Bayesian Methods for Variance Component Models

Li Sun; John S. J. Hsu; Irwin Guttman; Tom Leonard

Journal of the American Statistical Association, Vol. 91, No. 434. (Jun., 1996), pp. 743-752.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199606%2991%3A434%3C743%3ABMFVCM%3E2.0.CO%3B2-O

Journal of the American Statistical Association is currently published by American Statistical Association.

Bayesian Methods for Variance Component Models

Li SUN, John S. J. HSU, Irwin GUTTMAN, and Tom LEONARD

An exact Bayesian analysis can be performed for normal theory variance components models, using importance sampling. But remarkably accurate approximations are available using the Laplacian T-approximation introduced by Leonard, Hsu, and Ritter. Instead of maximizing the joint posterior density, conditional upon the parameter of interest, a device described by O'Hagan is used first. The Bayesian estimators are compared to the Lindley–Stein shrinkage estimators and the Lindley–Smith joint modal estimators. It is confirmed that joint modes can overcollapse toward prior hypotheses, when compared with more sensible Bayesian procedures. This is referred to as a "collapsing phenomenon." A numerical example from the one-way random-effects model is considered, and the risks of the different estimators are simulated under a variety of loss functions. It is concluded that although the Lindley–Stein estimator performs well, a full hierarchical Bayesian analysis performs at least equally well, while permitting more detailed finite-sample inference regarding any parameter of interest.

KEY WORDS: Analysis of variance; Empirical Bayes; Hierarchical Bayes; Importance sampling; Joint mode; Laplacian T approximation; Marginal mode; Maximum component loss; Risk function; Shrinkage estimator.

1. INTRODUCTION AND SUMMARY

Before the 1980s it was difficult to handle the computations for Bayesian variance components models, and indeed many authors concentrated on calculating posterior modes, using maximization techniques. With the advent of such powerful techniques as importance sampling, Markov chain iterations, and modern usages of Laplacian approximations, as facilitated by the speed, memory, and flexibility of more recent computer packages, it is now possible to provide detailed finite-sample inference for many such models.

As indicated by Leonard, Hsu, and Ritter (1994), Markov chain techniques (e.g., the Metropolis algorithm and Gibbs sampler) should be reserved for those complicated models that cannot be handled well using importance sampling. In this article we demonstrate that importance sampling can suffice for variance components models. Nonetheless, Laplacian techniques also become straightforward. These techniques are remarkably accurate and do not consume much computer time.

In Section 2, a quite general formulation of variance components models for normally distributed observations is stated. In Section 3, the Laplacian T-approximation, introduced by Leonard et al. (1994) is discussed. This is preferred, in the current situation, to the Laplacian techniques introduced by Leonard (1982), and Tierney and Kadane (1986), owing to a side effect of the "collapsing phenomenon" discussed later. In Section 4 a general Bayesian analysis is outlined, and the Laplacian and importance sampling techniques are developed in detail.

During the 1970s, joint posterior modes were quite popular for estimating the first-stage prior parameters. In Section 5 a serious "collapsing phenomenon" is described, showing that posterior modes should be used with extreme caution. Our main alternative suggestions to joint posterior modes are summaries of the marginal posterior densities of the first-stage parameters, as approximated via the techniques of Sections 3 and 4 or computed exactly using importance sampling. These have frequency properties similar to those of empirical Bayes estimators of the "Bayes–Stein" type and to estimators proposed by O'Hagan (1976). The latter involve a convenient substitution of the marginal posterior modes of the variance components.

In Sections 6 and 7 the collapsing phenomenon is further demonstrated by a consideration of frequency properties. These can be expressed algebraically when the number of first-stage parameters is large, and are calculated numerically when this number is finite.

In Section 8 the one-way random-effects model is considered in detail. We also consider the two-way random-effects model and split-plot designs, but for brevity of presentation, these models are not further investigated here.

In Section 9 the Laplacian approximations are validated by numerical comparisons with the exact results. In Section 10 frequency properties are investigated under a variety of loss functions, and in Section 11 some properties of "maximum component" loss are described.

Applications of these models include such diverse areas as the ozone layer (Reinsel and Tiao 1987), agricultural trials (Searle, Casella, and McCullogh 1992), genetic trait models in Animal Science (Foulley, Gianola, and Im 1990), psychometric models (Lord and Novick 1967), and signal processing in electrical engineering (Brennan 1994; Kalman 1960). An extensive literature relating to actuarial science business, credibility theory, and Bayesian graduation includes work of Fortney and Miller (1987), Hickman and Miller (1981), Miller and Fortney (1984), and Miller and Hickman (1975).

