model is

$$\begin{aligned} p(X_n|p_1) &= \int_0^{1-p_1} p_1^{y_{1+}} p_2^{y_{2+}} (1 - p_1 - p_2)^{nk - (y_{1+} + y_{2+})} \frac{1}{1 - p_1} g\left(\frac{p_2}{1 - p_1}\right) dp_2 \\ &= p_1^{y_{1+}} (1 - p_1)^{nk - y_{1+}} \int_0^1 s^{y_{1+}} (1 - s)^{nk - y_{1+} - y_{2+}} g(s) ds \\ &\propto p_1^{y_{1+}} (1 - p_1)^{nk - y_{1+}}, \end{aligned}$$

where  $y_{i+} = \sum_{j=1}^n y_{ij}$ . The Fisher information corresponding to this model is then

$$\Sigma^*(p_1) = \frac{nk}{p_1(1 - p_1)}. \quad (17)$$



Therefore,  $\pi_2^r(p_1)$  is also given by (15) when (16) holds. If (16) does not hold, the marginal reference priors will typically be different.

The Conditional Reference Prior of  $p_2$ . For any marginal subjective prior on  $p_1$ , the conditional reference prior of  $p_2$  is

$$\pi^r(p_2|p_1) \propto \frac{1}{\{p_2(1-p_1-p_2)\}^{\frac{1}{2}}}, \quad 0 < p_2 < 1-p_1.$$

### 3.2. NORMAL DISTRIBUTION

Consider the normal density with a mean  $\mu$  and a standard deviation  $\sigma$ ,

$$p(x; \mu, \sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty. \quad (18)$$

The Fisher information matrix for  $(\mu, \sigma)$  is

$$\Sigma = \Sigma(\mu, \sigma) = \text{diag}(\sigma^{-2}, 2\sigma^{-2}).$$

#### Case I. Marginal and Conditional Reference Priors For $\mu$

PROPOSITION 1. (a) If  $\pi^s(\sigma|\mu)$  is independent of  $\mu$ , the marginal reference prior under Option 1 is  $\pi_1^r(\mu) \propto 1$ .

(b) If  $\pi^s(\sigma|\mu)$  is independent of  $\mu$ , the marginal reference prior under Option 2 is  $\pi_2^r(\mu) \propto 1$ .

(c) For any given marginal prior of  $\sigma$ , the conditional reference prior of  $\mu$  is  $\pi^r(\mu|\sigma) \propto 1$ .

Proof. Note that  $|\Sigma|/|\Sigma_{22}| = \sigma^{-2}$ , and (a) follows from Theorem 1(b). For (b), since  $\pi^s(\sigma|\mu) = \pi^s(\sigma)$ , the marginal probability density  $p(X_n|\mu)$  is a scale-mixture of normals

and hence a location probability density. Consequently, the Fisher information for  $\mu$  will be constant. Part (c) is obvious.  $\square$

When  $\pi^s(\sigma|\mu)$  depends on  $\mu$ , Option 1 and Option 2 can generate different marginal priors for  $\mu$ .

*Case II. Marginal and Conditional Reference Priors For  $\sigma$ .*

PROPOSITION 2. (a) For any conditional prior of  $\mu$  given  $\sigma$ , the marginal prior for  $\sigma$  under Option 1 has the form  $\pi_1^r(\sigma) \propto 1/\sigma$ .

(b) If the prior distribution for  $\mu$  is normal with mean  $m$  and variance  $\tau^2$ , then the marginal prior for  $\sigma$  under Option 2 has the form

$$\pi_2^r(\sigma) \propto \left\{ \frac{n-1}{\sigma^2} + \frac{\sigma^2}{(\sigma^2 + n\tau^2)^2} \right\}^{\frac{1}{2}}. \quad (19)$$

(c) For any given marginal prior on  $\mu$ , the conditional prior for  $\sigma$  is also  $\propto 1/\sigma$ , independent of  $\mu$ .

Proof. Part (a) is immediate from Theorem 1 (a). Part (c) is easy. For (b), let  $X_n = (x_1, \dots, x_n)$  be a random sample from  $N(\mu, \sigma^2)$ . Define  $\bar{x}_n = (x_1 + \dots + x_n)/n$  and  $S^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2$ . The marginal model is

$$\begin{aligned} p(X_n|\sigma) &= \frac{1}{\sigma^n} \exp\left\{-\frac{S^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{n(\bar{x}_n - \mu)^2}{2\sigma^2}\right\} \frac{1}{(2\pi)^{\frac{1}{2}}\tau} \exp\left\{-\frac{(\mu - m)^2}{2\tau^2}\right\} d\mu \\ &\propto \frac{1}{\sigma^{n-1}(\sigma^2 + n\tau^2)^{1/2}} \exp\left\{-\frac{S^2}{2\sigma^2} - \frac{n(\bar{x}_n - m)^2}{2(\sigma^2 + n\tau^2)}\right\}. \end{aligned}$$

Therefore, writing  $\tilde{\sigma} = \sigma^2$ , we have

$$\frac{\partial^2}{\partial \tilde{\sigma}^2} \log\{p(X_n|\sigma)\} = \frac{n-1}{2\tilde{\sigma}^2} + \frac{1}{2(\tilde{\sigma} + n\tau^2)^2} - \frac{S^2}{\tilde{\sigma}^3} - \frac{n(\bar{x}_n - m)^2}{(\tilde{\sigma} + n\tau^2)^3}.$$

Under the marginal model, we have  $E_{\tilde{\sigma}} S^2 = (n-1)\tilde{\sigma}$  and  $E_{\tilde{\sigma}}(\bar{x}_n - m)^2 = \tilde{\sigma}/n + \tau^2$ . So the Fisher information for  $\tilde{\sigma}$  under the marginal model is then

$$\Sigma^*(\tilde{\sigma}) = (n-1)/(2\tilde{\sigma}^2) + \frac{1}{2}(\tilde{\sigma} + n\tau^2)^2.$$

The result is immediate. □

Note that  $\pi_2^r(\sigma)$  differs from  $\pi_1^r(\sigma)$ . When  $n \rightarrow \infty$ , however,  $\pi_2^r(\sigma)$  will converge to  $\pi_1^r(\sigma)$ .

