REFERENCE PRIORS
IN A VARIANCE COMPONENTS PROBLEM

by

James O. Berger   and   José M. Bernardo
Purdue and Duke Universities   Generalidad Valenciana
West Lafayette, IN, USA   Valenciana, SPAIN

Technical Report # 89-32C

Δ

Reference Priors
in a Variance Components Problem

JAMES O. BERGER and JOSÉ M. BERNARDO
Purdue and Duke Universities, USA, and Generalidad Valenciana, Spain

SUMMARY

The ordered group reference prior algorithm of Berger and Bernardo (1989b) is applied to the balanced variance components problem. Besides the intrinsic interest of developing good noninformative priors for the variance components problem, a number of theoretically interesting issues arise in application of the proposed procedure. The algorithm is described (for completeness) in an important special case, with a detailed heuristic motivation.

Keywords: NONINFORMATIVE PRIORS; MULTIPARAMETER; BAYESIAN INFERENCE.

1. INTRODUCTION

Determination of reasonable noninformative priors in multiparameter problems is not easy; common noninformative priors, such as Jeffreys's prior, can have features that have an unexpectedly dramatic effect on the posterior. In recognition of this problem, Bernardo (1979), proposed the *reference prior* approach to development of noninformative priors, the key feature of which was a possible dependence of the reference prior on specification of parameters of interest and nuisance parameters.

This approach was further extended in Berger and Bernardo (1989a, 1989b). The first paper introduced a technically important modification: the reference prior algorithm was utilized in two stages —first for the nuisance parameter and then for the parameter of interest, as in Bernardo (1979)— but it was applied on an increasing family of compact subspaces of the parameter space, allowing a crucial first stage prior normalization to be performed. The second paper greatly extended this idea to deal with multiple nuisance parameters and parameters of interest, allowing for iterative application of the reference prior algorithm to any sequence of groupings of the parameters. This paper also contained examples and extensive discussion concerning how and when to group the coordinates.

An interesting situation in which to illustrate the new algorithm is the balanced variance components model, which has for parameters a mean μ, and two variances τ^2 and σ^2; (see Section 4 for specifics.) The interest in considering this situation, besides the basic importance of the model in statistics, is that all the possible grouped reference priors can be determined

and compared, and that interesting technical issues arise in application of the algorithm. The grouped reference priors are given in Section 4, along with a detailed derivation in the most difficult (and theoretically interesting) case. This section also gives a useful calculational expression for the ensuing posterior moments, an expression requiring only one-dimensional numerical integration.

Section 2 of the paper presents needed notation and results about patterned information matrices. Section 3 reviews a special case of the grouped reference prior algorithm that was developed in Berger and Bernardo (1989b), and presents a detailed heuristic motivation for the algorithm. Section 5 presents concluding comments.

2. NOTATION AND PRELIMINARIES

The general algorithm will be presented for any parametric statistical problem in which the random observation X has density $p(x|\theta)$, where $\theta \in \Theta \subset R^k$ is the unknown parameter. We assume that the Fisher information matrix

$$H(\theta) = -E_{x|\theta}\left[\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log p(x|\theta)\right)\right]$$

exists and has rank k, so that

$$S(\theta) = H^{-1}(\theta)$$

also exists. Often, we will just write H and S.

We assume that the θ_i are separated into m groups of sizes $n_1, n_2, \ldots, n_m$, and that these groups are given by

$$\theta_{(1)} = (\theta_1, \ldots, \theta_{n_1}), \quad \theta_{(2)} = (\theta_{n_1+1}, \ldots, \theta_{n_1+n_2}), \ldots$$
$$\theta_{(i)} = (\theta_{N_{i-1}+1}, \ldots, \theta_{N_i}), \ldots \quad \theta_{(m)} = (\theta_{N_{m-1}+1}, \ldots, \theta_k),$$

where $N_j = \sum_{i=1}^{j} n_i$ for $j = 1, \ldots, m$. These are the groupings to which the reference prior algorithm will be applied. (The coordinates of θ can, of course, be reordered if necessary —see Section 3.2— to achieve the desired ordered grouping.) Also we shall define, for $j = 1, \ldots, m$,

$$\theta_{[j]} = (\theta_{(1)}, \ldots, \theta_{(j)}) \quad \text{and} \quad \theta_{[\sim j]} = (\theta_{(j+1)}, \ldots, \theta_{(m)}).$$

If we write S as

$$S = \begin{pmatrix} A_{11} & A_{21}^t & \cdots & A_{m1}^t \\ A_{21} & A_{22} & \cdots & A_{m2}^t \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{pmatrix}$$

so that A_{ij} is $(n_i \times n_j)$, and define

$S_j \equiv$ upper left $(N_j \times N_j)$ corner of S, with $S_m \equiv S$, and
$H_j \equiv S_j^{-1}$

then, the matrices

$h_j \equiv$ lower right $(n_j \times n_j)$ corner of H_j, $j = 1, \ldots, m$

will be of central importance. Note that, if one defines $B_j = (A_{j1}A_{j2}\cdots A_{j\,j-1})$ for $j = 2,\ldots,m$, of sizes $(n_j \times N_{j-1})$, then it is straightforward to verify that, for $j = 1,\ldots,m$

$$h_j = (A_{jj} - B_j H_{j-1} B_j^t)^{-1} \tag{2.1}$$

and

$$H_j = \begin{pmatrix} H_{j-1} + H_{j-1}B_j^t h_j B_j H_{j-1} & -H_{j-1}B_j^t h_j \\ -h_j B_j H_{j-1} & h_j \end{pmatrix}, \tag{2.2}$$

where any entry containing a factor of H_0 is to be omitted. Thus, one may iteratively calculate $H_1,\ldots,H_m$, and hence $h_1,\ldots,h_m$.

In the important special case where each $n_j = 1$, no matrix inversions are needed above, so that calculation of the h_j is trivial if S is available. An even greater simplification occurs if, in addition,

$$B_{i+1} = (c_i B_i, \quad A_{i+1\,i}) \tag{2.3}$$

for some constant c_i. Then, (2.1), (2.2), and (2.3) can be used to show that

$$h_{i+1} = \left[A_{i+1\,i+1} + c_i^2 A_{ii} - 2c_i A_{i+1\,i} - h_i (c_i A_{ii} - A_{i+1\,i})^2 \right]^{-1}. \tag{2.4}$$

This is particularly useful when (2.3) holds for all i, which often occurs in patterned covariance matrices, since then (2.4) can be used to iteratively determine all the h_i, starting with $h_1 = A_{11}^{-1}$, and defining $c_1 = 1$. Finally, if S is a block diagonal matrix, (i.e., $A_{ij} \equiv 0$ for all $i \neq j$) then $h_j \equiv A_{jj}^{-1}$, $j = 1,\ldots,m$.