Li Sun is Assistant Professor, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. John Hsu is Assistant Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106. Irwin Guttman is Professor and Chair, Department of Statistics, State University of New York, Buffalo, NY 14214. Tom Leonard is Professor, Department of Statistics, University of Wisconsin, Madison, WI 53706, and Department of Mathematics and Statistics, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom. The authors thank two referees and an associate editor for their many generous comments and suggestions.

Journal of the American Statistical Association
June 1996, Vol. 91, No. 434, Theory and Methods

2. MODEL FORMULATION

Consider an $M \times 1$ observation vector $\mathbf{y}$, which is taken to possess a multivariate normal distribution, conditional on its mean vector

$$\mathbf{X}\boldsymbol{\theta} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\gamma} \tag{1}$$

and covariance matrix $\phi\mathbf{I}_M$. Here $\mathbf{X} = (\mathbf{U}, \mathbf{Z})$ is a specified $M \times p$ design matrix, and $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters. Furthermore, $p = q_0 + q_1 + \cdots + q_s$, and the unknown $\boldsymbol{\theta}$ can be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T)^T = (\boldsymbol{\alpha}^T, \boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \ldots, \boldsymbol{\gamma}_s^T)^T$, where $\boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_s$ are $q_0 \times 1, q_1 \times 1, \ldots, q_s \times 1$ subvectors. Assume that $\boldsymbol{\alpha}$ comprises fixed effects and that $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_s$ are vectors that are a priori independent. The $\boldsymbol{\gamma}_j$ are assumed to be the parameters of primary interest. It is supposed that, $\boldsymbol{\gamma}_j$ possesses a distribution that is *multivariate normal* with mean vector $\mu_j\mathbf{e}_{q_j}$ and covariance matrix $\sigma_j^2\mathbf{I}_{q_j}$, where $\mathbf{e}_{q_j}$ denotes a $q_j \times 1$ vector of 1s and $\mathbf{I}_{q_j}$ is the $q_j \times q_j$ identity matrix. Then $\boldsymbol{\alpha}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_s$ can be interpreted as "first-stage parameters." Assume that $q_j \geq 3$, for $j = 1, \ldots, s$, so that the σ_j^2 can be identified from the current data, and meaningful Bayes–Stein-type shrinkage estimators can be obtained for the elements of the first-stage parameter vectors $\boldsymbol{\gamma}_j$, whenever proper distributions are assumed for the σ_j^2 in the prior assessment. But if the uniform distributions, described later, are assumed for the σ_j^2, then it is necessary to assume that $q_j > 3$ for $j = 1, 2, \ldots, s$, to ensure that the posterior distribution, described in Section 4, adequately reflects the random-effects assumptions.

This very general and flexible formulation covers a multitude of special cases. Although unbalanced designs are included in our general formulation, two simple balanced designs are described as follows:

- **One-way random-effects model.** For $j = 1, \ldots, m$ and $k = 1, \ldots, n$, observations y_{jk} have independent normal distributions, conditional on $\theta_1, \ldots, \theta_m$, with means θ_j and common variance ϕ. Furthermore, the θ_j are a random sample from a normal distribution with mean μ and variance σ^2.
- **Two-way random-effects model with interactions.** For $i = 1, \ldots, m, j = 1, \ldots, t$ and $k = 1, \ldots, n$, observations y_{ijk} have normal distributions, with means $\theta_i^A + \theta_j^B + \theta_{ij}^{AB}$ and variance ϕ. Here it can be convenient to regard the marginal effects θ_i^A and θ_j^B as fixed effects satisfying $\sum_{i=1}^m \theta_i^A = \sum_{j=1}^t \theta_j^B = 0$, and to take the interaction effects θ_{ij}^{AB} to constitute a random sample from a normal distribution with mean μ and variance σ^2. This device, introduced by Laird (1978), permits the reasonable estimation of just two variance components, ϕ and σ^2. Appropriate linear transformations of the marginal effects yield $q_0 = m + t - 2$ distinct fixed effects and hence a special case of (1). Sun (1992) instead regarded the marginal effects as random and then estimated two further variance components. Under his formulation, the $\mathbf{X}^T\mathbf{X}$ matrix is singular; however, his posterior distribution is still proper as a general condition discussed in Section 4 is satisfied.