### 3.3. GAMMA DISTRIBUTION

Consider the gamma density  $p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$ ,  $x > 0$ , where  $\alpha > 0$  and  $\beta > 0$ . The Fisher information matrix for  $(\alpha, \beta)$  is

$$\Sigma = \Sigma(\alpha, \beta) = \begin{pmatrix} \psi'(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{pmatrix}, \quad (20)$$

where  $\psi(\alpha) = d \log \{\Gamma(\alpha)\} / d\alpha$ .

**PROPOSITION 3.** (a) *For any conditional density of  $\beta$  given  $\alpha$ , the marginal reference prior for  $\alpha$  under Option 1 is given by*

$$\pi_1^r(\alpha) \propto \{\psi'(\alpha) - \alpha^{-1}\}^{\frac{1}{2}}. \quad (21)$$

(b) *Assume that the conditional distribution of  $\beta$  given  $\alpha$  is  $\text{Gamma}(a, b)$ . The marginal reference prior for  $\alpha$  under Option 2 is given by*

$$\pi_2^r(\alpha) \propto \{\psi'(\alpha) - n\psi'(n\alpha + a)\}^{\frac{1}{2}}. \quad (22)$$

(c) *For any given marginal prior on  $\alpha$ , the conditional reference prior for  $\beta$  is independent of  $\alpha$  and is given by  $\beta^{-1}$ .*

Proof. It is easy to see that  $|\Sigma| = \beta^{-2}\{\alpha\psi'(\alpha) - 1\}$  and  $\Sigma_{22} = \alpha\beta^{-2}$ . Thus  $|\Sigma|/\Sigma_{22} = \psi'(\alpha) - \alpha^{-1}$ , which implies (21). Part (c) is clear. For (b), let  $X_n = (x_1, \dots, x_n)$  be a random sample from a Gamma distribution. The marginal model is then

$$\begin{aligned} p(X_n|\alpha) &= \frac{(\prod_{i=1}^n x_i)^{\alpha-1}}{\{\Gamma(\alpha)\}^n} \int_0^\infty \beta^{n\alpha} \exp\left(-\beta \sum_{i=1}^n x_i\right) \frac{b^a}{\Gamma(a)} \beta^{a-1} \exp(-\beta b) d\beta \\ &\propto \frac{\Gamma(n\alpha + a)(\prod_{i=1}^n x_i)^\alpha}{\{\Gamma(\alpha)\}^n (\sum_{i=1}^n x_i + b)^{n\alpha+a}}. \end{aligned}$$

The second derivative of the logarithm of  $p(X_n|\alpha)$  is  $-n\psi'(\alpha) + n^2\psi'(n\alpha + a)$ , from which (b) follows immediately.  $\square$

Using an expansion for  $\psi'(\cdot)$ , (cf. Equation (1.45b) of Bowman and Shenton, 1988), we can show that  $\psi'(x) = x^{-1} + (2x^2)^{-1} + O(x^{-3})$  as  $x \rightarrow \infty$ . Therefore,  $n\psi'(n\alpha + a) \rightarrow \alpha^{-1}$  as  $n \rightarrow \infty$ . Consequently, the marginal reference prior for  $\alpha$  under Option 2 converges to the marginal reference prior for  $\alpha$  under Option 1.

### 3.4. NEYMAN-SCOTT PROBLEM

Suppose that  $y_{ij}, i = 1, \dots, n, j = 1, 2$ , are independent observations, and  $y_{ij}$  has a normal distribution with mean  $\mu_i$  and variance  $\sigma^2$ . We want to find the marginal reference prior distribution for  $\sigma^2$  or  $\sigma$ . The Fisher information matrix for  $(\mu_1, \dots, \mu_n, \sigma^2)$  is given by

$$\Sigma \equiv \Sigma(\mu_1, \dots, \mu_n, \sigma^2) = \text{diag}(2\sigma^{-2}, \dots, 2\sigma^{-2}, n\sigma^{-4}).$$

**PROPOSITION 4.** *For any conditional prior of  $(\mu_1, \dots, \mu_n)$  given  $\sigma^2$ , the marginal reference prior for  $\sigma^2$  under Option 1 is  $\pi_1^r(\sigma^2) \propto \sigma^{-2}$ . Equivalently, the marginal reference prior for  $\sigma$  is  $\pi_1^r(\sigma) \propto \sigma^{-1}$ .*

Proof. Let  $\theta_1 = \sigma^2$  and  $\theta_2 = (\mu_1, \dots, \mu_n)$ . We note that  $|\Sigma|/|\Sigma_{22}| = n\sigma^{-4}$ , which does

not depend on  $\theta_2$ . The first result follows from Theorem 1 (a). The second result is an immediate corollary.  $\square$

**PROPOSITION 5.** *Suppose that the prior distributions for  $\mu_i, i = 1, \dots, n$ , are independent normal with mean  $m_i$  and variance  $\tau^2$ . The marginal reference prior for  $\sigma^2$  under Option 2 then has the form*

$$\pi_2^r(\sigma^2) \propto \{\sigma^{-4} + (\sigma^2 + 2\tau^2)^{-2}\}^{\frac{1}{2}}. \quad (23)$$

*Equivalently, the marginal reference prior for  $\sigma$  under Option 2 has the form*

$$\pi_2^r(\sigma) \propto \{\sigma^{-2} + \sigma^2(\sigma^2 + 2\tau^2)^{-2}\}^{\frac{1}{2}}. \quad (24)$$

**Proof.** Define  $\bar{y}_i = (y_{i1} + y_{i2})/2$  and  $S^2 = \sum_{i=1}^n \sum_{j=1}^2 (y_{ij} - \bar{y}_i)^2$ . The marginal model is

$$\begin{aligned} p(X_n | \sigma^2) &\propto \frac{1}{\sigma^{2n}} \exp\left\{-\frac{S^2}{2\sigma^2}\right\} \prod_{i=1}^n \int_{-\infty}^{\infty} \exp\left\{-\frac{2(\bar{y}_i - \mu_i)^2}{2\sigma^2}\right\} \frac{1}{(2\pi)^{\frac{1}{2}}\tau} \exp\left\{-\frac{(\mu_i - m_i)^2}{2\tau^2}\right\} d\mu_i \\ &\propto \frac{1}{\sigma^n(\sigma^2 + 2\tau^2)^{\frac{1}{2}n}} \exp\left\{-\frac{S^2}{2\sigma^2} - \frac{\sum_{i=1}^n (\bar{y}_i - m_i)^2}{\sigma^2 + 2\tau^2}\right\}. \end{aligned}$$

