ORDERED GROUP REFERENCE PRIORS WITH APPLICATIONS TO MULTINOMIAL AND VARIANCE COMPONENTS PROBLEMS

by

James O. Berger
Department of Statistics
Purdue University
West Lafayette, IN, USA

and

José M. Bernardo
Generalidad Valenciana
SPAIN

Technical Report # 89–31C

Department of Statistics
Purdue University

December 1989

△

Tech. Rep. **01/89**, (November 2, 1989). (*)
Departamento de Estadística, Presidencia de la Generalidad.
Caballeros 2, E-46001 - Valencia, Spain.
Tel. (346) 386 36 65, Fax (346) 386 61 37.

Ordered Group Reference Priors with Applications to Multinomial and Variance Components Problems

JAMES O. BERGER and JOSÉ M. BERNARDO

Purdue and Duke Universities, USA, and *Generalidad Valenciana, Spain*

SUMMARY

Noninformative priors are developed, using the reference prior approach, for multiparameter problems in which there may be parameters of interest and nuisance parameters. For a given grouping of parameters and ordering of the groups (intuitively, according to inferential importance), an algorithm for determining the associated reference prior is presented. The algorithm is illustrated on a multinomial problem and on a variance components problem, with discussion of the variety and success of various groupings and ordering strategies.

Keywords: BAYESIAN INFERENCE; MULTIPARAMETER PROBLEMS; NONINFORMATIVE PRIORS.

AMS 1980 subject classifications. Primary 62F15; Secondary 62A15.

1. INTRODUCTION

In development of noninformative prior distributions, Bernardo (1979) explicitely recognized the importance of identifying the *parameters of interest* and the *nuisance parameters*, and tailoring the noninformative prior to this choice; a global noninformative prior distribution (e.g., that of Jeffreys, 1961) will not always be adequate for inferences about different parameters within a model. Many of the "counterexamples" to noninformative priors (e.g., those of Stein, 1959, or Dawid, Stone and Zidek, 1973) provide dramatic illustrations of this fact.

The *reference prior* approach of Bernardo (1979) addresses this problem by suggesting a two step reference prior. First, find the conditional reference prior for the nuisance parameters given the parameters of interest; then find the reference prior for the parameters of interest in the marginal model formed by integrating out the nuisance parameters. This procedure worked well in the examples considered in Bernardo (1979) and in subsequent works such as Bayarri (1981, 1985), Bernardo (1980, 1981, 1982, 1985), Bernardo and

(*) This work was supported by the U.S.-Spain. Joint Committee for Scientific and Technological Cooperation, Grant CC B8409-025, and by the National Science Foundation, Grant DMS 8702620. The second author is on leave of absence from the *Department of Statiscitcs, University of Valencia*, Spain.

Girón (1988) Eaves (1983, 1985), Ferrandiz (1982), Lindley (1988), Mendoza (1987, 1988) or Sendra (1982).

Recently, two limitations of the method have been observed. The first is somewhat technical, but often crucial. The conditional reference prior found in the first step in often improper, and yet is subsequently used to form the marginal model for the parameter of interest. Attempts to rigorously justify this step revealed a rather surprising necesssity: one must "normalize" even improper conditional reference priors. The normalization, and indeed the entire calculation, is done by a limiting operation on proper versions of the problem; Berger and Bernardo (1989a) illustrated this approach in the problem of estimating a product of normal means.

The second recent observation is that merely grouping the parameters of a model into "parameters of interest" and "nuisance parameters" may not go far enough. Allowing multiple groups "ordered" in terms of importance may be needed, with the reference prior being determined through a succession of analyses for the implied conditional problems. In fact, experience leads us to recommend providing a complete ordering of all parameters of a model, so that the reference prior is determined through a series of one-dimensional conditional steps.

In Section 2 of this paper, we introduce the general m-group reference prior algorithm. Sections 3 and 4 provide illustrative applications to, respectively, a multinomial and a variance components problem. Section 5 presents conclusions and discussion.

As background for the developments in the paper, it is perhaps helpful to briefy mention our overall philosophy concerning noninformative priors. This begins with the observation that noninformative priors seem to be enormously popular in applied Bayesian works; even the most avowed subjectivists seem to heavily use noninformative priors —perhaps with profuse apologies— when actually analyzing data. The second cornerstone of our philosophy is that no one has succeeded (or is ever likely to succeed) in defining unambiguously "noninformative" priors in an absolute sense. Our goal is the more modest goal of developing an algorithm for generation of priors that have a minimal impact on the Bayesian analysis when compared with the impact provided by the data. The concern is that, in higher dimensions, noninformative priors (such as Jeffreys' prior) can have hidden features that have a dramatic (and unrecognized) effect on the answer.

The reference prior approach, especially the new approach discussed herein of development through a series of one-dimensional conditional steps, seems to be remarkably successful in obtaining noninfluential priors in higher dimensions. We would certainly not claim that this reference prior approach is guaranteed to produce a prior with no undesirable characteristics, but its successes (and the lack of practically important counterexamples) are impresive.

2. NOTATION AND THE ALGORITHM

Section 2.1 presents needed notation. Section 2.2 develops the general m-group reference prior algorithm. Section 2.3 discusses the motivation for the algorithm.

2.1. *Notation*

We consider a 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

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

exists and has rank k, so that

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

also exists. Often, we will just write $\boldsymbol{H}$ and $\boldsymbol{S}$.

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 2.3— to achieve the desired ordered grouping.) Also we shall define, for $j = 1, \ldots, m$,

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

Finally, if we write $\boldsymbol{S}$ as

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

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

$\boldsymbol{S}_j \equiv$ upper left $(N_j \times N_j)$ corner of $\boldsymbol{S}$, with $\boldsymbol{S}_m \equiv \boldsymbol{S}$, and
$\boldsymbol{H}_j \equiv \boldsymbol{S}_j^{-1}$

then, the matrices

$\boldsymbol{h}_j \equiv$ lower right $(n_j \times n_j)$ corner of $\boldsymbol{H}_j$, $j = 1, \ldots, m$

will be of central importance. Expressions for these matrices are given in Appendix 1. In particular, $\boldsymbol{h}_1 \equiv \boldsymbol{H}_1 \equiv \boldsymbol{A}_{11}^{-1}$ and, if $\boldsymbol{S}$ is a block diagonal matrix, (i.e., $\boldsymbol{A}_{ij} \equiv \boldsymbol{0}$ for all $i \neq j$) then $\boldsymbol{h}_j \equiv \boldsymbol{A}_{jj}^{-1}$, $j = 1, \ldots, m$.