We may complete our prior assumptions in the more general case by taking $\boldsymbol{\alpha}, \mu_1, \ldots, \mu_s, \sigma_1^2, \ldots, \sigma_s^2$, and ϕ to be independent. For convenience, take $\boldsymbol{\alpha}$ and the μ_j to be uniformly distributed over their species of possible values. Furthermore, assume the prior density

$$\pi(\sigma_j^2) \propto (\sigma_j^2)^{-1/2(\nu_j+2)} \exp\{-\nu_j\zeta_j/2\sigma_j^2\} \qquad (0 < \sigma_j^2 < \infty)$$

for σ_j^2. Here ζ_j^{-1} denotes the prior mean of the precision σ_j^{-2} and ν_j can be interpreted as a prior "sample size." Note that the special case $\nu_j = -2$ and $\zeta_j = 0$ is of interest, because σ_j^2 is then uniformly distributed over $(0, \infty)$ (Leonard 1976; Strawderman 1971). Finally, we take $\nu_0\zeta_0/\phi$ to possess a chi-squared distribution with ν_ϕ degrees of freedom.

Box and Tiao (1968) addressed the problem of drawing inferences about elements of the $\boldsymbol{\gamma}_j$ in special cases. Other special cases have been considered by Fong (1992), Hill (1965), Reinsel (1985), and Reinsel and Tiao (1987). Here we demonstrate how to draw inferences under the foregoing very general formulation.

We will use the Laplacian T approximation (Leonard et al. 1994). Exact results are calculated using importance sampling (Geweke 1988, 1989; Leonard and Hsu 1992; Zellner and Rossi 1984), which is a simple restatement of straightforward Monte Carlo and involves independent simulations. It usually will be unnecessary to refer to the dependent simulations of the Gibbs sampler (Gelfand and Smith 1990; George, Makov, and Smith 1994), and indeed this technique should be reserved for more complicated hierarchical Bayes models; for example, with nonnormal error terms. Although it would be easy to describe a sequence of conditional distributions that would facilitate the Gibbs sampler in the current context, fast convergence of the importance sampling is more easily assured, for example, via the central limit theorem utilized by Geweke.

3. THE LAPLACIAN *T*-APPROXIMATION

Let $\pi_y(\boldsymbol{\theta})$ be the posterior density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ given data $\mathbf{y}$, and let $\eta = g(\boldsymbol{\theta})$ be the parameter of interest. Leonard et al. (1994) introduced a Laplacian T-approximation for the marginal posterior of η of the form

$$\pi_y^*(\eta) \propto |\mathbf{T}_\eta|^{-1/2} \pi_y(\boldsymbol{\theta}_\eta) \lambda_\eta^{-\omega/2} f(\eta|\omega, \theta_\eta^*, \mathbf{T}_\eta), \tag{2}$$

to the marginal posterior density of η, where

$$\mathbf{T}_\eta = \frac{\omega}{(\omega + p)\lambda_\eta} \mathbf{Q}_\eta,$$

$$\lambda_\eta = 1 - (\omega + p - 1)^{-1} \mathbf{l}_\eta^T \mathbf{Q}_\eta^{-1} \mathbf{l}_\eta,$$

$$\mathbf{Q}_\eta = \mathbf{U}_\eta + \frac{2}{\omega + p - 1} \mathbf{l}_\eta \mathbf{l}_\eta^T,$$

$$\mathbf{l}_\eta = \left. \frac{\partial \log \pi_y(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_\eta}, \tag{3}$$

$$\mathbf{U}_\eta = \left. \frac{\partial^2 \log \pi_y(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta}\boldsymbol{\theta}^T)} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_\eta}, \tag{4}$$

$$\boldsymbol{\theta}_\eta^* = \boldsymbol{\theta}_\eta + \mathbf{Q}_\eta^{-1} \mathbf{l}_\eta,$$

and $f(\eta|\omega, \theta_\eta^*, T_\eta)$ denotes the density of $\eta = g(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ possesses a multivariate t distribution with ω degrees of freedom, mean vector $\boldsymbol{\theta}_\eta^*$, and precision matrix $\mathbf{T}_\eta$. Here $\boldsymbol{\theta}_\eta$ represents some convenient approximation to the conditional posterior mean vector of $\boldsymbol{\theta}$, given η, and ω should be taken to roughly approximate the degrees of freedom of a generalized multivariate t approximation to the conditional distribution of $\boldsymbol{\theta}$ given η.