Therefore, writing  $\tilde{\sigma} = \sigma^2$ , we have

$$\frac{\partial^2}{\partial \tilde{\sigma}^2} \log\{p(X_n | \tilde{\sigma})\} = \frac{n}{2\tilde{\sigma}^2} + \frac{n}{2(\tilde{\sigma}^2 + 2\tau^2)^2} - \frac{S^2}{\tilde{\sigma}^3} - \frac{2\sum_{i=1}^n (\bar{y}_i - m_i)^2}{(\tilde{\sigma} + 2\tau^2)^3}.$$

Under the marginal model, we have  $E_{\tilde{\sigma}} S^2 = n\tilde{\sigma}$  and  $E_{\tilde{\sigma}} (\bar{y}_i - m_i)^2 = \frac{1}{2}\tilde{\sigma} + \tau^2$ . So the Fisher information for  $\tilde{\sigma}$  under the marginal model is then

$$\Sigma^*(\tilde{\sigma}) = n/(2\tilde{\sigma}^2) + \frac{1}{2}n(\tilde{\sigma} + 2\tau^2)^2.$$

This proves the first assertion. The second assertion follows immediately.  $\square$

Interestingly,  $\pi_1^r(\sigma)$  and  $\pi_2^r(\sigma)$  remain substantially different because of the strong prior input on the  $\mu_i$  in Option 2 even if  $n \rightarrow \infty$ . This prior input weakens if we take  $\tau^2$

very large, and in that case  $\pi_1^r(\sigma)$  and  $\pi_2^r(\sigma)$  approximately agree. Also, for any given marginal prior on  $(\mu_1, \dots, \mu_n)$ , the conditional prior of  $\sigma$  is independent of  $(\mu_1, \dots, \mu_n)$ , and is the same as  $\pi_1^r(\sigma)$ .

### 3.5. BIVARIATE NORMAL DISTRIBUTION

Suppose that  $(Y_1, Y_2)$  has a bivariate normal distribution with unknown mean vector  $(\theta_1, \theta_2)$ , variances equal to one and known correlation  $\rho$ . The joint density of  $(Y_1, Y_2)$  is

$$p(y_1, y_2 | \theta_1, \theta_2) = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{(y_1 - \theta_1)^2 - 2\rho(y_1 - \theta_1)(y_2 - \theta_2) + (y_2 - \theta_2)^2}{2(1-\rho^2)}\right\}.$$

The Fisher information matrix is the inverse of the covariance matrix.

PROPOSITION 6. (a) For any conditional prior of  $\theta_2$  given  $\theta_1$ ,  $\pi_1^r(\theta_1) \propto 1$ .

(b) Suppose that  $\theta_1$  and  $\theta_2$  are independent. Then, for any subjective prior for  $\theta_2$ ,  $\pi_2^r(\theta_1) \propto 1$ .

Proof. Part (a) is an immediate corollary of Theorem 1 (a). For any subjective prior for  $\theta_2$ , say  $\pi^s(\theta_2)$ , the Fisher information of  $\theta$  from the marginal model is given by

$$\begin{aligned} \Sigma^*(\theta_1) &= E \left\{ \frac{\frac{\partial}{\partial \theta_1} \int_{-\infty}^{\infty} p(Y_1, Y_2 | \theta_1, \theta_2) \pi^s(\theta_2) d\theta_2}{\int_{-\infty}^{\infty} p(Y_1, Y_2 | \theta_1, \theta_2) \pi^s(\theta_2) d\theta_2} \right\}^2 \\ &= \int \int \frac{\left\{ \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_1} p(y_1, y_2 | \theta_1, \theta_2) \pi^s(\theta_2) d\theta_2 \right\}^2}{\int_{-\infty}^{\infty} p(y_1, y_2 | \theta_1, \theta_2) \pi^s(\theta_2) d\theta_2} dy_1 dy_2 \\ &= \int \int \frac{\left[ \int_{-\infty}^{\infty} p(y_1, y_2 | \theta_1, \theta_2) \{ (y_1 - \theta_1) + \rho(y_2 - \theta_2) \} \pi^s(\theta_2) d\theta_2 \right]^2}{\int_{-\infty}^{\infty} p(y_1, y_2 | \theta_1, \theta_2) \pi^s(\theta_2) d\theta_2} dy_1 dy_2. \end{aligned}$$

By making the transformation  $t_1 = y_1 - \theta_1$ , we know that the right hand side does not depend on  $\theta_1$  and the Fisher information for  $\theta_1$  is in fact a constant. The result follows immediately.  $\square$

In general, the marginal priors under Option 1 and Option 2 are different. Here is an example, where  $\pi^s(\theta_2 | \theta_1)$  does depend on  $\theta_1$ . Suppose that  $\rho = 0$  and assume

that  $\pi^s(\theta_2|\theta_1)$  has a  $N(\theta_1, \theta_1^2)$  distribution. It is easy to see that  $Y_1|\theta_1$  has a  $N(\theta_1, 1)$  distribution and, independently,  $Y_2|\theta_1$  has a  $N(\theta_1, 1 + \theta_1^2)$  distribution. The Fisher information based on the observation  $Y_1$  is 1. Furthermore,

$$\frac{\partial^2}{\partial \theta_1^2} \log\{p(Y_2|\theta_1)\} = -\frac{2}{1 + \theta_1^2} + \frac{2\theta_1^2}{(1 + \theta_1^2)^2} - \frac{4\theta_1(Y_2 - \theta_1)}{(1 + \theta_1^2)^2} - \frac{4\theta_1^2(Y_2 - \theta_1)^2}{(1 + \theta_1^2)^3} + \frac{(Y_2 - \theta_1)^2}{(1 + \theta_1^2)^2}.$$

The Fisher information based on the observation  $Y_2$  is then  $(1 + \theta_1^2)^{-1} + 2\theta_1^2/(1 + \theta_1^2)^2$ .

Therefore,

$$\Sigma^*(\theta_1) = 1 + \frac{1}{1 + \theta_1^2} + \frac{2\theta_1^2}{(1 + \theta_1^2)^2}.$$

Clearly, the reference priors under Option 1 and Option 2 are thus different.