Finally, if $\Theta^* \subset \Theta$, we will define

$$\Theta^*(\theta_{[j]}) = \left\{\theta_{(j+1)} : (\theta_{[j]}, \theta_{(j+1)}, \theta_{[\sim(j+1)]}) \in \Theta^* \quad \text{for some} \quad \theta_{[\sim(j+1)]}\right\}; \tag{2.1.1}$$

we will use the common symbols

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

and will throughout the paper adopt the conventions that $\sum_{i=l}^{l-1}(\cdot) = 0$ and $\prod_{i=l}^{l-1}(\cdot) = 1$.

2.2. *The m-Group Reference Prior*

We suppose the θ_i have been ordered and divided into the m groups $\theta_{(1)}, \ldots, \theta_{(m)}$. (The ordering within the groups does not matter; see section 2.3 for discussion of the grouping and ordering.)

When the reference priors that are developed turn out to be proper (see, e.g., Section 3), 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 2.3 for discussion) a nested sequence $\Theta^1 \subset \Theta^2 \subset \ldots$ of compact subsects of Θ such that $\cup_{l=1}^{\infty} \Theta^l = \Theta$. 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 $\boldsymbol{h}_j(\theta)$ are given in Appendix 1.

Start: Define

$$\begin{aligned} \pi_m^l\left(\theta_{[\sim(m-1)]}|\theta_{[m-1]}\right) &= \pi_m^l\left(\theta_{(m)}|\theta_{[m-1]}\right) \\ &= \frac{|\boldsymbol{h}_m(\theta)|^{1/2} 1_{\Theta^l(\theta_{[m-1]})}(\theta_{(m)})}{\displaystyle\int_{\Theta^l(\theta_{[m-1]})} |\boldsymbol{h}_m(\theta)|^{1/2} d\theta_{(m)}}. \end{aligned} \tag{2.2.1}$$

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

$$\begin{aligned} &\pi_j^l\left(\theta_{[\sim(j-1)]}|\theta_{[j-1]}\right) \\ &\quad = \frac{\pi_{j+1}^l\left(\theta_{[\sim j]}|\theta_{[j]}\right) \exp\left\{\frac{1}{2} E_j^l\left[\left(\log|\boldsymbol{h}_j(\theta)|\right)|\theta_{[j]}\right]\right\} 1_{\Theta^l(\theta_{[j-1]})}\left(\theta_{(j)}\right)}{\displaystyle\int_{\Theta^l(\theta_{[j-1]})} \exp\left\{\frac{1}{2} E_j^l\left[\left(\log|\boldsymbol{h}_j(\theta)|\right)|\theta_{[j]}\right]\right\} d\theta_{(j)}} \end{aligned} \tag{2.2.2}$$

where

$$E_j^l\left[g(\theta)|\theta_{[j]}\right] = \int_{\left\{\theta_{[\sim j]}:(\theta_{[j]},\theta_{[\sim j]},)\in\Theta^l\right\}} g(\theta)\, \pi_{j+1}^l(\theta_{[\sim j]}|\theta_{[j]}) d\theta_{[\sim j]}. \tag{2.2.3}$$

(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

$$\pi^l(\theta) = \pi_1^l\left(\theta_{[\sim 0]}|\theta_{[0]}\right). \tag{2.2.4}$$

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

$$\pi(\theta) = \lim_{l\to\infty} \frac{\pi^l(\theta)}{\pi^l(\theta^*)} \tag{2.2.5}$$

where θ^* is some point in Θ^1.

Note: If the integrals and expectation in (2.2.1) and (2.2.2) 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., (2.2.5) is not needed).

The calculation of the m-group reference prior is greatly simplified under the condition

$$|\boldsymbol{h}_j(\theta)| \quad \text{depends only on } \theta_{[j]}, \quad \text{for} \quad j = 1, \ldots, m. \tag{2.2.6}$$

Lemma 2.2.1. *If* (2.2.6) *holds, then*

$$\pi^l(\theta) = \left(\prod_{i=1}^{m} \frac{|\boldsymbol{h}_i(\theta)|^{1/2}}{\int_{\Theta^l(\theta_{[i-1]})} |\boldsymbol{h}_i(\theta)|^{1/2} d\theta_{(i)}}\right) 1_{\Theta^l}(\theta). \tag{2.2.7}$$

Proof. Using (2.2.6) it is clear that

$$E_j^l\left[\log|\boldsymbol{h}_j(\theta)||\theta_{[j]}\right] = \log|\boldsymbol{h}_j(\theta)|.$$

The result is immediate from (2.2.2). ◁

2.3. *Motivation and Explanation*

2.3.1. Ordering and Grouping

What ordering should be chosen for the θ_i?. In nonhierarchical models (as considered here) we suggest the ordering be in terms of the inferential importance of the θ_i. For instance, in the variance components problem of Section 4, suppose that 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. (As argued in Bernardo, 1979, a cornerstone of

the reference prior approach in that the reference prior may change as one focuses on different parameters, even within the same study.)

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. At one time (e.g., Berger and Bernardo, 1989a), we advocated creating two groups, with $\theta_{(1)}$ being the "parameters of interest" and $\theta_{(2)}$ being the "nuisance parameters." Examples of unsuitable performance (to be discussed elsewhere) led us to consider additional groups, eventually leading to the present recommendation. (An example in which one might choose to group, is discussed in Section 3.4). Incidentally, within groups the ordering of the θ_i is immaterial.

2.3.2. Choice of the Θ^l

To reiterate, 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. (An exception will be seen in Section 4 — see also Berger and Bernardo, 1989a.)

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). This is based on the heuristic idea that the Θ^l should reflect the type of set on which we would state "noninformativeness" if we had to choose a compact set (though by choosing a nested infinite sequence we do not commit ourselves to any particular compact) and it is often the case that parametrizations are chosen so that one is "noninformative" about natural regions (e.g., rectangles) in that parametrization. This is admittedly quite vague and, to be honest, we are unhappy when the choice of $\{\Theta^l\}$ matters. Note that consideration of limits of compacts is also necessary in certain other approaches to development of noninformative priors; Cifarelli and Regazzini (1987) and Consonni and Veronese (1988) are two recent such references.

2.3.3. The Motivation for the k-Group Reference Prior Algorithm

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 not repeat the discussion here.

In Berger and Bernardo (1989a), a treatment of the case $k = 2$ is presented. The idea is to first find the reference prior for θ_2, at each given value of θ_1, calling this the "conditional reference prior" $\pi^l(\theta_2|\theta_1)$. Assuming asymptotic normality for the model, the argument in Bernardo (1979) leads (see also Berger and Bernardo, 1989a) to

$$\pi^l(\theta_2|\theta_1) \propto |\boldsymbol{h}_2(\theta_1, \theta_2)|^{1/2} 1_{\Theta^l(\theta_1)}(\theta_2). \tag{2.3.1}$$

Since this would subsequently be combined with a "marginal" reference prior for θ_1, it was realized that normalization would be important. Hence the Θ^l were introduced (in case

$|\boldsymbol{h}_2(\theta_1,\theta_2)|^{1/2}$ was not integrable), and $\pi^l(\theta_2|\theta_1)$ was actually defined as the normalized version of (2.3.1). This is directly analogous to "start" of the reference prior algorithm in Section 2.2, which gives $\pi^l(\theta_{(m)}|\theta_{(1)},\ldots,\theta_{(m-1)})$.