When $\boldsymbol{\theta}_\eta$ is the conditional posterior mode vector of $\boldsymbol{\theta}$, given η, (2) reduces to the Laplacian approximation introduced by Leonard (1982) and shown by Tierney and Kadane (1986) and Leonard et al. (1989) to possess saddlepoint accuracy and excellent finite-sample accuracy, in many special cases. It was previously used for hierarchical models by Kass and Steffey (1989).

But whenever the conditional posterior mode vector of $\boldsymbol{\theta}$, given η, does not closely approximate the conditional posterior mean vector, the alternative expression (2) can be superior to the original Laplacian approximation, with a judicious choice of ω. (See, e.g., a multiparameter Fisher–Behrens model and the numerical example presented in Leonard et al. 1994.)

In the special case where $\eta = \mathbf{a}^T\boldsymbol{\theta}$ is a linear combination of the θ's, the approximation (2) is equivalent to

$$\pi_y^*(\eta) \propto |\mathbf{T}_\eta|^{-1/2}\pi_y(\boldsymbol{\theta}_\eta)\lambda_\eta^{-(\omega+p)/2}t_\eta[\omega, \mathbf{a}^T\boldsymbol{\theta}_\eta^*, (\mathbf{a}^T\mathbf{T}_\eta^{-1}\mathbf{a})^{-1}]. \tag{5}$$

Here $t_\eta(\omega, \mu, \tau)$ denotes a generalized t density for η, with mean μ, precision τ, and ω degrees of freedom.

4. GENERAL BAYESIAN ANALYSIS

The model description in Section 2 implies assumptions for $\boldsymbol{\theta}$ of the following form. Let $\mathbf{A}\boldsymbol{\theta}$ possess, a priori and given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, a multivariate normal distribution with mean vector $\mathbf{G}\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In the prior assessment, we suppose that $\boldsymbol{\mu}$ is uniformly distributed over s-dimensional Euclidean space. Here $\mathbf{A}$ is the appropriate $(p - q_0) \times p$ matrix such that $\mathbf{A}\boldsymbol{\theta}$ consists of the last $p - q_0$ elements of $\boldsymbol{\theta}$ and the first q_0 elements of $\boldsymbol{\theta}$ (corresponding to $\boldsymbol{\alpha}$) are uniformly distributed. Furthermore, $\mathbf{G}$ is the appropriate $(p - q_0) \times s$ matrix, with a single unit entry in each row and zeros elsewhere. Finally, $\boldsymbol{\Sigma}$ is the appropriate $(p - q_0) \times (p - q_0)$ diagonal matrix, depending on $\sigma_1^2, \sigma_2^2, \ldots, \sigma_s^2$.

Consequently, the joint prior density of $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$, given $\boldsymbol{\Sigma}$, is

$$\pi(\boldsymbol{\mu}, \boldsymbol{\theta}|\boldsymbol{\Sigma}) \propto \exp\left[-\frac{1}{2}(\mathbf{A}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{A}\boldsymbol{\theta} - \mathbf{G}\boldsymbol{\mu})\right].$$

Integrating out $\boldsymbol{\mu}$, we find that the prior density of $\boldsymbol{\theta}$, given $\boldsymbol{\Sigma}$, is

$$\pi(\boldsymbol{\theta}|\boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{R}\mathbf{A}\boldsymbol{\theta}\right\},$$

where

$$\mathbf{R} = \boldsymbol{\Sigma}^{-1}[\mathbf{I}_{p-q_0} - \mathbf{G}(\mathbf{G}^T\boldsymbol{\Sigma}^{-1}\mathbf{G})^{-1}\mathbf{G}^T]\boldsymbol{\Sigma}^{-1}. \tag{6}$$

Consequently, the conditional posterior distribution of $\boldsymbol{\theta}$, given ϕ and $\boldsymbol{\Sigma}$, is multivariate normal, with mean vector

$$\boldsymbol{\theta}^* = \phi^{-1}(\phi^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{X}^T\mathbf{y} \tag{7}$$

and covariance matrix

$$\mathbf{D} = (\phi^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}, \tag{8}$$

where it is assumed that the inverse in (8) exists. If this is not the case, then our analysis can be readily extended, using generalized inverses. But this condition is, for example, satisfied by Sun's formulation of two-way random effects models, as discussed in Section 2. Furthermore, the posterior density of ϕ, and $\sigma_1^2, \ldots, \sigma_s^2$, is