We will use the common symbols

$$|A| = \text{determinant of } A. \quad 1_\Omega(y) = \begin{cases} 1 & \text{if } y \in \Omega \\ 0 & \text{otherwise,} \end{cases}$$

and will throughout the paper adopt the conventions that $\sum_{i=l}^{l-1}(\cdot) = 0$ and $\prod_{i=l}^{l-1}(\cdot) = 1$. Also, we will often use $p(u \mid v)$ to generically represent the conditional density of u given v.

3. THE m-GROUP REFERENCE PRIOR

3.1 *The Algorithm*

We suppose the θ_i have been ordered and divided into the m groups $\theta_{(1)},\ldots,\theta_{(m)}$. When the reference priors that are developed turn out to be proper, matters are straighforward. Often, however, they are improper, and care must be taken in their definition. In the improper case we proceed by specifying (see Section 4.2 for discussion) a nested sequence $\Theta^1 \subset \Theta^2 \subset \ldots$ of compact subsects of Θ such that $\cup_{l=1}^{\infty} \Theta^l = \Theta$. For simplicity, we shall assume that

$$\Theta^l = \Theta_{(1)}^l \times \Theta_{(2)}^l \times \cdots \times \Theta_{(m)}^l.$$

For a discussion on the existence of parametrizations in product form, see Kass (1989); also, the situation of arbirary Θ^l is considered in Berger and Bernardo (1989b).

A reference prior is determined on each compact Θ^l, for which the result is typically a proper prior, followed by performing a limiting operation. Specifically, one follows the following algorithm. Note that expressions for the $h_j(\theta)$ have been given in Section 2.

Start: Define

$$\pi_m^l \left(\theta_{[\sim(m-1)]} | \theta_{[m-1]} \right) = \pi_m^l \left(\theta_{(m)} | \theta_{[m-1]} \right)$$

$$= \frac{|h_m(\theta)|^{1/2} 1_{\Theta_{(m)}^l}}{\int_{\Theta_{(m)}^l} |h_m(\theta)|^{1/2} d\theta_{(m)}}. \tag{3.1.1}$$

Iteration: For $j = m - 1, m - 2, \ldots, 1$, define

$$\pi_j^l \left(\theta_{[\sim(j-1)]} | \theta_{[j-1]}\right)$$

$$= \frac{\pi_{j+1}^l \left(\theta_{[\sim j]} | \theta_{[j]}\right) \exp\left\{\frac{1}{2} E_j^l \left[(\log |\boldsymbol{h}_j(\theta)|) | \theta_{[j]}\right]\right\} 1_{\Theta_{(j)}^l}}{\int_{\Theta_{(j)}^l} \exp\left\{\frac{1}{2} E_j^l \left[(\log |\boldsymbol{h}_j(\theta)|) | \theta_{[j]}\right]\right\} d\theta_{(j)}} \quad (3.1.2)$$

where, letting $\Theta_{[\sim j]}^l = \Theta_{(j+1)}^l \times \cdots \times \Theta_{(m)}^l$,

$$E_j^l \left[g(\theta) | \theta_{[j]}\right] = \int_{\Theta_{[\sim j]}^l} g(\theta) \, \pi_{j+1}^l (\theta_{[\sim j]} | \theta_{[j]}) d\theta_{[\sim j]}. \quad (3.1.3)$$

(Note that it is easy to check, by integrating in turn over $\theta_{(m)}, \theta_{(m-1)}, \ldots, \theta_{(j)}$, that π_j^l defines a probability distribution.) For $j = 1$, interpret $\theta_{[\sim 0]}$ as θ and $\theta_{[0]}$ as vacuous, and write

$$\pi^l(\theta) = \pi_1^l \left(\theta_{[\sim 0]} | \theta_{[0]}\right). \quad (3.1.4)$$

Finish: Define the m-group reference prior, assuming it yields a proper posterior, by

$$\pi(\theta) = \lim_{l \to \infty} \frac{\pi^l(\theta)}{\pi^l(\theta^*)} \quad (3.1.5)$$

where θ^* is some point in Θ^1.

If the integrals and expectation in (3.1.1) and (3.1.2) are finite when the "l" is removed (i.e., when Θ^l is replaced by Θ everywhere), then the reference prior is defined simply by π_1 (i.e., (3.1.5) is not needed). It should also be observed that the condition

$$|\boldsymbol{h}_j(\theta)| \quad \text{depends only on } \theta_{[j]}, \quad (3.1.6)$$

results (for the iteration corresponding to j) in (3.1.2) being replaced by

$$\pi_j^l(\theta_{[\sim(j-1)]} | \theta_{[j-1]}) = \pi_{j+1}^l(\theta_{[\sim j]} | \theta_{[j]}) \frac{|\boldsymbol{h}_j(\theta)|^{1/2} 1_{\Theta_{(j)}^l}}{\int_{\Theta_{(j)}^l} |\boldsymbol{h}_j(\theta)|^{1/2} d\theta_{(j)}}. \quad (3.1.7)$$

3.2. Motivation and Discussion

We suggest that in nonhierarchical models (as considered here) the ordering of the $\theta_{(i)}$ should be in terms of their inferential importance. For instance, in the variance components scenario, to be considered in Section 4, if inference concerning the "between" variance, τ^2, is the primary goal, with the population mean, μ, and "within" variance σ^2 being nuisance parameters, then the suggested ordering would be $\theta_1 = \tau^2, \theta_2 = \mu, \theta_3 = \sigma^2$, or maybe $\theta_1 = \tau^2, \theta_2 = \sigma^2, \theta_3 = \mu$. For inference concerning μ, on the oher hand, (μ, τ^2, σ^2) or (μ, σ^2, τ^2) would be the suggested orderings. On the issue of grouping of coordinates, our advice is: do not group without a very good reason. Thus the k-group reference prior (each stage having $n_i = 1$) is generally recommended. Incidentaly, within groups the ordering of the θ_i is immaterial.

When the reference priors are proper there is no need to consider compact Θ^l. And even when improper, the reference prior is often unaffected by the particular sequence $\{\Theta^l\}$ chosen. When needed, our typical choice of the $\{\Theta^l\}$ is simply a collection of nested rectangles in Θ (or other appropriate shape if Θ is not an "infinite" rectangle).