Reverting to the two parameter case for simplicity, the natural next step is to form the marginal model for θ_1, by integrating out θ_2 with respect to $\pi^l(\theta_2|\theta_1)$, and then to find the reference prior for θ_1 in this marginal model. This approach unfortunately requires the determination of $\boldsymbol{H}(\theta_1)$ for the convolution of $p(x|\theta_1,\theta_2)$ and $\pi^l(\theta_2|\theta_1)$. Such is frequently not available in closed form, limiting the usefulness of the approach. Thus, we consider $Z=(X_1,\ldots,X_t)$, where the X_i are i.i.d. $p(x\,|\,\theta_1,\theta_2)$, so that

$$p^*(z\,|\,\theta_1,\theta_2)=\prod_{i=1}^{t}p(x_i\,|\,\theta_1,\theta_2),$$

and derive the marginal model

$$p^l(z\,|\,\theta_1)=\int p^*(z\,|\,\theta_1,\theta_2)\pi^l(\theta_2\,|\,\theta_1)\,d\theta_2.$$

The reference prior argument (see Bernardo, 1979) suggests that the reference prior for θ_1 should be

$$\pi^l(\theta_1)\propto\ \exp\left\{\int p^l(z\,|\,\theta_1)\log\pi^l(\theta_1\,|\,z)\,dz\right\}, \tag{2.3.2}$$

where

$$\pi^l(\theta_1\,|\,z)=\frac{p^l(z\,|\,\theta_1)\pi^l(\theta_1)}{\int p^l(z\,|\,\theta_1)\pi^l(\theta_1)\,d\theta_1}$$

is the posterior corresponding to the prior $\pi^l(\theta_1)$. This, of course, only defines $\pi^l(\theta_1)$ implicitely, but as $t\to\infty$ (the asymptotic step in the reference prior development) the right hand side of (2.3.2) will typically converge to

$$\exp\left\{\tfrac{1}{2}\int\pi^l(\theta_2|\theta_1)\log|\boldsymbol{h}_1(\theta_1,\theta_2)|d\theta_2\right\}.$$

Interpreting this as the marginal reference prior for θ_1, the natural next step is to multiply this by $\pi^l(\theta_2|\theta_1)$, obtaining the overall reference prior

$$\pi^l(\theta_1,\theta_2)\propto\pi^l(\theta_2|\theta_1)\exp\left\{\tfrac{1}{2}\int\pi^l(\theta_2|\theta_1)\log|\boldsymbol{h}_1(\theta_1,\theta_2)|d\theta_2\right\}.$$

But this is just the numerator in (2.2.2) (when $m=k=2$ and for $j=1$). The denominator in (2.2.2) is just the appropriate normalizing constant. (Details of this argument can be found in Berger and Bernardo (1989b). Even there, however, the argument is intended to just be heuristic, in that no attempt is made to construct conditions under which the limiting arguments are sure to be valid. The point is that, while we consider this to be the motivation, we actually *define* the reference prior via the algorithm in Section 2.2.)

Further stages (when $m > 2$) are handled in exactly the same manner yielding (2.2.2) as the stage-to-stage updating formula. The net result is $\pi^l(\theta)$, the m-stage reference prior on the compact Θ^l.

The final step of the development is to pass to the limit in l. Again as in Bernardo (1979), the most general way to do this is to define the l-posterior

$$\pi^l(\theta|x) = \frac{p(x|\theta)\pi^l(\theta)}{\int p(x|\theta)\pi^l(\theta)d(\theta)}$$

and the reference posterior

$$\pi(\theta|x) = \lim_{l\to\infty} \pi^l(\theta|x) \tag{2.3.3}$$

(the limit perhaps being in a distributional sense). If there exists a prior such that

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)d\theta}$$

then $\pi(\theta)$ would be called the reference prior.

Under reasonable conditions, however, $\pi(\theta)$ can be more simply calculated as

$$\pi(\theta) = \lim_{l\to\infty} \frac{\pi^l(\theta)}{\pi^l(\theta^*)} \tag{2.3.4}$$

where $\theta^* \in \Theta^1$ is any fixed point. (The main condition is that the posterior obtained from this $\pi(\theta)$ be proper.) When $\pi(\theta)$ can be obtained through (2.3.3), but not through (2.3.4), the situation is somewhat pathological (see Section 4.2 for an example), and we would not then have much faith in the entire heuristic argument. Hence we opt for the simple definition (2.3.4). The above argument is meant to be only heuristic, and does not provide our definition of a grouped reference prior; the definition is given via the algorithm in Section 2.2. (For a partial indication of the difficulties of making the heuristic argument precise, see Berger, Bernardo and Mendoza, 1989).

3. THE MULTINOMIAL DISTRIBUTION

Calculation of m-group reference priors for the multinomial distribution is comparatively simple because all distributions involved turn out to be proper, and the integrations in (2.2.1) and (2.2.2) can be done in closed form. In Section 3.1 some preliminary formulas are given; Section 3.2 develops the m-group reference prior; Section 3.3 investigates properties of the reference prior; Section 3.4 is discussion.

3.1. *Preliminaries*

We write the multinomial density for $(k+1)$ cells as

$$p(r_1,\ldots,r_k|\theta_1,\ldots,\theta_k) = \frac{n!}{\left(\prod_{i=1}^{k} r_i!\right)(n-r)!}\left(\prod_{i=1}^{k}\theta_i^{r_i}\right)(1-\delta_k)^{n-r}, \tag{3.1.1}$$

where r_i is the observed frequency in cell i, θ_i is the probability of cell i, n is the total number of observations,

$$r = \sum_{i=1}^{k} r_i, \qquad \text{and} \qquad \delta_j = \sum_{i=1}^{j}\theta_i.$$

(Note that, in our notation, we will suppress the cell count and probability for the $(k+1)$st cell.)