### 3.6. BETA DISTRIBUTION

Consider the beta density

$$p(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad (25)$$

where  $\alpha > 0$  and  $\beta > 0$  are unknown parameters. The Fisher information matrix for  $(\alpha, \beta)$  is

$$\Sigma = \Sigma(\alpha, \beta) = \begin{pmatrix} G(\alpha) - G(\alpha + \beta) & -G(\alpha + \beta) \\ -G(\alpha + \beta) & G(\beta) - G(\alpha + \beta) \end{pmatrix}, \quad (26)$$

where  $G(\alpha) = \frac{d^2}{d\alpha^2} \log\{\Gamma(\alpha)\}$  is the poly-gamma function. A computational formula for  $G$  is  $G(x) = \sum_{j=0}^{\infty} (x+j)^{-2}$  (Bowman and Shenton, 1988).

We now try to find the conditional reference prior of  $\beta$  given  $\alpha$ . Note that the Fisher information matrix does not satisfy the condition in Theorem 2. We will see that the conclusion of Theorem 2 fails in this case. Let  $l_i < u_i$  be two sequences of constants satisfying

$$l_i \rightarrow 0 \text{ and } u_i \rightarrow \infty. \quad (27)$$

Let  $K_i(\alpha) = \int_{l_i}^{u_i} \{G(\beta) - G(\alpha + \beta)\}^{\frac{1}{2}} d\beta$ . When  $\alpha = 1$ , it is easy to show that  $G(\beta) - G(\beta + 1) = \beta^{-2}$  and  $K_i(1) = \int_{l_i}^{u_i} \beta^{-1} d\beta = \log(u_i) - \log(l_i)$ . For an arbitrary  $\alpha > 0$ , exact computation of  $K_i(\alpha)$  is quite complicated, but we have the following expansion.

LEMMA 1. For fixed  $\alpha$  and as  $i \rightarrow \infty$ ,

$$K_i(\alpha) = \sqrt{\alpha} \log(u_i) - \log(l_i) + O(1),$$

where  $O(1)$  is a bounded constant.

Proof. See the Appendix. □

PROPOSITION 7. Assume that (27) holds and that  $u_i l_i \rightarrow 1$  as  $i \rightarrow \infty$ . For any given marginal prior of  $\alpha$ , the conditional reference prior of  $\beta$  given  $\alpha$  is

$$\pi^r(\beta|\alpha) \propto \frac{1}{\sqrt{\alpha + 1}} \{G(\beta) - G(\alpha + \beta)\}^{\frac{1}{2}}.$$

Proof. From Lemma 1,

$$\lim_{i \rightarrow \infty} \frac{K_i(1)}{K_i(\alpha)} = \lim_{i \rightarrow \infty} \frac{\log(u_i) - \log(l_i)}{\sqrt{\alpha} \log(u_i) - \log(l_i) + O(1)} = \frac{2}{\sqrt{\alpha + 1}}.$$

The result then follows from (11). □

This fact illustrates that, when  $|\Sigma_{22}(\theta_1, \theta_2)|$  does not have the form (12),  $\pi^r(\theta_2|\theta_1)$  may not be proportional to  $|\Sigma_{22}(\theta_1, \theta_2)|^{\frac{1}{2}}$ . Furthermore,  $\pi^r(\theta_2|\theta_1)$  may depend on the choice of the compact supports  $[l_i, u_i]$ . For example, if  $l_i = u_i^{-\sqrt{\alpha}}$ , then  $\pi^r(\beta|\alpha) \propto \alpha^{-\frac{1}{2}} \{G(\beta) - G(\alpha + \beta)\}^{\frac{1}{2}}$ , although such a choice of the compact sets would be rather unusual.

#### 4. WHEN TWO PARAMETERS ARE KNOWN TO BE INDEPENDENT



#### 4.1. BASIC ALGORITHM

Other types of possible partial information may be available. For example, one might believe that  $\theta_1$  and  $\theta_2$  are independent. Then one wants, as a reference prior, the product of marginal reference priors,  $\pi_1^r(\theta_1)$  and  $\pi_2^r(\theta_2)$ . It is not clear how to define these, but Option 1 in Section 2.1 suggests the following iterative algorithm.

Step 0. Choose any initial nonzero marginal prior density for  $\theta_2$ , say  $\pi_2^{(0)}(\theta_2)$ .

Step 1. Define an interim prior density for  $\theta_1$  by

$$\pi_1^{(1)}(\theta_1) \propto \exp\left\{\frac{1}{2} \int \pi_2^{(0)}(\theta_2) \log(|\Sigma|/|\Sigma_{22}|) d\theta_2\right\}.$$

Step 2. Define an interim prior density for  $\theta_2$  by

$$\pi_2^{(1)}(\theta_2) \propto \exp\left\{\frac{1}{2} \int \pi_1^{(1)}(\theta_1) \log(|\Sigma|/|\Sigma_{11}|) d\theta_1\right\}.$$

Now replace  $\pi_2^{(0)}$  in *Step 0* by  $\pi_2^{(1)}$  and repeat *Step 1* and *Step 2*, to obtain  $\pi_1^{(2)}$  and  $\pi_2^{(2)}$ . Consequently, we generate two sequences  $\{\pi_1^{(i)}\}_{i \geq 1}$  and  $\{\pi_2^{(i)}\}_{i \geq 1}$ . The desired marginal priors will be the limits  $\pi_j^r = \lim_{i \rightarrow \infty} \pi_j^{(i)}$ ,  $j = 1, 2$ , if the limits exist. In applying the iterative algorithm, it may be necessary to operate on compact sets, and then let the sets grow.

We do not know the extent to which this algorithm converges in general. We have studied several specific situations, and convergence was achieved quickly. For instance, in the two-parameter Weibull model the equations iterate to the usual reference prior given in Sun (1997). It would clearly be of interest to establish conditions under which convergence is guaranteed. For many important situations, it is possible to deduce the result of the above algorithm directly without actually going through the iterations. Here are two sufficient conditions under which this can be done.