In Bernardo (1979) the motivation for the reference prior approach is discussed. The idea is basically to choose the prior which, in a certain asymptotic sense, maximizes the information in the posterior that is provided by the data. We will now summarize the argument for the simple case $k = 2$. We stress that the argument below is heuristic and that, rather than trying to make it precise, we choose to *define* the grouped reference prior as that determined by the algorithm in Section 3.1. (See Berger, Bernardo and Mendoza (1989) for an indication of some of the difficulties in making the heuristic argument precise.)

Let x be the result of an experiment which provides information about θ_1, according to the probability model $p(x|\theta_1, \theta_2)$, $(\theta_1, \theta_2) \in \Theta$, where θ_2 is some nuisance parameter, and let Θ^l be a compact approximation to the (joint) parameter space Θ. Let $z_t = \{x_1, \ldots, x_t\}$ be the result from t conditionally independent replications of the original experiment; the amount of information $I_t^{\theta_1}\{p(\theta_1, \theta_2)\}$ to be expected about θ_1 from z_t when the prior is $p(\theta_1, \theta_2)$ is defined to be (Shannon, 1948; Lindley, 1956)

$$I_t^{\theta_1}\{p(\theta_1, \theta_2)\} = \int p(z_t) \int p(\theta_1|z_t) \log \frac{p(\theta_1|z_t)}{p(\theta_1)} d\theta_1 dz_t.$$

Using a variational argument, it may be shown under certain conditions that, for any fixed $p(\theta_1|\theta_2)$, this is maximized by a prior $\pi_t(\theta_1)$ which satisfies

$$\pi_t(\theta_1) \propto \exp\left\{\int p(z_t|\theta_1) \log p(\theta_1|z_t) dz_t\right\}. \tag{3.2.1}$$

Note that (3.2.1) only defines $\pi_t(\theta_1)$ implicitely for given t, since $p(\theta_1 \mid z_t)$ depends on $\pi_t(\theta_1)$. As $t \to \infty$, perfect knowledge is approached, so that $\pi_t(\theta_1)$ approaches that prior which, given $p(\theta_2|\theta_1)$, maximizes the *missing information* about θ_1; this is referred to as the *reference* prior.

Moreover, as $t \to \infty$, and under appropriate regularity conditions, $p(\theta_1|z_t)$ is approached by the asymptotic *marginal* posterior distribution of θ_1, $N\{\theta_1|\hat{\theta}_1, S_1(\hat{\theta}_1, \hat{\theta}_2)\}$, where $S_1(\theta_1, \theta_2)$ is the upper left corner of $S(\theta_1, \theta_2)$, the inverse of Fisher's information matrix $H(\theta_1, \theta_2)$.

It follows that, for each conditonal prior $p(\theta_2|\theta_1)$, and for t large enough, one may write

$$\pi_t(\theta_1) \propto \exp\left\{\int p(z_t|\theta_1) \log N\{\theta_1|\hat{\theta}_1, S_1(\hat{\theta}_1, \hat{\theta}_2)\} dz_t\right\}$$

$$\simeq \exp\left\{\int \int p(\hat{\theta}_1, \hat{\theta}_2|\theta_1) \log N\{\theta_1|\hat{\theta}_1, S_1(\hat{\theta}_1, \hat{\theta}_2)\} d\hat{\theta}_1 d\hat{\theta}_2\right\}$$

$$= \exp\left\{\int \int \left\{\int p(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2) p(\theta_2|\theta_1) d\theta_2\right\} \log N\{\theta_1|\hat{\theta}_1, S_1(\hat{\theta}_1, \hat{\theta}_2)\} d\hat{\theta}_1 d\hat{\theta}_2\right\}$$

$$= \exp\left\{\int \int p(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2) \int p(\theta_2|\theta_1) \log N\{\theta_1|\hat{\theta}_1, S_1(\theta_1, \theta_2)\} d\theta_2 d\hat{\theta}_1 d\hat{\theta}_2\right\}$$

$$\simeq \exp\left\{\int p(\theta_2|\theta_1) \log |S_1(\theta_1, \theta_2)|^{-1/2} d\theta_2\right\}$$

since, as $t \to \infty$, $p(\hat{\theta}_1, \hat{\theta}_2|\theta_1, \theta_2)$ will concentrate on (θ_1, θ_2) and

$$\exp\left\{\log N\{\theta_1|\hat{\theta}_1, S_1(\hat{\theta}_1, \hat{\theta}_2)\}\Big|_{(\hat{\theta}_1, \hat{\theta}_2) = (\theta_1, \theta_2)}\right\} = |S_1(\theta_1, \theta_2)|^{-1/2}.$$

Furthermore, since $h_1 \equiv H_1 \equiv S_1^{-1}$, $|h_1|^{1/2} = |S_1|^{-1/2}$ and hence,

$$\pi_t(\theta_1) \propto \exp\left\{\tfrac{1}{2} \int p(\theta_2|\theta_1) \log |h_1(\theta_1, \theta_2)| d\theta_2\right\}$$

or, normalizing over Θ^l,

$$\pi^l(\theta_1) = \frac{\exp\left\{\frac{1}{2}\int_{\Theta^l_{(2)}} p(\theta_2|\theta_1)\log|\mathbf{h}_1(\theta_1,\theta_2)|d\theta_2\right\}1_{\Theta^l_{(1)}}}{\int_{\Theta^l_{(1)}}\exp\left\{\frac{1}{2}\int_{\Theta^l_{(2)}} p(\theta_2|\theta_1)\log|\mathbf{h}_1(\theta_1,\theta_2)|d\theta_2\right\}d\theta_1}, \qquad (3.2.2)$$

which provides an explicit expression for the reference prior $\pi^l(\theta_1)$ corresponding to a given $p(\theta_2|\theta_1)$.

It is natural to choose $p(\theta_2|\theta_1)$ to be the *conditional* reference prior $\pi(\theta_2|\theta_1)$, i.e. that which, given θ_1, maximizes the missing information about θ_2. By the same argument leading to (3.2.1), this is obtained from

$$\pi_t(\theta_2|\theta_1) \propto \exp\left\{\int p(z_t|\theta_2,\theta_1)\log p(\theta_2|\theta_1,z_t)dz_t\right\}.$$