$$\pi(\phi, \sigma_1^2, \ldots, \sigma_s^2|\mathbf{y}) \propto \pi(\phi, \sigma_1^2, \ldots, \sigma_s^2)l^*(\phi, \sigma_1^2, \ldots, \sigma_s^2|\mathbf{y}), \tag{9}$$

where the first contribution to the right side denotes the prior density,

$$\begin{aligned} l^*(\phi, \sigma_1^2, \ldots, \sigma_s^2|\mathbf{y}) &\propto \phi^{-1/2}|\boldsymbol{\Sigma}|^{-1/2} \\ &\times |\mathbf{G}^T\boldsymbol{\Sigma}^{-1/2}\mathbf{G}|^{-1/2}|\phi^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{A}^T\mathbf{R}\mathbf{A}|^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}\phi^{-1}S_R^2 - \frac{1}{2}\phi^{-1}\hat{\boldsymbol{\theta}}^T\mathbf{A}^T\mathbf{R}\mathbf{A}\right. \\ &\qquad \left.(\phi^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\theta}}\right\}, \end{aligned}$$

and $\hat{\boldsymbol{\theta}}$ and S_R^2 denote the least squares vector of $\boldsymbol{\theta}$ and the residual sum of squares.

Note that, the posterior density of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \gamma_1, \ldots, \gamma_s)^T$, unconditional on $\boldsymbol{\mu}$ and the variances, is the product of generalized multivariate t-densities

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}) &\propto [\nu_0\zeta_0 + S_R^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]^{-(\nu_0+M)/2} \\ &\times \prod_{j=1}^{s}[\nu_j\zeta_j + \boldsymbol{\gamma}_j^T(I_{q_j} - q_j^{-1}\mathbf{e}_{q_j}\mathbf{e}_{q_j}^T)\boldsymbol{\gamma}_j]^{-(q_j+\nu_j-1)/2}, \end{aligned} \tag{10}$$

where the ν's and ζ's were introduced in Section 2. The question now arises as to how to approximate the unconditional posterior mean of elements of $\boldsymbol{\theta}$, or linear combinations thereof. Lindley (1971) and Leonard (1972), in special cases, recommended maximizing (10) to find the unconditional posterior mode vector of $\boldsymbol{\theta}$. Lindley and Smith (1972) recommended maximizing the joint posterior mode vector of $\boldsymbol{\theta}$ and the variance components. Efron and Morris (1973) were the first to state that this can provide estimators with quite inferior frequency properties. A related problem is that the conditional posterior mode vector of $\boldsymbol{\theta}$, given $\mathbf{a}^T\boldsymbol{\theta} = \eta$, is unlikely to closely approximate the conditional posterior mean vector of $\boldsymbol{\theta}$, given η.

Following O'Hagan (1976), a more promising suggestion is to replace ϕ and $\sigma_1^2, \ldots, \sigma_s^2$ in (8) and (9) by their (marginal) posterior modes; that is, the values maximizing the posterior density (10). We refer to O'Hagan's suggestion as $\tilde{\boldsymbol{\theta}}$. This is an estimator of the formal "Bayes–Stein" type

and motivates us to suggest the following general methods of analysis:

- **Laplacian T approximations.** Apply the general approximation (5) to the posterior density of any linear transformation $\eta = \mathbf{a}^T\boldsymbol{\theta}$ of the parameter vector $\boldsymbol{\theta}$. Note that the posterior density of $\boldsymbol{\theta}$ conditional on the variance components is maximized, subject to the constraint $\mathbf{a}^T\boldsymbol{\theta} = \eta$ when $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_\eta$, where

$$\tilde{\boldsymbol{\theta}}_\eta = \boldsymbol{\theta}^* + \kappa\mathbf{D}\mathbf{a}, \tag{11}$$

with

$$\kappa = (\mathbf{a}^T\mathbf{D}\mathbf{a})^{-1}(\eta - \mathbf{a}^T\boldsymbol{\theta}^*), \tag{12}$$

and $\boldsymbol{\theta}^*$ and $\mathbf{D}$ are defined in (7) and (8). We thus recommend initially approximating the conditional mean vector of $\boldsymbol{\theta}$, given only η, by the $\tilde{\boldsymbol{\theta}}_\eta$, which replaces $\boldsymbol{\theta}^*$ in (11) and (12) by the O'Hagan estimator $\tilde{\boldsymbol{\theta}}$. Furthermore, replace all variance components appearing in $\mathbf{D}$ [see (6) and (8)] by corresponding values maximizing (9). The approximation (5) will then be different from the ordinary Laplacian approximation introduced by Leonard (1982). All of the details of this approximation, as described in Section 3, follow in straightforward fashion; for example, (3) and (4) relate to the first derivatives and second derivatives, with respect to $\boldsymbol{\theta}$, of the logarithm of the joint posterior density (10). Note that the posterior means of the elements of $\boldsymbol{\theta}$ will no longer assume weighted average forms of the Bayes–Stein type.