We assume that the θ_i have been ordered and grouped as discussed in Section 2.3.1 (see also Sections 3.4 and 5), and freely use the associated Section 2 notation. Calculation yields

$$\boldsymbol{H}(\theta_1,\ldots,\theta_k) = n\ \text{diag}\left\{\theta_1^{-1},\ldots,\theta_k^{-1}\right\} + n(1-\delta_k)^{-1}\boldsymbol{1}_k,$$

where diag{} stands for the diagonal matrix with given entries, and $\boldsymbol{1}_k$ stands for the $(k\times k)$ matrix of all ones. Further calculation yields

$$\boldsymbol{S}(\theta_1,\ldots,\theta_k) = \frac{1}{n}\ \text{diag}\left\{\theta_1,\ldots,\theta_k\right\} - \frac{1}{n}\theta^t\theta.$$

From this, it is clear that

$$\boldsymbol{S}_j = \frac{1}{n}\text{diag}\left\{\theta_1,\ldots,\theta_{N_j}\right\} - \frac{1}{n}\theta_{[j]}^t\theta_{[j]}$$

and, since $\boldsymbol{S}_j$ has the same structure as $\boldsymbol{S}$, it must be the case that

$$\boldsymbol{H}_j \equiv \boldsymbol{S}_j^{-1} = n\ \text{diag}\left\{\theta_1^{-1},\ldots,\theta_{N_j}^{-1}\right\} + n(1-\delta_{N_j})^{-1}\boldsymbol{1}_{N_j}.$$

Furthermore, an easy calculation yields

$$\begin{aligned}|\boldsymbol{h}_j| &= \text{determinant of the lower right } (n_j\times n_j) \text{ corner of } \boldsymbol{H}_j \\ &= n^{n_j}\left(\prod_{i=N_{j-1}+1}^{N_j}\theta_i^{-1}\right)(1-\delta_{N_{j-1}})\,(1-\delta_{N_j})^{-1}.\end{aligned} \tag{3.1.2}$$

Thus we have available, in closed form, all the quantities needed to apply the reference prior algorithm. For use in the following, define the constants

$$C_{2l-1} = \frac{\pi^l}{(l-1)!} \qquad \text{and} \qquad C_{2l} = \frac{(2\pi)^l}{[(2l-1)(2l-3)\cdots(1)]} \tag{3.1.3}$$

for all positive integers l.

3.2. *The Multinomial m-Group Reference Prior and Posterior*

All distributions that will be encountered have finite mass, so that there is no need to consider a compact sequence $\{\Theta^l\}$. Hence all formulas in Section 2.2 will be applied with the "l" superscripts removed. Note also that, here, (2.1.1) becomes

$$\Theta(\theta_{[j]}) = \left\{ \begin{matrix} \theta_{(j+1)} : \text{ all elements of } \theta_{(j+1)} \text{ are positive} \\ \text{their sum is less than } (1 - \delta_{N_j}) \end{matrix} \right\}.$$

The following lemma provides the crucial calculational development for the reference prior.

Lemma 3.2.1. *For* $j = 1, \ldots, m$,

$$\frac{|\boldsymbol{h}_j(\theta)|^{1/2}}{\int\limits_{\Theta(\theta_{[j-1]})} |\boldsymbol{h}_j(\theta)|^{1/2} d\theta_{(j)}} = C_{n_j}^{-1} \left(\prod_{i=N_{j-1}+1}^{N_j} \theta_i^{-1/2} \right) (1 - \delta_{N_{j-1}})^{(1-n_j)/2} (1 - \delta_{N_j})^{-1/2}. \tag{3.2.1}$$

Proof. Apply Lemma A2.2 in Appendix 2, with $s_i = -\frac{1}{2}$ and $t = -\frac{1}{2}$, and use equation (A2.1). ◁

Theorem 3.2.2. *The m- group reference prior is given by*

$$\pi(\theta) = \left(\prod_{i=1}^{m} C_{n_i}^{-1} \right) \left(\prod_{i=1}^{k} \theta_i^{-\frac{1}{2}} \right) \left(\prod_{i=1}^{m-1} (1 - \delta_{N_i})^{-\frac{1}{2} n_{i+1}} \right) (1 - \delta_{N_m})^{-\frac{1}{2}}. \tag{3.2.2}$$

The m-group reference posterior is

$$\pi(\theta | r_1, \ldots, r_k) \propto \left(\prod_{i=1}^{k} \theta_i^{r_i - \frac{1}{2}} \right) \left(\prod_{i=1}^{m-1} (1 - \delta_{N_i})^{-\frac{1}{2} n_{i+1}} \right) (1 - \delta_{N_m})^{n-r-\frac{1}{2}}. \tag{3.2.3}$$

Proof. Since the $|\boldsymbol{h}_j(\theta)|$ satisfy (2.2.6), Lemma 2.2.1 yields

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

From (2.2.4), (2.2.2), and Lemma 3.2.1 it follows that

$$\pi(\theta) = \left(\prod_{j=1}^{k} \theta_i^{-\frac{1}{2}} \right) \prod_{j=1}^{m} \left[C_{n_j}^{-1} (1 - \delta_{N_{j-1}})^{(1-n_j)/2} (1 - \delta_{N_j})^{-\frac{1}{2}} \right].$$

Telescoping the product yields (3.2.2), and (3.2.3) is immediate from (3.1.1). ◁

Special Cases. Two interesting special cases are the 1-group and the k-group reference priors.

Case 1: The one-Group Reference Prior. If $m = 1$, (3.2.2) yields

$$\pi(\theta) = C_k^{-1} \left(\prod_{i=1}^{k} \theta_i^{-1/2} \right) (1 - \delta_k)^{-1/2}, \tag{3.2.4}$$

which is, of course, Jeffreys's noninformative prior.

Case 2. The k-Group Reference Prior. If $m = k$ (i.e., all group sizes are $n_i = 1$), then (3.2.2) yields

$$\pi(\theta) = (\pi^{-k}) \prod_{i=1}^{k} \left[\theta_i^{-1/2} (1 - \delta_i)^{-1/2} \right]. \tag{3.2.5}$$

This is actually the reference prior that we will recommend for typical use.

3.3. *Properties of the Reference Prior and Posterior*

3.3.1. Marginal Distributions

The marginal probability distribution of $(r_1, \ldots, r_l)$ is also multinomial, with cell probabilities $\theta_1, \ldots, \theta_l$ and sample size n. (All other observations are lumped together into the new $(l+1)$st cell, which is the union of the $(l+1)$st through $(k+1)$st cells in the original multinomial.) It is of considerable interest to see whether or not the m-group reference prior "marginalizes" consistently, in that the reference prior for the "collapsed" $(l+1)$-cell multinomial be the same as that obtained by finding the marginal distribution of $\theta_1, \ldots, \theta_l$ from the original m-group reference prior. The following lemma provides the answer. For simplicity we assume, with the exception of (3.3.3), that the marginalization is done over groups.