Again, as $t \to \infty$, and under appropriate regularity conditions, $p(\theta_2|\theta_1,z_t)$ is approximated by the asymptotic *conditional* posterior distribution of θ_2 given θ_1, $N\{\theta_2|\tilde{\theta}_2,\mathbf{h}_2^{-1}(\hat{\theta}_1,\hat{\theta}_2)\}$ where $\mathbf{h}_2(\theta_1,\theta_2)$ is the lower right corner of the information matrix $\mathbf{H}(\theta_1,\theta_2)$. Hence, for large enough t, one may write

$$\pi_t(\theta_2|\theta_1) \propto \exp\left\{\int p(z_t|\theta_1,\theta_2)\log N\{\theta_2|\tilde{\theta}_2,\mathbf{h}_2^{-1}(\hat{\theta}_1,\hat{\theta}_2)\}dz_t\right\}$$

$$= \exp\left\{\int\int p(\hat{\theta}_1,\hat{\theta}_2|\theta_1,\theta_2)\log N\{\theta_2|\tilde{\theta}_2,\mathbf{h}_2^{-1}(\hat{\theta}_1,\hat{\theta}_2)\}d\hat{\theta}_1 d\hat{\theta}_2\right\}$$

$$\simeq |\mathbf{h}_2(\theta_1,\theta_2)|^{1/2}$$

since, as $t \to \infty$, $p(\hat{\theta}_1,\hat{\theta}_2|\theta_1,\theta_2)$ will concentrate on (θ_1,θ_2) and

$$\exp\left\{\log N\{\theta_1|\theta_1,\mathbf{h}_2^{-1}(\theta_1,\theta_2)\}\Big|_{(\hat{\theta}_1,\hat{\theta}_2)=(\theta_1,\theta_2)}\right\} = |\mathbf{h}_2(\theta_1,\theta_2)|^{1/2}.$$

Therefore, after normalizing,

$$\pi^l(\theta_2|\theta_1) = \frac{|\mathbf{h}_2(\theta_1,\theta_2)|^{1/2}1_{\Theta^l_{(2)}}}{\int_{\Theta^l_{(2)}}|\mathbf{h}_2(\theta_1,\theta_2)|^{1/2}d\theta_2}. \qquad (3.2.3)$$

Finally, the joint reference prior necessary to obtain a reference posterior for θ_1 will be defined to be $\pi^l(\theta_1,\theta_2) = \pi^l(\theta_2|\theta_1)\pi^l(\theta_1)$ on Θ^l; hence, if $\{\Theta^l\}$ is a sequence of compacts expanding to Θ, the desired reference prior is defined to be

$$\pi(\theta_1,\theta_2) = \lim_{l\to\infty}\frac{\pi^l(\theta_1,\theta_2)}{\pi^l(\theta_1^*,\theta_2^*)}, \qquad (3.2.4)$$

where (θ_1^*,θ_2^*) is some point of Θ^1; the denominator is necessary to cancel out irrelevant multiplicative constants which may diverge as $l \to \infty$.

Equation 3.2.3 is directly analogous to "start" of the reference prior algorithm in Section 3.1, which gives $\pi^l(\theta_{(m)}|\theta_{(1)},\ldots,\theta_{(m-1)})$; Equation 3.2.2 provides the motivation for the "iteration" steps, and Equation 3.2.4 is a version of the "finish" element in the algorithm.

Further stages (when $m > 2$) are handled in exactly the same manner yielding (3.1.2) as the stage-to-stage updating formula. The net result is $\pi^l(\theta)$, the m-stage reference prior on the compact Θ^l.

Ordered Grouping	Reference Prior
$\{(\mu, \sigma^2, \tau^2)\}$	$\sigma^{-2}(n\tau^2 + \sigma^2)^{-3/2}$
$\{(\mu, \sigma^2), \tau^2\}$	$\sigma^{-5/2}(n\tau^2 + \sigma^2)^{-1}$
$\{(\mu, \tau^2), \sigma^2\}$	$\tau^{-3C_n/2}\sigma^{-2}\psi\left(\dfrac{\tau^2}{\sigma^2}\right)$
$\{\sigma^2, (\mu, \tau^2)\}$	$\sigma^{-1}(n\tau^2 + \sigma^2)^{-3/2}$
$\{\tau^2, (\mu, \sigma^2)\}$	$\tau^{-1}\sigma^{-2}(n\tau^2 + \sigma^2)^{-1/2}\psi\left(\dfrac{\tau^2}{\sigma^2}\right)$
$\{\mu, (\sigma^2, \tau^2)\}$, $\{(\sigma^2 . \tau^2), \mu\}$ $\{\mu, \sigma^2, \tau^2\}$ $\{\sigma^2, \mu, \tau^2\}$, $\{\sigma^2 . \tau^2 . \mu\}$	$\sigma^{-2}(n\tau^2 + \sigma^2)^{-1}$
$\{\mu, \tau^2, \sigma^2\}$, $\{\tau^2, \mu, \sigma^2\}$, $\{\tau^2 . \sigma^2 . \mu\}$	$\tau^{-C_n}\sigma^{-2}\psi\left(\dfrac{\tau^2}{\sigma^2}\right)$

$C_n = \{1 - \sqrt{n-1}(\sqrt{n} + \sqrt{n-1})^{-3}\}$, $\psi(\tau^2/\sigma^2) = [(n-1) + (1 + n\tau^2/\sigma^2)^{-2}]^{1/2}$.
Grouping of parameters is indicated by parentheses.

Table 1. *Reference priors for the variance components problem*

4. THE VARIANCE COMPONENTS PROBLEM

The determination of m-group reference priors for the balanced variance components problem is of interest, not only methodologically, but also because it provides an interesting illustration of the techniques (and possible difficulties) of the general limiting derivation of $\pi(\theta)$ via (3.1.5). Section 4.1 presents the model and the m-group reference priors. Section 4.2 discusses some of the interesting technical issues that arose in the development. Section 4.3 briefly discusses using the reference priors in posterior calculations.

4.1. *The Model and Reference Priors*

We consider the balanced variance components model

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1, \ldots, p \text{ and } j = 1, \ldots, n,$$

where the α_i are i.i.d. $N(\alpha_i | 0, \tau^2)$ and, independently, the ε_{ij} are i.i.d. $N(\varepsilon_{ij} | 0, \sigma^2)$. The parameters (μ, τ^2, σ^2) are unknown.

Since there are only $k = 3$ parameters it is easy to list *all* m-group reference priors. The possible ordered groupings are given in Table 1, along with the associated reference priors. Note that Jeffreys's prior is that associated with the single group $\{(\mu, \sigma^2, \tau^2)\}$; the prior suggested by Box and Tiao (1973, p. 251) is that associated with $\{\mu, (\sigma^2, \tau^2)\}$. Observe that C_n is typically very near 1, and that

$$\sqrt{n-1} \le \psi(\frac{\tau^2}{\sigma^2}) \le \sqrt{n};$$

thus, replacing C_n by 1 and ψ by a constant is reasonable for all but very small n.