**THEOREM 3.** (a) If  $|\Sigma|/|\Sigma_{22}|$  does not depend on  $\theta_2$ , then the marginal reference priors are

$$\pi_1^r(\theta_1) \propto (|\Sigma|/|\Sigma_{22}|)^{\frac{1}{2}} \quad \text{and} \quad \pi_2^r(\theta_2) \propto \exp\left\{\frac{1}{2} \int \pi_1^r(\theta_1) \log(|\Sigma|/|\Sigma_{11}|) d\theta_1\right\}. \quad (28)$$

(b) If  $|\Sigma|/|\Sigma_{11}|$  does not depend on  $\theta_1$ , then the marginal reference priors are

$$\pi_2^r(\theta_2) \propto (|\Sigma|/|\Sigma_{11}|)^{\frac{1}{2}} \quad \text{and} \quad \pi_1^r(\theta_1) \propto \exp\left\{\frac{1}{2} \int \pi_2^r(\theta_2) \log(|\Sigma|/|\Sigma_{22}|) d\theta_2\right\}. \quad (29)$$

*Proof.* Under the assumption in (a),  $\pi_1^r(\theta_1)$  does not depend on the choice of  $\pi_2^{(0)}$  in *Step 0*. □

The reference priors under the independence assumption are, in general, different from the reference prior or the reverse reference prior (Berger and Bernardo, 1992). The following result gives a condition under which they are the same. Its proof is obvious, and is omitted.

**THEOREM 4.** If the Fisher information matrix of  $(\theta_1, \theta_2)$  is of the form

$$\Sigma(\theta_1, \theta_2) = \text{diag}\{g_1(\theta_1)h_1(\theta_2), g_2(\theta_1)h_2(\theta_2)\},$$

then the independent marginal reference priors are

$$\pi_1^r(\theta_1) \propto \{g_1(\theta_1)\}^{\frac{1}{2}} \quad \text{and} \quad \pi_2^r(\theta_2) \propto \{h_2(\theta_2)\}^{\frac{1}{2}}. \quad (30)$$

Under the conditions of the theorem, when either  $\theta_1$  or  $\theta_2$  is the parameter of interest, the reference priors have the same form:  $\pi(\theta_1, \theta_2) \propto \{g_1(\theta_1)h_2(\theta_2)\}^{\frac{1}{2}}$  (cf. Datta and Ghosh, 1995). Therefore, the reference prior and the reverse reference prior are also as in (30).

#### 4.2. EXAMPLES FOR INDEPENDENT PRIORS

Example 1: *normal distribution*. Clearly, when  $\mu$  and  $\sigma$  are independent, the marginal reference priors are  $\pi_1^r(\mu) \propto 1$  and  $\pi_2^r(\sigma) \propto 1/\sigma$ .

Example 2: *gamma distribution*. It is easy to see that  $|\Sigma|/|\Sigma_{22}|$  does not depend on  $\beta$ . From Theorem 3, the marginal reference priors are  $\pi_1^r(\alpha) \propto \{\psi'(\alpha) - 1/\alpha\}^{\frac{1}{2}}$  and

$$\pi_2^r(\beta) \propto \exp\left[\frac{1}{2} \int \{\psi'(\alpha) - \alpha^{-1}\}^{\frac{1}{2}} \log\left\{\frac{\alpha\psi'(\alpha) - 1}{\beta^2\phi'(\alpha)}\right\} d\alpha\right] \propto \beta^{-1}.$$

This is also the unrestricted reference prior when  $\alpha$  is the parameter of interest (Sun and Ye, 1996).

Example 3: *bivariate binomial distribution*. Crowder and Sweeting (1989) consider the following bivariate binomial distribution, whose probability density is given by

$$f(r, s|p, q) = \binom{m}{r} p^r (1-p)^{m-r} \binom{r}{s} q^s (1-q)^{r-s},$$

where  $0 < p, q < 1$ , and  $s$  and  $r$  are nonnegative integers satisfying  $0 \leq s \leq r \leq m$ . The Fisher information matrix for  $(p, q)$  is given by

$$\Sigma = m \text{diag}[\{p(1-p)\}^{-1}, p\{q(1-q)\}^{-1}].$$

Clearly, the Jeffreys prior is proportional to  $\{(1-p)q(1-q)\}^{-\frac{1}{2}}$ . Based on the assumptions that  $p$  and  $q$  are independent, that  $\theta = pq$  and  $\phi = p(1-q)(1-pq)^{-1}$  are independent, and some invariance considerations, Crowder and Sweeting (1989) derived the noninformative prior,  $\pi_{CS}(p, q) \propto \{p(1-p)q(1-q)\}^{-1}$ . Polson and Wasserman (1990) derived, as the reference prior when either  $p$  or  $q$  is the parameter of interest,  $\pi_r(p, q) \propto \{p(1-p)q(1-q)\}^{-\frac{1}{2}}$ . From Theorem 3, this is also the reference prior based on independence of  $p$  and  $q$ .

### 4.3. A CLINICAL TRIAL: ECMO

Ware (1989) considered a Bayesian solution of a clinical trial. Ten patients were given standard therapy and six survived. On the other hand, nine patients were treated with ECMO (extra corporeal membrane oxygenation) and all nine survived. Let  $p_1$  be the probability of success under standard therapy and  $p_2$  be the probability of success under ECMO. It is desired to compare the two treatments. Let  $\eta_i = \log\{p_i/(1 - p_i)\}$ ,  $i = 1, 2$ , and  $\delta = \eta_2 - \eta_1$ . The quantity of interest is then the posterior probability that  $\delta > 0$ , where  $\eta_1$  is a nuisance parameter. This example was reanalyzed by Kass and Greenhouse (1989), who considered 84 different proper prior distributions, all involving the independence assumption. They said that the independence assumption is somewhat subtle and reasonable.

A follow-up to Kass and Greenhouse's study was given in Lavine *et al.* (1991), who studied bounds on the posterior probability that  $\delta > 0$  under various priors with and without the independence constraint. Berger and Moreno (1994) also treated the example from a robust Bayesian viewpoint. Lavine *et al.* (1991) and Berger and Moreno (1994) all showed that, without the independence assumption, the infima of the posterior probability that  $\delta > 0$  for a reasonable class of priors might be very small. They also thus suggested use of the independence assumption (assuming, of course, that it was plausible in the application).