  Leonard et al. (1994) made two suggestions for choosing ω, and one of these is applied in Section 9. This approximate procedure consumes much less computer time as compared to exact procedures based on importance sampling and was shown by Leonard et al. to be remarkably accurate in the context of a multiparameter Fisher–Behrens problem.
- **Exact procedures (importance sampling).** We recommend importance sampling from a generalized multivariate t approximation to the posterior density of $\log\phi, \log\sigma_1^2, \ldots, \log\sigma_s^2$. The exact density can be obtained from (9). Let $\boldsymbol{\xi}$ and $\boldsymbol{\Lambda}$ denote the *posterior mode vector* and *posterior information matrix* of these $s + 1$ parameters. Then the procedures described by Leonard and Hsu (1992) may be used, by simulating from a generalized multivariate t distribution with ω^* degrees of freedom, mean vector $\boldsymbol{\xi}$, and precision matrix $\omega^*\boldsymbol{\Lambda}/(\omega^* + s + 1)$. The degrees of freedom ω^* should be chosen pragmatically, to ensure fast and steady convergence, or adaptively, to reduce the standard error of simulation. Choosing ω^* equal to 20 or 30 often suffices. This permits exact computation of the posterior expectation or distribution of any parameter of interest; for example, $\eta = \mathbf{a}^T\boldsymbol{\theta}$. In some special cases, variations on this scheme are available by taking logs of linear combinations of the variance components and then using rejection methods to handle the constraints on the new parameters.

5. THE COLLAPSING PHENOMENON

Consider the special case of the one-way random-effects model of Section 2 where ϕ is known, $y_{i1} = y_i$, for $i = 1, \ldots, m$, and $n = 1$. Let μ and σ^2 be a priori independent with μ uniformly distributed over $(-\infty, \infty)$ and $\nu\zeta/\sigma^2$ possessing a chi-squared distribution with ν degrees of freedom. Then estimates for the θ_i might be considered that take the form

$$\tilde{\theta}_i = (\phi^{-1}y_i + \tilde{\sigma}^{-2}y_\cdot)/(\phi^{-1} + \tilde{\sigma}^{-2}), \tag{13}$$

where

$$\tilde{\sigma}^2 = \left[\nu\zeta + \sum_{i=1}^{m}(\tilde{\theta}_i - y_\cdot)^2\right] \Big/ (\nu^* + m), \tag{14}$$

with the dot notation denoting average with respect to that subscript and possible choices of ν^* are discussed later (see, for example, Scott and Smith 1969).

For example, $\nu^* = \nu - 1$ leads to estimates maximizing the joint posterior density of $\theta_1, \ldots, \theta_m$, and $\nu^* = \nu + 2$ corresponds to maximizing the joint posterior density of $\theta_1, \ldots, \theta_m, \mu$, and σ^2, with respect to all $m + 2$ parameters. Note that (13) gives the posterior mean of θ_i, given only σ^2, but with σ^2 replaced by $\tilde{\sigma}^2$.

Any estimator of the form

$$\hat{\sigma}^2 = \max\{(m - c)^{-1}S^2 - \phi, 0\}, \tag{15}$$

with $S^2 = \sum_{i=1}^{m}(y_i - \bar{y})^2$, will be consistent for σ^2, as $m \to \infty$, with c fixed, where this consistency refers to the joint distribution of $y_1, y_2, \ldots$, conditional only on μ and σ^2. Replacing $\tilde{\sigma}^2$ in (13) by $\hat{\sigma}^2$ leads to the empirical Bayes estimators

$$\theta_i^* = \begin{cases} y_i - \frac{(m-c)}{S^2}\phi(y_i - y_\cdot) & \text{if } S^2 \geq (m - c)\phi, \\ \bar{y} & \text{if } S^2 \leq (m - c)\phi. \end{cases} \tag{16}$$

Note that empirical Bayes estimators replace prior parameters by estimators based on the current data set, which are themselves free from further prior assumptions. They contrast with hierarchical Bayes estimators, which place further distributions on the (first-stage) prior parameters.