As indicated in Section 3.2, we are most favorably disposed towards the last two reference priors in Table 1, since they correspond to the various 3-group reference priors (each group having only a single element). Note that, among the 3-group reference priors, only the order of σ^2 and τ^2 affects the answer; thus there are only two 3-group reference priors instead of the possible six. Thus all that need be specified, in order to determine the 3-group reference prior, is whether σ^2 or τ^2 is deemed to be of more importance.

4.2. *Determination of the Reference Priors*

To implement the algorithm in Section 3.1, compact sets Θ^l must be selected. In deriving the reference priors in Table 1, nested boxes of the form

$$\Theta^l = (a_l, b_l) \times (c_l, d_l) \times (e_l, f_l), \qquad \text{for} \quad (\mu, \sigma^2, \tau^2), \tag{4.2.1}$$

were chosen, where $a_l \to -\infty, c_l$ and $e_l \to 0$, and the upper endpoints $\to \infty$. This would, intuitively, correspond to a presumption of prior independence among the parameters. In most cases, the precise choice of the endpoints in (4.2.1) was immaterial to the result. Disturbing exceptions were the third and last reference priors in Table 1, where the reference prior actually depends on

$$\eta = \lim_{l \to \infty} \left[\frac{\log d_l}{\log c_l^{-1}} \right].$$

If this limit does not exist, there is no reference prior for these situations. If the limit does exist, the third and last reference priors are, in general, as given in Table 1 but with C_n replaced by

$$C_n(\eta) = 1 - \lambda(\eta^2 - \lambda)(\eta - \lambda)^{-2}, \tag{4.2.2}$$

where $\lambda = \sqrt{n-1}/\sqrt{n}$. Recalling that (c_l, d_l) is the range for σ^2, the implication is that we must specify the relative rate at which we are "noninformative" about $\log \sigma^2$, as $\sigma^2 \to 0$ and $\sigma^2 \to \infty$, to determine the reference prior. In Table 1 we made the natural choice $\eta = 1$, but the need to make such an extra choice is clearly unfortunate.

For the third prior in Table 1, it is indeed not even possible to choose a value of η such that $\eta \leq \lambda(\sqrt{6\lambda + 3} - 1)/(3\lambda + 1)$, for then it can be shown that $C_n(\eta) \geq 4/3$ and the prior will have a nonintegrable singularity at $\tau^2 = 0$, a singularity which persists in the posterior; these values of η thus lead to unusable reference priors. Note that $\eta = 1$ does yield a proper posterior.

Alternatives to Θ_l in (4.2.1) can also be considered. One reasonable choice is

$$\Theta^l = \left\{ (\mu, \sigma^2, \tau^2); \ \mu \in (a_l, b_l), \ \sigma^2 \in (c_l, d_l), \ \frac{\tau^2}{\sigma^2} \in (e_l, f_l) \right\}.$$

The point is that it is sometimes natural to be "noninformative" about the ratio τ^2/σ^2 rather than just τ^2 (cf. Hill, 1965).

If such Θ^l are used, the reference priors are as in Table 1, except for the second, third, and last cases, which become $\sigma^{-3}(n\tau^2 + \sigma^2)^{-1}$, $\tau^{-3}\sigma^{-2}\psi(\tau^2/\sigma^2)$, and $\tau^{-2}\sigma^{-2}\psi(\tau^2/\sigma^2)$ respectively. These last two priors have nonintegrable singularities at $\tau^2 = 0$, which persist in the posterior, and hence are not usable.

We shall present here the development of the last reference prior in Table 1, for the ordered grouping $\{\mu, \tau^2, \sigma^2\}$. The analyses for all other cases in Table 1 are similar, but simpler, and are hence omitted.

We apply the algorithm in Section 3.1, with the Θ^l defined by (4.2.1). Note that the Fisher information matrix for (μ, τ^2, σ^2) is

$$
H(\mu, \tau^2, \sigma^2) = \begin{pmatrix} \dfrac{pn}{(n\tau^2 + \sigma^2)} & 0 & 0 \\[2ex] 0 & \dfrac{pn^2}{2(n\tau^2 + \sigma^2)^2} & \dfrac{pn}{2(n\tau^2 + \sigma^2)^2} \\[2ex] 0 & \dfrac{pn}{2(n\tau^2 + \sigma^2)^2} & \left[\dfrac{p(n-1)}{2\sigma^4} + \dfrac{p}{2(n\tau^2 + \sigma^2)^2}\right] \end{pmatrix},
$$

so that

$$
S(\mu, \tau^2, \sigma^2) = \begin{pmatrix} \dfrac{(n\tau^2 + \sigma^2)}{pn} & 0 & 0 \\[2ex] 0 & \left[\dfrac{2\sigma^4}{pn^2(n-1)} + \dfrac{2(n\tau^2 + \sigma^2)^2}{pn^2}\right] & -\dfrac{2\sigma^4}{pn(n-1)} \\[2ex] 0 & -\dfrac{2\sigma^4}{pn(n-1)} & \dfrac{2\sigma^4}{p(n-1)} \end{pmatrix}
$$

In the notation of Section 2, the 3-group case $\{\mu, \tau^2, \sigma^2\}$ corresponds to $\theta_{(1)} = \theta_1 = \mu$, $\theta_{(2)} = \theta_2 = \tau^2$, and $\theta_{(3)} = \theta_3 = \sigma^2$. Also, S satisfies (2.3) with $B_2 = 0$, $c_2 = 1$, and $A_{32} = -2\sigma^4/[pn(n-1)]$, so that

$$
h_1 = \sigma_{11}^{-1} = \frac{pn}{n\tau^2 + \sigma^2}
$$

and (2.4) yields

$$
\begin{aligned}
h_2 &= \left[A_{22} + A_{11} - h_1 A_{11}^2\right]^{-1} \\[1ex]
&= \left[\frac{2\sigma^4}{pn^2(n-1)} + \frac{2(n\tau^2 + \sigma^2)^2}{pn^2}\right]^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
h_3 &= \left[A_{33} + A_{22} - 2A_{32} - h_2(A_{22} - A_{32})^2\right]^{-1} \\[1ex]
&= \left[\frac{p(n-1)}{2\sigma^4} + \frac{p}{2(n\tau^2 + \sigma^2)^2}\right].
\end{aligned}
$$