For this problem, both the Jeffreys prior and the Berger and Bernardo (1992) reference prior will give a dependent prior for  $\delta$  and  $\eta_1$ . We now derive the reference prior under the independence assumption. First, the Fisher information matrix of  $(p_1, p_2)$  is

$$\Sigma(p_1, p_2) = \text{diag}[n_1/\{p_1(1 - p_1)\}, n_2/\{p_2(1 - p_2)\}],$$

where  $n_1 = 10$  and  $n_2 = 9$ . Thus the Fisher information matrix of  $(\eta_1, \delta)$  is given by

$$\begin{pmatrix} \frac{n_1 e^{\eta_1}}{(1+e^{\eta_1})^2} + \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2} & \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2} \\ \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2} & \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2} \end{pmatrix}. \quad (31)$$

We can now apply Theorem 3, because  $|\Sigma|/|\Sigma_{22}| = n_1 e^{\eta_1}/(1+e^{\eta_1})^2$  is independent of  $\delta$ .

Note that the two marginal priors are proper, so it is not necessary to use a compact set limiting argument for the derivation.

PROPOSITION 8. (a) *Under the constraint that the marginal priors for  $\delta$  and  $\eta_1$  be independent, the marginal reference priors are of the form*

$$\pi_1^r(\eta_1) = e^{\eta_1/2}/\{\pi(1+e^{\eta_1})\}; \quad (32)$$

$$\pi_2^r(\delta) \propto \exp\left(-\frac{1}{2\pi} \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} \log\left[1 + \frac{n_1}{n_2} \{(1-t)e^{-\delta/2} + te^{\delta/2}\}^2\right] dt\right). \quad (33)$$

(b) *The marginal densities in (a) are both symmetric and  $\exp(\eta_1/2)$  has a folded Cauchy  $(0,1)$  density.*

Proof. The density in (32) follows from the fact that

$$\int_{-\infty}^{\infty} \frac{e^{\eta_1/2}}{1+e^{\eta_1}} d\eta_1 = \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} dt = \{\Gamma(\frac{1}{2})\}^2 = \pi.$$

For (33),

$$\begin{aligned} \pi_2^r(\delta) &\propto \exp\left[-\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\eta_1/2}}{1+e^{\eta_1}} \log\left\{\frac{\frac{n_1 e^{\eta_1}}{(1+e^{\eta_1})^2} + \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2}}{\frac{n_1 e^{\eta_1}}{(1+e^{\eta_1})^2} \frac{n_2 e^{\eta_1+\delta}}{(1+e^{\eta_1+\delta})^2}}\right\} d\eta_1\right] \\ &= \exp\left[-\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\eta_1/2}}{1+e^{\eta_1}} \log\left\{\frac{n_1 e^{\eta_1} (1+e^{\eta_1+\delta})^2 + n_2 e^{\eta_1+\delta} (1+e^{\eta_1})^2}{e^{\eta_1} e^{\eta_1+\delta}}\right\} d\eta_1\right] \\ &= \exp\left[-\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\eta_1/2}}{1+e^{\eta_1}} \log\left\{n_1 e^{-(\eta_1+\delta)} (1+e^{\eta_1+\delta})^2 + n_2 e^{-\eta_1} (1+e^{\eta_1})^2\right\} d\eta_1\right]. \end{aligned}$$

Making the transformation  $t = e^{\eta_1}/(1+e^{\eta_1})$ , we have

$$\pi_2^r(\delta) \propto \exp\left[-\frac{1}{2\pi} \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} \log\left\{\frac{n_1(1-t)}{te^\delta} \left(1 + \frac{te^\delta}{1-t}\right)^2 + \frac{n_2(1-t)}{t} \left(1 + \frac{t}{1-t}\right)^2\right\} dt\right]$$

$$\begin{aligned}
&= \exp\left(-\frac{1}{2\pi} \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} \log\left[\frac{n_1\{(1-t)e^{-\delta/2} + te^{\delta/2}\}^2 + n_2}{t(1-t)}\right] dt\right) \\
&\propto \exp\left(-\frac{1}{2\pi} \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} \log\left[n_1\{(1-t)e^{-\delta/2} + te^{\delta/2}\}^2 + n_2\right] dt\right).
\end{aligned}$$

Formula (33) then follows. Part (b) is clear. □

Kass and Greenhouse (1989) found that the posterior probability  $P(\delta > 0|data)$  is approximately 0.95 based on the independence of the proper prior they favoured. For their independent prior in the  $\delta$  and  $\eta_1$  parameterization,  $P(\delta > 0|data)$  was approximately 0.99. Figure 1 compares the independent reference prior density and the resulting posterior density of  $\delta$ . The resulting posterior probability that  $\delta > 0$  is about 0.99. It is interesting that the noninformative prior analysis yields the same conclusion as the Kass and Greenhouse (1989) subjective analysis for the same parameterization, even though it can be shown that the reference priors are considerably more diffuse than the subjective priors of Kass and Greenhouse. Note finally that, even though  $\pi_2^r(\delta)$  can be expressed only in terms of an integral, this is not a problem in that computation must be done by Monte Carlo integration in any case.

## 5. DISCUSSION

We have proposed two options to find the marginal reference prior for  $\theta_1$  when the conditional prior for  $\theta_2$  is known. Option 2 was felt to be the most natural approach, but difficulty in its implementation will usually necessitate use of the easy Option 1. Table 1 summarizes, for the examples in this paper, when the two options are known to yield the same answer. Note, however, that, for all examples considered in which the two options give different answers with the exception of the Neyman-Scott problem,  $\pi_1^r(\cdot)$  and  $\pi_2^r(\cdot)$  agree asymptotically, as  $n \rightarrow \infty$ . This lends further support to general

use of the simple Option 1. In the Neyman-Scott problem, the two marginal reference priors do remain different as  $n \rightarrow \infty$ , but, since the number of unknown parameters grows with  $n$ , this is perhaps not unexpected.

The conditional reference prior,  $\pi^r(\theta_2|\theta_1)$ , is usually given by (10), and does not then depend on the specified marginal prior for  $\theta_1$ . However, as shown in the example of the beta distribution in Section 3.6,  $\pi^r(\theta_2|\theta_1)$  can differ from (10).

In dealing with the partial prior knowledge that  $\theta_1$  and  $\theta_2$  are independent, an iterative application of the reference algorithm was proposed. While this can be trivially implemented in many important special cases, its general applicability and convergence require further study.