The estimator in (16) was first recommended by Lindley (1962) as a refinement to Stein's (1956) estimators. The iterative procedures resulting from (13) and (14) may be compared to techniques based on the EM algorithm (Dempster, Laird, and Rubin 1977) for the marginal posterior modes μ_M and σ_M^2 of μ and σ^2. The EM algorithm yields the equations $\mu_M = y$ and

$$\sigma_M^2 = \left\{\nu\zeta + \sum_{i=1}^{m}(\tilde{\theta}_i^M - y_\cdot)^2 + m\tilde{v}_M\right\} \Big/ (\nu_M^* + m), \tag{17}$$

where $\nu_M^* = \nu + 2$, $\tilde{\theta}_i^M$ is found by replacing $\tilde{\sigma}^2$ in (13) by σ_M^2, and

$$\tilde{v}_M = (\phi^{-1} + \sigma_M^2)^{-1}. \tag{18}$$

Note that (17) adjusts (14) via a term $\tilde{v}_M$, which is based on the conditional posterior variance of each θ_i, given μ and σ^2. Combining (13), (17), and (18) shows that σ_M^2 satisfies the cubic equation

$$(\nu_M^* + m)\sigma_M^2 = \nu\zeta + \sigma_M^4 S^2/(\sigma_M^2 + \phi)^2 + m\phi\sigma_M^2/(\sigma_M^2 + \phi). \quad (19)$$

But combining (13) and (14) shows that $\tilde{\sigma}^2$ satisfies

$$(\nu_M^* + m)\tilde{\sigma}^2 = \nu\zeta + \tilde{\sigma}^4 S^2/(\tilde{\sigma}^2 + \phi)^2. \quad (20)$$

Equations (19) and (20) can yield quite different numerical solutions. For example, σ_M^2 satisfying (19) has the same limiting behavior, as $m \to \infty$, as $\hat{\sigma}^2$ in (15), which justifies the Lindley–Stein estimator (16). But as $m \to \infty$, (20) can be reduced to either $\tilde{\sigma}^2 = 0$ or a quadratic in $\tilde{\sigma}^2$. By considering the solution of this quadratic, we find that $\tilde{\sigma}^2$ has the same limiting behavior, as $m \to \infty$, as

$$\tilde{\sigma}_\infty^2 = \begin{cases} \frac{1}{2}m^{-1}S^2 - \phi + \sqrt{\frac{1}{4}m^{-2}S^4 - m^{-1}S^2\phi} & \\ \qquad\qquad \text{if } S^2 > 4m\phi, & (21) \\ 0 \quad \text{if } S^2 < 4m\phi. & \end{cases}$$

Note that for finite m, the expression in (21) provides the exact solution to (14), when $\nu = -2, \zeta = 0$ and $\nu^* = \nu + 2$. Thus it is a joint modal estimator under a uniform prior for σ^2.

Under our distributional assumptions for $y_1, \ldots, y_m$, conditional on μ and σ^2 but not on $\theta_1, \ldots, \theta_m$, the estimators $\hat{\sigma}^2$ and σ_M^2 are strongly consistent for σ^2, as $m \to \infty$, with r, ν, and ζ fixed. But $\tilde{\sigma}^2$ and $\tilde{\sigma}_\infty^2$ are strongly consistent for σ_0^2, as $m \to \infty$, with ν and ζ fixed where

$$\sigma_0^2 = \begin{cases} \frac{1}{2}(\sigma^2 - \phi) + \frac{1}{2}\sqrt{(\sigma^2 + \phi)(\sigma^2 - 3\phi)} & \text{for } \sigma^2 \geq 3\phi, \\ 0 \quad \text{for } \sigma^2 < 3\phi. & \end{cases}$$

Because σ_0^2 is always less than σ^2, this defines the *collapsing phenomenon.* It is quite striking that when $\sigma^2 < 3\phi, \tilde{\sigma}^2$ is strongly consistent for zero.

We emphasize these results by the following summary: (a) $\tilde{\sigma}^2 \xrightarrow{d} \sigma_0^2 < \sigma^2\ (m \to \infty)$, and (b) $\sigma_M^2 \xrightarrow{d} \sigma^2\ (m \to \infty)$, where the convergence in distribution incorporates prior assumptions regarding $\theta_1, \ldots, \theta_m$.