Start: To begin,

$$
\begin{aligned}
\pi_3^l(\sigma^2 \mid \mu, \tau^2) &= \frac{|h_3(\mu, \tau^2, \sigma^2)|^{1/2} 1_{(c_l, d_l)}(\sigma^2)}{\int_{c_l}^{d_l} |h_3(\mu, \tau^2, \sigma^2)|^{1/2} d\sigma^2} \\[2ex]
&= \frac{\left[\dfrac{p(n-1)}{2\sigma^4} + \dfrac{p}{2(n\tau^2 + \sigma^2)^2}\right]^{1/2} 1_{(c_l, d_l)}(\sigma^2)}{\Psi_1(\tau^2 \mid c_l, d_l)},
\end{aligned}
$$

where

$$
\Psi_1(\tau^2 \mid c_l, d_l) = \int_{c_l}^{d_l} \left[\frac{p(n-1)}{2\sigma^4} + \frac{p}{2(n\tau^2 + \sigma^2)^2}\right]^{1/2} d\sigma^2.
$$

Iteration for $j = 2$.

$$
E^l\left[\log|h_2| \mid (\mu, \tau^2)\right] = \frac{\Psi_2(\tau^2 \mid c_l, d_l)}{\Psi_1(\tau^2 \mid c_l, d_l)},
$$

where

$$\Psi_2(\tau^2|c_l, d_l) = \int_{c_l}^{d_l} \left(\log \left[\frac{2\sigma^4}{pn^2(n-1)} + \frac{2(n\tau^2 + \sigma^2)^2}{pn^2} \right]^{-1} \right)$$

$$\times \left[\frac{p(n-1)}{2\sigma^4} + \frac{p}{2(n\tau^2 + \sigma^2)^2} \right]^{1/2} d\sigma^2.$$

Hence

$$\pi_2^l(\tau^2, \sigma^2|\mu) = \frac{\left[\frac{p(n-1)}{2\sigma^4} + \frac{p}{2(n\tau^2 + \sigma^2)^2} \right]^{1/2} 1_{(c_l, d_l)}(\sigma^2)}{\Psi_1(\tau^2|c_l, d_l)}$$

$$\times \frac{\exp\left\{ \frac{1}{2}\Psi_2(\tau^2|c_l, d_l)/\Psi_1(\tau^2|c_l, d_l) \right\} 1_{(c_l, f_l)}(\tau^2)}{K(c_l, d_l, e_l, f_l)},$$

where

$$K(c_l, d_l, e_l, f_l) = \int_{e_l}^{f_l} \exp\left\{ \frac{1}{2}\Psi_2/\Psi_1 \right\} d\tau^2.$$

Iteration for $j = 1$. Since neither h_1, nor π_2^l depend on μ,

$$E^l \left[\log |h_1(\mu, \tau^2, \sigma^2)| \mid \mu \right] = k^*(c_l, d_l, e_l, f_l).$$

Hence

$$\pi^l(\mu, \tau^2, \sigma^2) = \pi_1^l(\mu, \tau^2, \sigma^2) = \frac{\pi_2^l(\tau^2, \sigma^2 \mid \mu)}{(b_l - a_l)} 1_{(a_l, b_l)}(\mu).$$

Finish. Choosing, say, the fixed point $(\mu, \tau^2, \sigma^2) = (0, 1, 1)$, one obtains (ignoring multiplicative constants)

$$\pi(\mu, \tau^2, \sigma^2) = \lim_{l \to \infty} \left\{ \frac{\left[\frac{p(n-1)}{2\sigma^4} + \frac{p}{2(n\tau^2 + \sigma^2)^2} \right]^{1/2}}{\Psi_1(\tau^2|c_l, d_l)/\Psi_1(1|c_l, d_l)} \right.$$

$$\left. \times \exp\left\{ \frac{\Psi_2(\tau^2|c_l, d_l)}{2\Psi_1(\tau^2|c_l, d_l)} - \frac{\Psi_2(1|c_l, d_l)}{2\Psi_1(1|c_l, d_l)} \right\} \right\}. \tag{4.2.3}$$

To determine this limit, note first that changing variables from σ^2 to $v = \tau^2/\sigma^2$ in the integrals defining Ψ_1 and Ψ_2 yields

$$\Psi_1(\tau^2|c_l, d_l) = \sqrt{\frac{p}{2}} \int_{\tau^2/d_l}^{\tau^2/c_l} \frac{1}{v} \left[n - 1 + \frac{1}{(1+nv)^2} \right]^{1/2} dv,$$

$$\Psi_2(\tau^2|c_l, d_l) = -\sqrt{\frac{p}{2}} \int_{\tau^2/d_l}^{\tau^2/c_l} \left[\log \tau^4 + \log\left(\frac{2}{pn^2} \right) + \log\left(\frac{1}{(n-1)v^2} + (n + v^{-1})^2 \right) \right]$$

$$\times \frac{1}{v} \left[n - 1 - \frac{1}{(1+nv)^2} \right]^{1/2} dv.$$

Divide the integrals up into integrals over the regions $(\tau^2/d_l, \varepsilon)$, $(\varepsilon, \varepsilon^{-1})$, and $(\varepsilon^{-1}, \tau^2/c_l)$. Clearly

$$
\int_{\tau^2/d_l}^{\varepsilon} \frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv = \int_{\tau^2/d_l}^{\epsilon}\left[\frac{\sqrt{n}}{v} + O(1)\right] dv
$$

$$
= \sqrt{n}\left[\left(\log\frac{\varepsilon}{\tau^2}\right) + \log d_l\right] + O(\epsilon).
$$

Likewise,

$$
\int_{\tau^2/d_l}^{\varepsilon} \log\left(\frac{1}{(n-1)v^2} + (n+v^{-1})^2\right) \frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv
$$

$$
= \int_{\tau^2/d_l}^{\varepsilon}\left[\log\left(\frac{n}{(n-1)v^2}\right) + O(v)\right]\left[\frac{\sqrt{n}}{v} + O(1)\right] dv
$$

$$
= \left(\log\frac{n}{n-1}\right)\sqrt{n}\left(\log\frac{\varepsilon}{\tau^2} + \log d_l\right) + \sqrt{n}\left[(\log\tau^2 - \log d_l)^2 - (\log\varepsilon)^2\right] + O(\varepsilon).
$$