Lemma 3.3.1. *For the prior $\pi(\theta)$ in (3.2.2), the marginal reference prior for $\theta_{(1)}, \ldots, \theta_{(j)}$ is*

$$\pi_{[j]}(\theta_{(1)}, \ldots, \theta_{(j)}) = \left(\prod_{i=1}^{j} C_{n_i}^{-1} \right) \left(\prod_{i=1}^{N_j} \theta_i^{-\frac{1}{2}} \right) \left(\prod_{i=1}^{j-1} (1 - \delta_{N_i})^{-\frac{1}{2} n_{i+1}} \right) (1 - \delta_{N_j})^{-1/2} . \tag{3.3.1}$$

The marginal reference prior for $\theta_{(1)}$ is

$$\pi_{[1]}(\theta_{(1)}) = C_{n_1}^{-1} \left(\prod_{i=1}^{n_1} \theta_i^{-\frac{1}{2}} \right) (1 - \delta_{n_1})^{-\frac{1}{2}} , \tag{3.3.2}$$

while that for $\theta_1, \ldots, \theta_l$, when $l < n_1$, is

$$\pi_{[1,l]}(\theta_1, \ldots, \theta_l) = \frac{C_{n_1 - l}}{C_{n_1}} \left(\prod_{i=1}^{l} \theta_i^{-\frac{1}{2}} \right) (1 - \delta_l)^{(n_1 - l - 1)/2} . \tag{3.3.3}$$

Proof. The expression for $\pi_{[j]}$ follows from Lemma A2.2 in Appendix 2, applied iteratively for $l = m, m-1, \ldots, j+1$ with all $s_i = -\frac{1}{2}$ and $t = -\frac{1}{2}$; note that

$$t + n_l + \sum_{i=N_{l-1}+1}^{N_l} s_i = -\tfrac{1}{2} + n_l - \tfrac{1}{2}n_l = -\tfrac{1}{2} + \tfrac{1}{2}n_l,$$

and

$$\prod_{i=N_{l-1}+1}^{N_l} B\left(s_i + 1, \quad t + N_i - i + 1 + \sum_{p=i+1}^{N_l} S_p\right) = \prod_{i=1}^{n_l} B\left(\tfrac{1}{2}, \tfrac{i}{2}\right) = C_{n_l}.$$

The expression for $\pi_{[1]}$ is immediate. The formula for $\pi_{[1,l]}$ follows from iterative application of Lemma A2.1 in Appendix 2, for $p = n_1, \ldots, l+1$, with $\theta = \theta_p$, $s = -1/2$, $t = (n_1 - p - 1)/2$, and $\delta = \delta_{p-1}$. ◁

Typically, of course, one will be interested in the marginal posteriors, rather than the marginal priors. These are immediate from the marginal priors, however; simply multiply by the likelihood from the corresponding marginal multinomial distribution. For instance, the marginal posterior for $\theta_{(1)}$ is

$$\pi_{[1]}(\theta_{(1)}|r_1, \ldots, r_{n_1}) \propto \left(\prod_{i=1}^{n_1} \theta_i^{r_i - \frac{1}{2}}\right)(1 - \delta_{n_1})^{n - \sum_{i=1}^{n_1} r_i - \frac{1}{2}}.$$

(This could, of course, also have been obtained by calculating the marginal density of $\theta_{(1)}$ from (3.2.3).)

To return to the question posed at the beginning of this Section, we see that (3.3.1) and (3.3.2) are of exactly the same form as (3.2.2). Hence, if we reduce consideration to the first j groups of parameters $\theta_{(1)}, \ldots, \theta_{(j)}$, the answers obtained by marginalizing from the original m-group reference prior are identical to the answers obtained by treating the j groups as a "new" multinomial problem. This property may be viewed to be valuable, because of the following well known example.

Example. Suppose we have the multinomial model with $\theta_1, \ldots, \theta_{n_1}$. Consider the 1-group reference prior, which is Jeffreys's prior, here

$$\pi(\theta_1, \ldots, \theta_{n_1}) \propto \left(\prod_{i=1}^{n_1} \theta_i^{-1/2}\right)(1 - \delta_{n_1})^{-1/2}.$$

The posterior means of the θ_i are then

$$E[\theta_i|r_1, \ldots, r_{n_1}] = \frac{r_i + \frac{1}{2}}{n + \frac{1}{2}(n_1 + 1)}.$$

Now suppose one notices that the $(n - \sum_{i=1}^{n_1} r_i)$ observations in the $(n_1 + 1)$st cell could have been further subdivided into $n_2 + 1$ new categories. (Say one discovers a new classification scheme into n_2 categories for elements in this cell.) Then one has the apparent

option of adding $r_{n_1+1}, \ldots, r_{n_1+n_2}$ and $\theta_{n_1+1}, \ldots, \theta_{n_1+n_2}$ to the multinomial model. If one did so and used Jeffreys's prior, which would then be

$$\pi(\theta_1, \ldots, \theta_{n_1+n_2}) \propto \left(\prod_{i=1}^{n_1+n_2} \theta_i^{-1/2}\right)(1-\delta_{n_1+n_2})^{-1/2},$$

a calculation shows that the posterior means of the θ_i would now be

$$E[\theta_i | r_1, \ldots, r_{n_1+n_2}] = \frac{r_i + \frac{1}{2}}{n + \frac{1}{2}(n_1+n_2+1)}.$$

The creation of new cells can thus have a pronounced effect on posterior beliefs about existing cells.

The m-group reference prior is essentially immune to this difficulty, since the marginal prior (and posterior) for, say, $\theta_{(1)}$ is the same no matter how many additional groups (or cells) are added. This needs two qualifications, however. The first is that the marginalization property does not hold for all groups; it holds only for an initial sequence $(\theta_{(1)}, \ldots, \theta_{(j)})$. Of course, by construction it is $\theta_{(1)}$ that is supposed to be of interest, so that this should not be an objection.

The second limitation of the marginalization property is that it does not hold *within*, say, $\theta_{(1)}$. This can be seen from (3.3.3), where the marginal reference prior is *not* of the form (3.3.2), and is hence different from that which would have been obtained had the problem been originally confined to $(\theta_1, \ldots, \theta_l)$. We defer further discussion of this issue to Section 3.4.

3.3.2. Moments of the Reference Priors

For comparing and understanding the group reference priors, it is useful to have expressions for their moments.