#### APPENDIX. PROOF OF LEMMA 1

Let

$$J_\alpha(\beta) = \sum_{j=1}^{\infty} \{(\beta + j)^{-2} - (\alpha + \beta + j)^{-2}\}.$$

Then  $G(\beta) - G(\alpha + \beta) = \beta^{-2} - (\alpha + \beta)^{-2} + J_\alpha(\beta) = \alpha(\alpha + 2\beta)\{\beta^2(\alpha + \beta)^2\}^{-1} + J_\alpha(\beta)$ . Define  $h(x) = (\beta + x)^{-2} - (\alpha + \beta + x)^{-2}$ , for  $x \geq 0$ . Since  $h'(x) = -2\{(\beta + x)^{-3} - (\alpha + \beta + x)^{-3}\} < 0$  for  $x > 0$ ,  $h(x)$  is a monotone decreasing function of  $x$ . Thus,

$$\begin{aligned} J_\alpha(\beta) &\geq \int_1^\infty \{(\beta + x)^{-2} - (\alpha + \beta + x)^{-2}\} dx \\ &= (\beta + 1)^{-1} - (\alpha + \beta + 1)^{-1} \\ &= \alpha\{(\beta + 1)(\alpha + \beta + 1)\}^{-1}. \end{aligned}$$

On the other hand, for any  $x \geq 1$  and  $\alpha > 0$ ,

$$\left\{ \frac{1}{x^2} - \frac{1}{(x + \alpha)^2} \right\} - \left\{ \frac{1}{x^2 - \frac{1}{4}} - \frac{1}{(x + \alpha)^2 - \frac{1}{4}} \right\}$$

$$= \{(x + \alpha)^2 - x^2\} \left[ \frac{1}{x^2(x + \alpha)^2} - \frac{1}{(x^2 - \frac{1}{4})\{(x + \alpha)^2 - \frac{1}{4}\}} \right],$$

which is negative. Thus, for any  $j \geq 1$ ,

$$\begin{aligned} \frac{1}{(\beta + j)^2} - \frac{1}{(\alpha + \beta + j)^2} &\leq \frac{1}{(\beta + j)^2 - \frac{1}{4}} - \frac{1}{(\alpha + \beta + j)^2 - \frac{1}{4}} \\ &= \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} \left\{ \frac{1}{(\beta + x)^2} - \frac{1}{(\alpha + \beta + x)^2} \right\} dx. \end{aligned}$$

Therefore,

$$J_\alpha(\beta) \leq \int_{\frac{1}{2}}^{\infty} \left\{ \frac{1}{(\beta + x)^2} - \frac{1}{(\alpha + \beta + x)^2} \right\} dx = \frac{\alpha}{(\beta + \frac{1}{2})(\alpha + \beta + \frac{1}{2})}.$$

Consequently,  $H_i^L(\alpha) \leq K_i(\alpha) \leq H_i^U(\alpha)$ , where

$$\begin{aligned} H_i^L(\alpha) &= \sqrt{\alpha} \int_{l_i}^{u_i} \left\{ \frac{\alpha + 2\beta}{\beta^2(\alpha + \beta)^2} + \frac{1}{(\beta + 1)(\alpha + \beta + 1)} \right\}^{\frac{1}{2}} d\beta; \\ H_i^U(\alpha) &= \sqrt{\alpha} \int_{l_i}^{u_i} \left\{ \frac{\alpha + 2\beta}{\beta^2(\alpha + \beta)^2} + \frac{1}{(\beta + \frac{1}{2})(\alpha + \beta + \frac{1}{2})} \right\}^{\frac{1}{2}} d\beta. \end{aligned}$$

For any  $0 < \epsilon < 1$  small enough and  $i$  large enough such that  $l_i < \epsilon < 1/\epsilon < u_i$ , we have

$$\begin{aligned} H_i^U(\alpha) &= \sqrt{\alpha} \left( \int_{l_i}^{\epsilon} \frac{1}{\beta} \left\{ \frac{\alpha + 2\beta}{(\alpha + \beta)^2} + \frac{\beta^2}{(\frac{1}{2} + \beta)(\alpha + \frac{1}{2} + \beta)} \right\}^{\frac{1}{2}} d\beta + O(1) \right. \\ &\quad \left. + \int_{1/\epsilon}^{u_i} \frac{1}{\beta} \left[ \frac{2 + \alpha\beta^{-1}}{\beta(1 + \alpha\beta^{-1})^2} + \frac{1}{\{1 + \frac{1}{2}\beta^{-1}\}\{1 + (\alpha + \frac{1}{2})\beta^{-1}\}} \right]^{\frac{1}{2}} d\beta \right) \\ &= \sqrt{\alpha} \left[ \int_{l_i}^{\epsilon} \frac{1}{\sqrt{\alpha\beta}} \{1 + O(\beta)\} d\beta + O(1) + \int_{1/\epsilon}^{u_i} \frac{1}{\beta} \{1 + O(\frac{1}{\beta})\} d\beta \right] \\ &= \sqrt{\alpha} \log(u_i) - \log(l_i) + O(1). \end{aligned}$$

Similarly, we have  $H_i^L(\alpha) = \sqrt{\alpha} \log(u_i) - \log(l_i) + O(1)$ . This completes the proof.

#### ACKNOWLEDGEMENTS

Sun's research was partially supported by a National Security Agency grant and a Research Board award from the University of Missouri-System. Berger's research was



supported under a National Science Foundation grant. The manuscript was prepared using computer facilities supported in part by a National Science Foundation awarded to the Department of Statistics at the University of Missouri–Columbia. The authors gratefully acknowledge the very constructive comments of the editors (Professors A.P. Dawid and D.M. Titterington), an associate editor, and two referees.

#### REFERENCES

- BERGER, J. O. AND BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Statist. Assoc.*, **84**, 200-207.
- BERGER, J. O. AND BERNARDO, J. M. (1992). On the development of reference priors (with Discussion). In *Bayesian Statistics, 4*, Eds. J. M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, pp 35-60, Oxford University Press.
- BERGER, J. O. AND MORENO, E. (1994). Bayesian robustness in bidimensional models: prior independence. *J. Statistical Planning and Inference*, **40**, 161-176.
- BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B*, **41**, 113-147.
- BOWMAN, K.O. AND SHENTON, L.R. (1988). *Properties of Estimators for the Gamma Distribution*. New York: Marcel Dekker.
- CROWDER, M. AND SWEETING, T. (1989). Bayesian inference for a bivariate binomial distribution. *Biometrika*, **76**, 599-603.
- DATTA, G.S. AND GHOSH, M. (1995). Some remarks on noninformative priors. *J. Am. Statist. Assoc.*, **90**, 1357-1363.
- GHOSH, J.K. AND MUKERJEE, R. (1992). Noninformative priors (with Discussion).