For the range $(\varepsilon^{-1}, \tau^2/c_l)$ we get

$$
\int_{\varepsilon^{-1}}^{\tau^2/c_l} \frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv = \int_{\varepsilon^{-1}}^{\tau^2/d_l}\left(\frac{\sqrt{n-1}}{v} + O\left(\frac{1}{v^3}\right)\right) dv
$$

$$
= \sqrt{n-1}\left(\log\varepsilon\tau^2 - \log c_l\right) + O(\varepsilon^2)
$$

and

$$
\int_{\varepsilon^{-1}}^{\tau^2/c_l} \log\left(\frac{1}{(n-1)v^2} + (n+v^{-1})^2\right) \frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv
$$

$$
= \int_{\varepsilon^{-1}}^{\tau^2/c_l} \log\left(n^2 + 0\left(\frac{1}{v}\right)\right)\left[\frac{\sqrt{n-1}}{v} + O\left(\frac{1}{v^3}\right)\right] dv
$$

$$
= \sqrt{n-1}(\log n^2)(\log\varepsilon\tau^2 - \log c_l).
$$

To complete the specification, define

$$
K_1(\varepsilon) = \int_{\varepsilon}^{\varepsilon^{-1}} \frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv,
$$

$$
K_2(\varepsilon) = \int_{\varepsilon}^{\varepsilon^{-1}} \log\left(\frac{1}{(n-1)v^2} + (n+v^{-1})^2\right)\frac{1}{v}\left[n - 1 + \frac{1}{(1+nv)^2}\right]^{1/2} dv.
$$

Collecting terms yields

$$\Psi_1(\tau^2|c_l, d_l) = \sqrt{\frac{p}{2}}\Big\{ \sqrt{n}\log d_l - \sqrt{n-1}\log c_l + (\log \varepsilon)\left(\sqrt{n} + \sqrt{n-1}\right)$$

$$+ (\log \tau^2)(\sqrt{n-1} - \sqrt{n}) + K_1(\varepsilon) + O(\varepsilon)\Big\}$$

$$\Psi_2(\tau^2|c_l, d_l) = -\left[\log \tau^4 + \log \frac{2}{pn^2}\right]\Psi_1(\tau^2|c_l, d_l)$$

$$- \sqrt{\frac{p}{2}}\Big\{ \sqrt{n}(\log d_l)^2 - (\log d_l)2\sqrt{n}(\log \tau^2) - 2\sqrt{n-1}\log n \log c_l$$

$$+ \sqrt{n}\log \frac{n}{(n-1)} \log d_l + K^*(\varepsilon, \tau^2) + O(\varepsilon)\Big\},$$

where K^* is a function of ε and τ^2.

As $l \to \infty$, $\log d_l$ and $\log c_l^{-1}$ converge to infinity, while all other terms in Ψ_1 and Ψ_2 remain constant. (The $O(\varepsilon)$ terms are uniform in l.) Hence, it follows immediately that

$$\lim_{l \to \infty} \frac{\Psi_1(\tau^2|c_l, d_l)}{\Psi_1(1|c_l, d_l)} = 1. \tag{4.2.4}$$

More difficult is the limit of Ψ_2/Ψ_1. Defining

$$R(\varepsilon, \tau^2) = (\log \varepsilon)\left(\sqrt{n} + \sqrt{n-1}\right) + (\log \tau^2)\left(\sqrt{n-1} - \sqrt{n}\right) + K_1(\varepsilon),$$

$$S(\varepsilon, \tau^2, c_l, d_l) = \sqrt{n}(\log d_l)^2 - (\log d_l)2\sqrt{n}(\log \tau^2) + \sqrt{n}\log\left(\frac{n}{n-1}\right)\log d_l$$

$$- 2\sqrt{n-1}(\log n)\log c_l + K^*(\varepsilon, \tau^2),$$

we have

$$\frac{\Psi_2(\tau^2|c_l, d_l)}{\Psi_1(\tau^2|c_l, d_l)} + \left[\log \tau^4 + \log \frac{2}{pn^2}\right]$$

$$= \frac{-\{S(\varepsilon, \tau^2, c_l, d_l) + O(\varepsilon)\}}{\{\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l + R(\varepsilon, \tau^2) + O(\varepsilon)\}}$$

$$= \frac{-\{S(\varepsilon, \tau^2, c_l, d_l) + O(\varepsilon)\}}{\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l}\left[1 - \frac{R(\varepsilon, \tau^2) + O(\varepsilon)}{\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l} + O\left(\frac{1}{(\log d_l/c_l)^2}\right)\right]$$

$$= \frac{-\left[\sqrt{n}(\log d_l)^2 + \sqrt{n}\left(\log \frac{n}{n-1}\right)\log d_l\right]}{\left(\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l\right)} + \frac{[R(\varepsilon, \tau^2) + O(\varepsilon)]\sqrt{n}(\log d_l)^2}{\left(\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l\right)^2}$$

$$+ \frac{(2\sqrt{n}(\log \tau^2)(\log d_l) + 2\sqrt{n-1}(\log n)(\log c_l))}{\left(\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l\right)} + O\left(\frac{1}{(\log d_l/c_l)}\right).$$

Thus

$$\frac{\Psi_2(\tau^2|c_l, d_l)}{\Psi_1(\tau^2|c_l, d_l)} - \frac{\Psi_2(1|c_l, d_l)}{\Psi_1(1|c_l, d_l)}$$

$$= -\log \tau^4 + \frac{\left[R(\varepsilon, \tau^2) - R(\varepsilon, 1) + O(\varepsilon)\right]\sqrt{n}(\log d_l)^2}{\left(\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l\right)^2}$$

$$+ \frac{2\sqrt{n}(\log \tau^2)\log d_l}{\left(\sqrt{n}(\log d_l) - \sqrt{n-1}\log c_l\right)} + O\left(\frac{1}{(\log d_l/c_l)}\right)$$

$$= -\log \tau^4 - \frac{\left[(\log \tau^2)(1 - \sqrt{n-1}/\sqrt{n}) + O(\varepsilon)\right]}{\left[1 - (\sqrt{n-1}/\sqrt{n})(\log c_l/\log d_l)\right]^2}$$

$$+ \frac{2(\log \tau^2)}{\left[1 - (\sqrt{n-1}/\sqrt{n})(\log c_l/\log d_l)\right]} + O\left(\frac{1}{(\log d_l/c_l)}\right).$$

Noting that the $O(\varepsilon)$ term was uniform in l, and assuming that

$$\eta = \lim_{l\to\infty} \frac{\log d_l}{\log c_l^{-1}}$$

exists, we thus have, defining $\lambda = \sqrt{n-1}/\sqrt{n}$,

$$\lim_{l\to\infty} \left\{ \frac{\Psi_2(\tau^2|c_l, d_l)}{\Psi_1(\tau^2|c_l, d_l)} - \frac{\Psi_2(1|c_l, d_l)}{\Psi_1(1|c_l, d_l)} \right\}$$

$$= -\log \tau^4 + \log \tau^2 \left[\frac{(\lambda - 1)}{(1 + \eta^{-1}\lambda)^2} + \frac{2}{(1 + \eta^{-1}\lambda)} \right] + \frac{O(\epsilon)}{(1 + \eta^{-1}\lambda)^2}$$

$$= -(\log \tau^2)\left[1 - \frac{\lambda(\eta^2 - \lambda)}{(\eta + \lambda)^2}\right] + \frac{O(\epsilon)}{(1 + \eta^{-1}\lambda)^2}.$$

Observing that $O(\varepsilon)$ can be made arbitrarily small, this equation, together with (4.2.3) and (4.2.4), yields the reference prior in Table 1, with C_n replaced by (4.2.2).