Lemma 3.3.2. *For $N_{j-1}+1 \le l \le N_j$ and $\pi(\theta)$ defined by (3.2.2),*

$$E^{\pi}[\theta_l^s] = \prod_{i=1}^{j}\prod_{p=1}^{s}\left[1 + \frac{n_i}{(2p-1)}\right]^{-1}.$$

If all $n_i = 1$,

$$E^{\pi}[\theta_j^s] = \left[\frac{(2s)!}{(2^s s!)^2}\right]^j.$$

If $s = 1$ (but n_i is arbitrary)

$$E^{\pi}[\theta_l] = \prod_{i=1}^{j}(1+n_i)^{-1},$$

Finally, the mean for the $(k+1)$st cell is

$$E^{\pi}[1 - \delta_{N_m}] = \prod_{i=1}^{m}(1+n_i)^{-1}.$$

Proof. This is similar to the proof of Lemma A2.2 in Appendix 2, using Lemma A2.1 iteratively. ◁

3.4. *Discussion*

The multinomial scenario dramatically demonstrates how the m-group reference prior can "decouple" groups of coordinates. Thus the inferences obtained for $\theta_{(1)}$ will depend only on $(r_1, \ldots, r_{n_1})$ and n, and not on what happens in other cells (or how many other cells there are). This is a natural property when, indeed, $\theta_{(1)}$ is of interest and the other parameters are nuisance parameters. (Note that standard noninformative priors, such as Jeffreys's prior, do not have this property.)

The desirability of this property can, however, be questioned. It requires an asymmetric treatment of the θ_i, and in problems where there is a small number of fixed "indistinguishable" cells, such asymmetry may be unappealing.

To dramatize the difference, consider the two extremes of the 1-group and the k-group reference priors in (3.2.4) and (3.2.5). From Lemma 3.3.2, one sees, for instance, that the prior means of the θ_i are $(1+k)^{-1}$ for the 1-group (Jeffreys's) reference prior, but are 2^{-i} for the k-group reference prior. Thus the 1-group reference prior treats the θ_i equally, while the k-group reference prior gives exponentially decreasing mass to the θ_i as i increases.

This situation clearly demonstrates the impossibility of unambiguously defining "noninformative". Initially it seems reasonable to insist that a noninformative prior for a multinomial problem be exchangeable, and to require that it have the marginalization property; but these requirements are completely incompatible!. Through consideration of a variety of examples we have convinced ourselves that the marginalization property is typically more important, and hence that the k-group reference prior in typically more attractive, but some flexibility is clearly required. In particular, if one has a small number of cells of interest, between which exchangeability seems very natural, it would clearly be tempting to use a 2-group reference prior, guaranteeing the marginalization property for the group of parameters of interest, while preserving exchangeability within the group. Thus, one might well want to be "subjectively noninformative".

If it is only $\theta_{(1)}$ that is of interest, note that there is no reason to even formally consider use of an m-group reference prior. The result will simply be that obtained by collapsing the original multinomial to the (n_1+1)-cell multinomial, with cell probabilities determined by $\theta_{(1)}$, and then using Jeffreys's prior for $\theta_{(1)}$. Thus, in practice, one needs to formally use the m-group reference prior only if more than the first group is of interest. Of course, this will typically be the case if the recommended full k-group reference prior in (3.2.5) is utilized and several of the θ_i are of interest.

4. VARIANCE COMPONENTS

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 (2.2.5). 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

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

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 $\sqrt{n-1} \leq \psi(\tau^2/\sigma^2) \leq \sqrt{n}$; thus, replacing C_n by 1 and ψ by a constant is reasonable for all but very small n.

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

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

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

As indicated in Section 2.3.1, 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*

In Appendix 3, the development of the reference prior for the ordered grouping $\{\mu, \sigma^2, \tau^2\}$ is presented. This analysis is typical of the analyses for all cases in which C_n does not occur. The analysis for the C_n cases is considerably more technical; see Berger and Bernardo (1989b) for details.

To implement the algorithm in Section 2.2, 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. (As mentioned in Section 2.3.3, we could have defined $\pi(\theta)$ as a distribution yielding the "reference posterior" defined in (2.3.3). Here the posterior defined by (2.3.3) would exist, but would be degenerate at $\tau^2 = 0$. This is clearly not very attractive.) 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.

4.3. *Posterior Calculations*

Note that the likelihood function is proportional to

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

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

$$\pi(\mu, \sigma^2, \tau^2) = \frac{\left[\dfrac{(n-1)}{\sigma^4} + \dfrac{n^{\rho-2}}{(n\tau^2+\sigma^2)^\rho}\right]^{1/2}}{\sigma^\alpha \tau^\beta (n\tau^2+\sigma^2)^\gamma}$$

for certain constants α, β, γ, and ρ.

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

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

a form which clearly includes all posterior moments of the parameters. Interestingly, this calculation requires only one-dimensional numerical integration. To see this, first transform to the variables (μ, σ^2, v), where $v = \tau^2/\sigma^2$. It is then straightforward to integrate over μ, and then σ^2, in closed form. Only the final integral over v requires numerical integration. See Berger and Bernardo (1989b) for details.

5. CONCLUSIONS

We have considered m-group reference priors as a possible solution to the clearcut need in multiparameter problems for developing noninformative priors with limited dependencies between groups of parameters (especially parameters of interest and nuisance parameters). There are different possible views on the success of the solution.

The least commital view is that the m-group reference prior method succeeds in generating a variety of interesting possible noninformative priors. For instance, (3.2.5) is very interesting for its marginalization property (and is, to our knowledge, new), while several of the priors in Table 1 have not been seen before. As candidates for in-depth study or for Bayesian sensitivity studies, these can be very useful noninformative priors (especially because of their ability to "decouple" parameters). In this regard, the 1-group (Jeffreys's) and k-group reference priors are likely to exhibit the greatest differences and, if a Bayesian analysis yields essentially the same answer for either prior there is reason to be confident in the answer.

The more optimistic view about m-group reference priors, in particular about k-reference priors (each group has only one parameter), is that they provide the best available "automatic" priors for general use. Our preference for the k-reference prior is, for the most part, empirically based. In all examples we have considered (including many of the "counterexamples" to Jeffreys's or other noninformative priors), the k-reference priors have yielded very sensible results.

Our enthusiasm for k-reference priors is slightly tempered by two issues we have touched on. First, they can be technically difficult or ambiguous to derive, especially when limits over $\{\Theta^l\}$ are needed. This can obviously reduce their pragmatic appeal, (although derivation of the k-group reference prior could be considered to be the theoreticians job, in which case the user is not affected).

The second difficulty with k-group reference priors is that they can depend on the ordering of the parameters and it can be difficult to decide on a complete ordering, especially for the nuisance parameters. One possibility is to order the parameters of interest, but group all nuisance parameters together (or maybe have several groups, when natural). Ordering within groups does not matter, and grouping nuisance parameters in rarely harmful, so this is often a sensible resolution of the ordering problem. A second possible solution is to try several different orderings, and see if it matters. (Note, indeed, that the ordering or groupings of nuisance parameters frequently is immaterial, as in the multinomial problem.) A third possible solution is to use the average of all the k-group reference priors from feasible orderings.