- In *Bayesian Statistics*, **4**, Eds. J.M. Bernardo, J.O. Berger, A.P Dawid and A.F.M. Smith, pp 195-210, Oxford Univ. Press.
- KASS, R.E. AND GREENHOUSE, J. (1989). "Investigating therapies of potentially great benefit: a Bayesian perspective," comment on "Investigating therapies of potentially great benefit: ECMO," by J.H. Ware, *Statistical Science*, **4**, 310-314.
- KASS, R.E. AND WASSERMAN, L. (1996). The selection of prior distributions by formal rules, *J. Am. Statist. Assoc.*, **91**, 1343-1370.
- LAVINE, M., WASSERMAN, L., AND WOLPERT, R.L. (1991). Bayesian inference with specified prior marginals. *J. Am. Statist. Assoc.*, **86**, 964-971.
- POLSON, N. AND WASSERMAN, L. (1990). Prior distributions for the bivariate binomial. *Biometrika*, **77**, 901-904.
- Sun, D. (1997). A note on noninformative priors for Weibull distributions. *Journal of Statistical Planning and Inference*, in press.
- Sun, D. and Ye, K. (1996). Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika*, **83**, 55-65.
- WARE, J. H. (1989). Investigating therapies of potentially great benefit: ECMO (with Discussion). *Statistical Science*, **4**, 298-340.

Table 1: Comparison of Marginal Reference Priors from Two Options

Distribution	$(\theta_1, \theta_2)$	$\pi_2^r(\theta_1) = \pi_1^r(\theta_1)$ ?
Multinomial $(p_1, p_2)$	$(p_1, p_2)$	Yes, if (16) holds
Normal $(\mu, \sigma^2)$	$(\mu, \sigma)$	Yes, if $\pi^s(\sigma \mu) = \pi^s(\sigma)$
	$(\sigma, \mu)$	No
Gamma $(\alpha, \beta)$	$(\alpha, \beta)$	No
Neyman–Scott	$\{\sigma, (\mu_1, \dots, \mu_n)\}$	No
Normal <sub>2</sub> $(\theta_1, \theta_2, 1, 1, \rho)$ $\rho$ is known	$(\theta_1, \theta_2)$	Yes, if $\theta_1$ and $\theta_2$ are independent

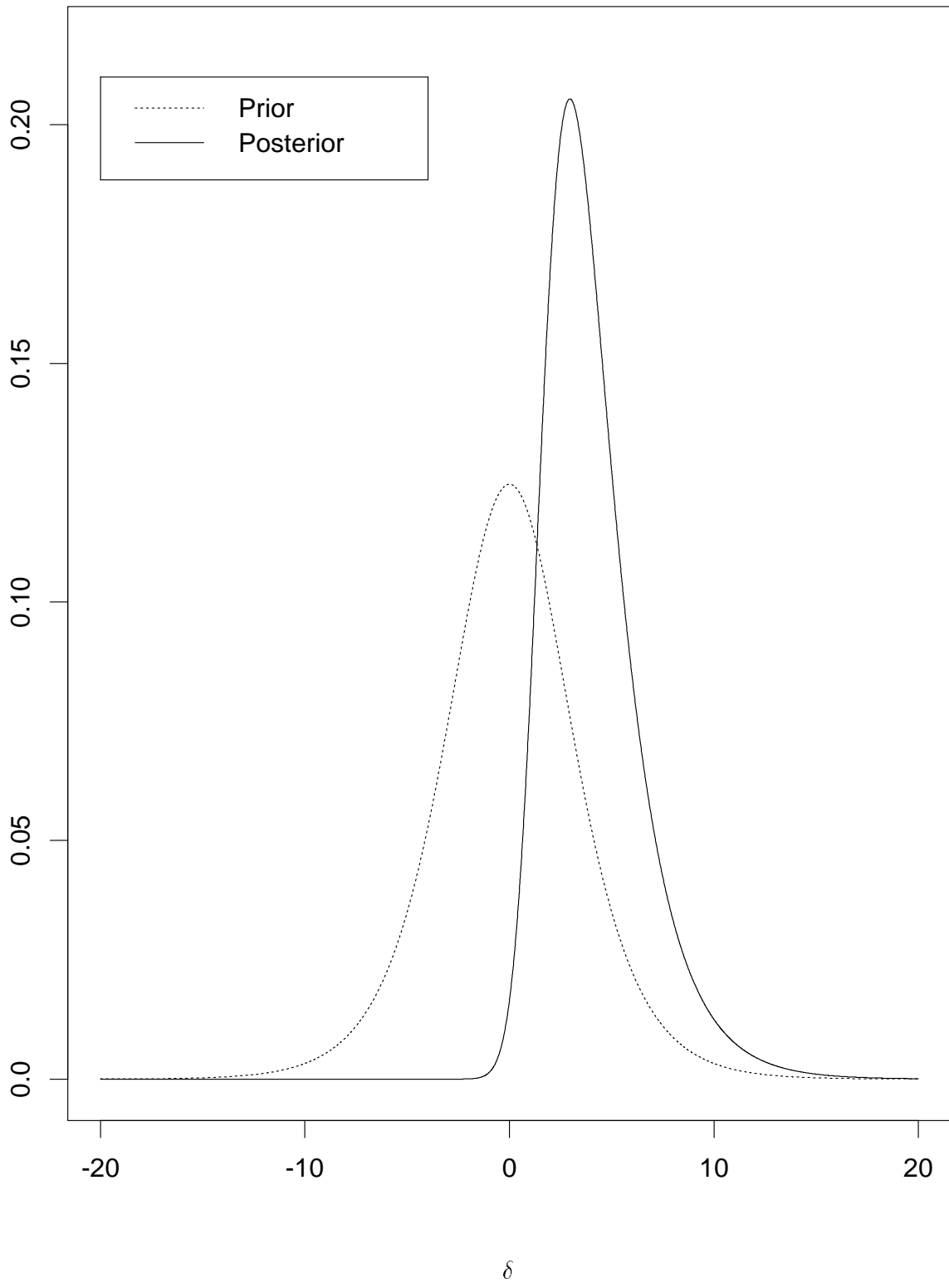


Fig. 1: Prior and Posterior Densities of  $\delta$  from ECMO data