4.3. *Posterior Calculations*

Note that the likelihood function is proportional to

$$l(\mu, \sigma^2, \tau^2) = \sigma^{-(n-1)p}(n\tau^2 + \sigma^2)^{-p/2} \exp\left\{ -\frac{1}{2}\left[\frac{np(\mu - \overline{\overline{x}})^2}{n\tau^2 + \sigma^2} \right.\right.$$

$$\left.\left. + \frac{n\Sigma(\overline{x}_i - \overline{\overline{x}})^2}{n\tau^2 + \sigma^2} + \frac{\Sigma\Sigma(x_{ij} - \overline{x}_i)^2}{\sigma^2} \right]\right\}.$$

Also, all of the reference priors can be written in the form

$$\pi(\mu, \sigma^2, \tau^2) = \sigma^{-\alpha}\tau^{-\beta}(n\tau^2 + \sigma^2)^{-\gamma}\psi^*\left(\frac{\tau^2}{\sigma^2}\right)$$

for certain constants $\alpha, \beta, \gamma,$ and ψ^* either equal to ψ in Table 1, or equal to 1.

Finally, suppose one is interested in evaluating the posterior expectation of a function of the form

$$\rho(\mu, \sigma^2, \tau^2) = \mu^r \sigma^s \varphi\left(\frac{\tau^2}{\sigma^2}\right),$$

a form which clearly includes all posterior moments (and cross-moments) of $\mu, \sigma,$ and τ. The following lemma shows that this calculation reduces to one-dimensional integration for $r = 0, 1,$ or 2. (Integer $r \geq 3$ also can be handled using only one-dimensional integration, but the formula gets messier.)

Lemma 4.1. *The posterior expectation of* $\rho(\mu, \sigma^2, \tau^2)$ *for* $r = 0, 1,$ *or* $2,$ *is*

$$
E\left[\rho(\mu, \sigma^2, \tau^2) \Big| \ data\right] = \frac{\Gamma\left(\frac{1}{2}(c - s)\right)}{2^{s/2}\Gamma(c/2)}
$$

$$
\times \frac{\int_0^\infty [S(v)]^{s-c} \left\{ \overline{x}^r + \frac{(1 + nv)S^2(v)}{np(c - 2 - s)} 1_2(r) \right\} v^{\frac{-\beta}{2}}(1 + nv)^{-\gamma^*} \phi(v)\psi^*(v)\, dv}{\int_0^\infty [S(v)]^{-c} v^{\frac{-\beta}{2}}(1 + nv)^{-\gamma^*} \psi^*(v)\, dv},
$$

where $\gamma^* = \gamma + \frac{1}{2}(p - 1),$ $c = \alpha + \beta + 2\gamma + np - 5,$ $1_2(r)$ *equals* 1 *if* $r = 2$ *and equals* 0
otherwise, and

$$
S(v) = \left(\frac{n \sum_i (\overline{x}_i - \overline{\overline{x}})^2}{(1 + nv)} + \sum_i \sum_j (x_{ij} - \overline{x}_i)^2 \right)^{1/2}.
$$

Proof. First transform to the variables (μ, σ^2, v), where $v = \tau^2/\sigma^2$. It is then straightforward
to integrate over μ, followed by σ^2, yielding the result. ◁

5. CONCLUDING REMARKS

The variance components example clearly illustrates the difficulties that can be encountered
in applying the grouped reference prior algorithm:

(i) the grouping and ordering of the groups will frequently result in different reference priors,

(ii) the limiting process given in the algorithm can be difficult to carry out, and

(iii) the limit can depend on the compact sets chosen (and may not even yield a proper
posterior.)

Before addressing these points, it should be noted that it is rather rare to encounter the second
and third difficulties; indeed, the variance components problem is the most pathological we
have seen for the reference prior theory.

The dependence of the reference prior on the group chosen and their order is, we feel,
unavoidable. Many examples exist which illustrate that no single noninformative prior will
work well for all functions of a given high-dimensional parameter. As more fully discussed
in Berger and Bernardo (1989b), our own preference is actually to use the reference prior
corresponding to single element groups, with the groups ordered according to the inferential
importance of the parameters. That different orderings of the nuisance parameters can yield
different answers even has positive aspects; one can then conduct a sensitivity study over the
choice of the noninformative prior.

As to the possible technical difficulty of implementing the grouped reference prior al-
gorithm, again most examples we have seen are much easier than the variance components
example. Also, the determination of the reference prior can be thought of as the theoretician's
work, to be done for all common models of statistical importance.

Difficulties with existence or uniqueness of the limit in the algorithm are more troubling.
It would be nice if such never occurred, and indeed occurrence of these problems is quite
rare, but there are no guarantees. As in Berger and Bernardo (1989a), one can surmount the
non-uniqueness problem (when present) if there is a natural sequence of compact sets that one
prefers.

At the very least, the grouped reference prior algorithm can be thought of as a method
for generating interesting candidate noninformative priors, either for sensitivity studies or for

investigation of their performance. While our attitude is that study of the performance of noninformative priors is certainly to be encouraged, we have found the group reference priors to generally be highly satisfactory, and we would feel reasonably confident in using them in situations in which further study is impossible.

ACKNOWLEDGEMENTS

REFERENCES

Berger, J. O. and Bernardo, J. M. (1989a). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200–207.

Berger, J. O. and Bernardo, J. M. (1989b). Ordered group reference priors with applications to multinomial and variance components problems. *Tech. Rep.* **01/89**, *Dept. Estadística.* Presidencia de la Generalidad Valenciana, Spain.

Berger, J. O., Bernardo, J. M., and Mendoza, M. (1989). On priors that maximize expected information. *Recent Developments in Statistics and their Applications* (J. Klein and J.C. Lee, eds.). Seoul: Freedom Academy. 1–20.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass.: Addison-Wesley.

Hill, B. M. (1965). Inference about variance components in the one-way model. *J. Amer. Statist. Assoc.* **60**, 806–825.

Jeffreys, H. (1961). *Theory of Probability.* Oxford: University Press.

Kass, R. E. (1989). The geometry of asymptotic inference. *Statist. Sci.* **4**, (in press).

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist.* **27**, 986–1005.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423 and 623–656.