Many other issues could be raised. One of the most important is that of inference about functions $\varphi(\theta_1, \ldots, \theta_k)$ of the parameters. Reference prior theory (see, e.g., Bernardo, 1979 or Berger and Bernardo, 1989a) requires a reparameterization, with $\varphi(\theta_1, \ldots, \theta_k)$ defined as the "new" $\theta_{(1)}$, before the reference prior can be determined. Luckily shortcuts appear to be available, so that it is not necessary to completely redo the reference prior development for every function that is of interest. This work will be reported elsewhere.

ACKNOWLEDGEMENTS

We are grateful to Ke-Ying Ye for assisting in the determination of Table 1.

REFERENCES

Bayarri, M. J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trab. Estadist.* **32**, 18–31.

Bayarri, M. J. (1985). Bayesian inference on the parameters of the Beta distribution. *Statistics and Decisions* Suppl. Issue **2**, 17–22.

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). Reference priors in a variance components problem. *Proccedings of the Indo-USA Workshop on Bayesian Analysis in Statistics and Econometrics* (P. Goel, ed.). Bangalore, India (in press).

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).

Bernardo, J. M. (1980). A Bayesian analysis of classical hypothesis testing. *Bayesian Statistics 1* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) Valencia: University Press, 605–647 (with discussion).

Bernardo, J. M. (1981). Reference decisions. *Symposia Matematica* **25**, 85–94.

Bernardo, J. M. (1982). Contraste de modelos probabilísticos desde una perspectiva Bayesiana. *Trab. Estadist.* **33**, 16–30.

Bernardo, J. M. (1985). On a famous problem of induction. *Trab. Estadist.* **36**, 24–30.

Bernardo J. M. and Girón, F. J. (1988). A Bayesian analysis of simple mixture problems. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) Oxford: University Press, 67–78.

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

Cifarelli, D. M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations. *Probability and Bayesian Statistics* (R. Viertl ed.). New York: Plenum, 83–95.

Consonni, G. and Veronese, P. (1988). A note on coherent invariant distributions as non-informative priors for exponential and location-sacale families. *Studi Statistici* **19**, Universitá L. Bocconi, Milan.

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradox in Bayesian and structural inference. *J. Roy. Statist. Soc. B* **35**, 189–233 (with discussion).

Eaves, D. M. (1983). On Bayesian nonlinear regression with an enzyme example. *Biometrika* **70**, 373–379.

Eaves, D. M. (1985). On maximizing the missing information about a hypothesis. *J. Roy. Statist. Soc. B* **47**, 263–265.

Ferrandiz, J. R. (1982). Una solución Bayesiana a la paradoja de Stein. *Trab. Estadist.* **33**, 31–46.

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

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

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

Lindley, D. V. (1988). Statistical inference concerning Hardy-Weinberg equilibrium. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) Oxford: University Press, 307–326.

Mendoza, M. (1987). A Bayesian analysis of a generalized slope ratio bioassay. *Probability and Bayesian Statistics* (R. Viertl, ed.) London: Plenum Press, 357–364.

Mendoza, M. (1988). Inferences about the ratio of linear combinations of the coefficients in a multiple regression model. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) Oxford: University Press, 705–711.

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

Sendra, M. (1982). Distribución final de referencia para el problema de Fieller-Creasy. *Trab. Estadist.* **33**, 55–72.

Stein, M. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.* **30**, 877–880.

APPENDIX 1

Calculation of the $\boldsymbol{h}_i$

With the notation of Section 2.1, define $\boldsymbol{B}_j = (\boldsymbol{A}_{j1}\boldsymbol{A}_{j2}\cdots\boldsymbol{A}_{j\,j-1})$, $j = 2,\ldots,m$, of sizes $(n_j \times N_{j-1})$. It is straightforward to verify that, for $j = 1,\ldots,m$

$$\boldsymbol{h}_j = (\boldsymbol{A}_{jj} - \boldsymbol{B}_j\boldsymbol{H}_{j-1}\boldsymbol{B}_j^t)^{-1} \tag{A1.1}$$

and

$$\boldsymbol{H}_j = \begin{pmatrix} \boldsymbol{H}_{j-1} + \boldsymbol{H}_{j-1}\boldsymbol{B}_j^t\boldsymbol{h}_j\boldsymbol{B}_j\boldsymbol{H}_{j-1} & -\boldsymbol{H}_{j-1}\boldsymbol{B}_j^t\boldsymbol{h}_j \\ -\boldsymbol{h}_j\boldsymbol{B}_j\boldsymbol{H}_{j-1} & \boldsymbol{h}_j \end{pmatrix}, \tag{A1.2}$$

where any entry containing a factor of $\boldsymbol{H}_0$ is to be omitted. Thus one may calculate the matrices $\boldsymbol{H}_1,\ldots,\boldsymbol{H}_m$, and hence $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_m$, iteratively.

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

$$\boldsymbol{B}_{i+1} = (c_i\boldsymbol{B}_i, \quad \boldsymbol{A}_{i+1\,i}) \tag{A1.3}$$

for some constant c_i. Then, (A1.1), (A1.2), and (A1.3) can be used to show that

$$\boldsymbol{h}_{i+1} = \left[\boldsymbol{A}_{i+1\,i+1} + c_i^2\boldsymbol{A}_{ii} - 2c_i\boldsymbol{A}_{i+1\,i} - \boldsymbol{h}_i\,(c_i\boldsymbol{A}_{ii} - \boldsymbol{A}_{i+1\,i})^2\right]^{-1}. \tag{A1.4}$$

This is particularly useful when (A1.3) holds for all i, which often occurs in patterned covariance matrices, since then (A1.4) can used to iteratively determine all the $\boldsymbol{h}_i$, starting with $\boldsymbol{h}_1 = \boldsymbol{A}_{11}^{-1}$, and defining $c_1 = 1$.

APPENDIX 2

Integrations for the Multinomial Problem

For convenience, define

$$B(\alpha,\beta) = \int_0^1 \xi^{\alpha-1}(1-\xi)^{\beta-1}d\xi,$$

and note the easily verified facts that

$$\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \frac{B(\alpha,\beta+1)}{B(\alpha,\beta)} = \frac{\beta}{\alpha+\beta}$$

Using these facts and observing that

$$B\left(\tfrac{1}{2},\tfrac{1}{2}\right) = \pi \quad \text{and} \quad B\left(\tfrac{1}{2},1\right) = 2,$$

induction on l establishes the useful expression (see (3.1.3))

$$C_l = \prod_{i=1}^{l} B\left(\tfrac{1}{2},\tfrac{i}{2}\right). \tag{A2.1}$$

Lemma A2.1. *If $s > -1$ and $t > -1$, then*

$$\int_0^{1-\delta} \theta^s (1-\theta-\delta)^t d\theta = (1-\delta)^{s+t+1} B(s+1, t+1).$$

Proof. Change variable to $y = \theta/(1-\delta)$. ◁

Lemma A2.2. *If $t > -1$ and all $s_i > -1$, then*

$$\int_{\Theta(\theta_{[j-1]})} \left(\prod_{i=N_{j-1}+1}^{N_j} \theta_i^{s_i} \right) (1-\delta_{N_j})^t d\theta_{(j)}$$

$$= \left(1-\delta_{N_{j-1}}\right)^{t+n_j+\Sigma_{i=N_{j-1}+1}^{N_j} s_i} \prod_{i=N_{j-1}+1}^{N_j} B\left(s_i+1,\ t+N_j-i+1+\Sigma_{l=i+1}^{N_j} s_l\right).$$

Proof. Observe that

$$1-\delta_{N_j} = 1-\delta_{N_{j-1}} - \Sigma_{i=N_{j-1}+1}^{N_j} \theta_i,$$

and then apply Lemma A2.1 iteratively for $i = N_j, N_j-1, \ldots, N_{j-1}+1$. ◁

APPENDIX 3

Reference Priors for Variance Components

We apply the algorithm in Section 2.2, with the Θ^l defined by (4.2.1). Note that the Fisher information matrix for (μ, σ^2, τ^2) is

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

so that

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

We shall analize the 3-group case $\{\mu, \sigma^2, \tau^2\}$. In the notation of Section 2.1, $\theta_{(1)} = \theta_1 = \mu$, $\theta_{(2)} = \theta_2 = \sigma^2$, and $\theta_{(3)} = \theta_3 = \tau^2$. Also, $\boldsymbol{S}$ satisfies (A1.3), with $\boldsymbol{B}_2 = 0$, $c_2 = 1$, and $\boldsymbol{A}_{32} = -2\sigma^4/[pn(n-1)]$, so that

$$\boldsymbol{h}_1 = \sigma_{11}^{-1} = \frac{pn}{n\tau^2+\sigma^2}$$

and (A1.4) yields

$$\begin{aligned} \boldsymbol{h}_2 &= \left[\boldsymbol{A}_{22} + \boldsymbol{A}_{11} - \boldsymbol{h}_1 \boldsymbol{A}_{11}^2\right]^{-1} \\ &= \frac{p(n-1)}{2\sigma^4} \\ \boldsymbol{h}_3 &= \left[\boldsymbol{A}_{33} + \boldsymbol{A}_{22} - 2\boldsymbol{A}_{32} - \boldsymbol{h}_2 (\boldsymbol{A}_{22} - \boldsymbol{A}_{32})^2\right]^{-1} \\ &= \frac{pn^2}{2(n\tau^2 + \sigma^2)^2} \,. \end{aligned}$$

Of course, $\boldsymbol{h}_3$ could have been obtained directly from $\boldsymbol{H}$.

Start: To begin,

$$\begin{aligned} \pi_3^l(\tau^2|\mu,\sigma^2) &= \frac{\left|\boldsymbol{h}_3(\mu,\sigma^2,\tau^2)\right|^{1/2} 1_{(e_l,f_l)}(\tau^2)}{\displaystyle\int_{e_l}^{f_l} \left|\boldsymbol{h}_3(\mu,\sigma^2,\tau^2)\right|^{1/2} d\tau^2} \\ &= \frac{n(n\tau^2+\sigma^2)^{-1} 1_{(e_l,f_l)}(\tau^2)}{\log\left[(nf_l+\sigma^2)/(ne_l+\sigma^2)\right]} \,. \end{aligned}$$

Iteration for $j = 2$: Since $\boldsymbol{h}_2$ does not depend on τ^2,

$$\begin{aligned} E^l\left[\log|\boldsymbol{h}_2|\,\middle|\mu,\sigma^2\right] &= \int (\log|\boldsymbol{h}_2|)\pi_3^l(\tau^2|\mu,\sigma^2)d\tau^2 \\ &= \log|\boldsymbol{h}_2| = \log\frac{p(n-1)}{2\sigma^4}. \end{aligned}$$

Hence

$$\begin{aligned} \pi_2^l(\sigma^2,\tau^2|\mu) &= \frac{n(n\tau^2+\sigma^2)^{-1}1_{(e_l,f_l)}(\tau^2)}{\log\left[(nf_l+\sigma^2)/(ne_l+\sigma^2)\right]} \cdot \frac{\left[p(n-1)/(2\sigma^4)\right]^{1/2} 1_{(c_l,d_l)}(\sigma^2)}{\displaystyle\int_{c_l}^{d_l}\left[p(n-1)/(2\sigma^4)\right]^{1/2} d\sigma^2} \\ &= \frac{n(n\tau^2+\sigma^2)^{-1}1_{(e_l,f_l)}(\tau^2)}{\log\left[(nf_l+\sigma^2)/(ne_l+\sigma^2)\right]} \cdot \frac{1_{(c_l,d_l)}(\sigma^2)}{\sigma^2\log(d_l/c_l)} \,. \end{aligned}$$

Iteration for $j = 1$: Since neither $\boldsymbol{h}_1(\mu,\sigma^2,\tau^2)$ nor $\pi_2^l(\sigma^2,\tau^2|\mu)$ depend on μ,

$$E^l\left[\log|\boldsymbol{h}_1(\mu,\sigma^2,\tau^2)|\,\middle|\mu\right] = K(c_l,d_l,e_l,f_l).$$

Hence

$$\begin{aligned} \pi^l(\mu,\sigma^2,\tau^2) &= \pi_1^l(\mu,\sigma^2,\tau^2) \\ &= \pi_2^l(\sigma^2,\tau^2|\mu)(b_l-a_l)^{-1}1_{a_l,b_l}(\mu). \end{aligned}$$

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

$$\begin{aligned} \pi(\mu,\sigma^2,\tau^2) &= \lim_{l\to\infty} \frac{\pi_2^l(\sigma^2,\tau^2|\mu)1_{(a_l,b_l)}(\mu)}{\pi_2^l(1,1|0)} \\ &= \lim_{l\to\infty} \frac{(n+1)}{\sigma^2(n\tau^2+\sigma^2)} \cdot \frac{\log\left[(nf_l+1)/(ne_l+1)\right]}{\log\left[(nf_l+\sigma^2)/(ne_l+\sigma^2)\right]} 1_{\Theta^l}(\mu,\sigma^2,\tau^2) \\ &= \frac{(n+1)}{\sigma^2(n\tau^2+\sigma^2)} \qquad \text{(since } f_l \to \infty \text{ and } c_l \to 0\text{).} \end{aligned}$$

The proportionality constant $(n+1)$ is irrelevant, so $\sigma^{-2}(n\tau^2+\sigma^2)^{-1}$ is the reference prior.