6. MEAN SQUARED ERROR PROPERTIES

We now investigate the average mean squared errors (MSE's), conditional on $\theta_1, \ldots, \theta_m$, of the estimators $\tilde{\theta}_1, \ldots, \tilde{\theta}_m$. We regard frequency properties as part of the Bayesian paradigm. Bayesian concepts can be used to construct meaningful procedures. But the latter can then be further validated by considering their potential long-run performance. One problem with MSE is that it averages the components of loss. In our computations of Section 10, we also refer to the risk function under the loss function

$$L(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \max_{j=1,\ldots,m} (\tilde{\theta}_j - \theta_j)^2. \quad (22)$$

Sun (1992) proved the following lemma.

Lemma 1. Whenever

$$Q = Q(\tilde{\boldsymbol{\theta}}) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} (\tilde{\theta}_i - \theta_i)^2,$$

exists, the average MSE of $\tilde{\theta}_1, \tilde{\theta}_2, \ldots,$ approaches

$$r^2(\boldsymbol{\theta}) = \begin{cases} \phi\left[\frac{(Q-\phi)(Q+\phi+R)}{(Q+\phi)(Q-\phi+R)}\right] & \text{if } Q \geq 3\phi, \\ Q \quad \text{if } Q < 3\phi, & \end{cases}$$

where

$$R = \sqrt{(Q+\phi)(Q-3\phi)}.$$

Because the maximum likelihood estimators $y_1, \ldots, y_m$ have average risk ϕ, this demonstrates that the posterior modes $\tilde{\theta}_1, \ldots, \tilde{\theta}_m$ have smaller limiting average MSE than $y_1, \ldots, y_m$ only if $Q > 3\phi$ or $Q < \phi$. But the Lindley–Stein estimators (16), together with O'Hagan estimators based on μ_M and σ_M^2, have a limiting-average MSE

$$r_M(\boldsymbol{\theta}) = \frac{Q\phi}{(Q+\phi)}$$

and thus are substantially superior to $y_1, \ldots, y_m$ whenever at least two of the θ_i's are unequal.

7. SAMPLING VARIANCE ϕ UNKNOWN

Consider next the one-way random-effects model, with replications, described in Section 2. Together with the prior assumptions for μ and σ^2 of Section 4, assume that $\nu_0\zeta_0/\phi$ has a chi-squared distribution with ν_0 degrees of freedom independently of μ and σ^2. Let S_{W}^2 and S_{B}^2 denote the usual within-group and between-group sum of squares. It is straightforward to show that as $m \to \infty$, with n fixed, the Lindley–Smith joint modal estimators possess the same limiting behavior as the estimators

$$\tilde{\theta}_i = (1 - \tilde{\rho})y_{i\cdot} + \tilde{\rho}y_{\cdot\cdot}, \quad (23)$$

where

$$\tilde{\rho} = \begin{cases} 1 - \{2(1+n^{-1})\}^{-1}\sqrt{1 - 4(1-n^{-2})F^{-1}} & \\ \qquad\qquad \text{if } F \geq 4(1-n^{-2}), & \\ 1 \quad \text{if } F < 4(1-n^{-2}), & \end{cases}$$

and F is the usual F statistic. The estimators in (23) are also the joint posterior modes of $\theta_1, \ldots, \theta_m$, unconditional on μ and σ^2, for finite m, and under uniform priors for ϕ and σ^2; that is, $\pi(\phi, \sigma^2) \propto 1$. But estimators based on marginal modes for ϕ and σ^2 (see Sec. 2) replace $\tilde{\rho}$ by

$$\rho^* = \min\{F^{-1}, 1\}.$$

Under our distributional assumptions, given μ, σ^2, and ϕ, the quantities

$$(n-1)M_{\mathrm{W}} = m^{-1}S_{\mathrm{W}}^2,$$

$$M_{\mathrm{B}} = (m-1)^{-1}S_{\mathrm{B}}^2,$$

and F are strongly consistent, as $m \to \infty$, with n fixed for $(n-1)\phi, n\sigma^2+\phi$, and $1+n\sigma^2/\phi$. The joint modal estimators for ϕ and σ^2 have the same limiting form as the estimators

$$\tilde{\phi} = n^{-1}[(n-1)M_{\mathrm{W}} + \tilde{\rho}^2 M_{\mathrm{B}}]$$

and

$$\tilde{\sigma}^2 = n^{-1}(1-\tilde{\rho})^2 M_{\mathrm{B}},$$

and these are not always strongly consistent for ϕ and σ^2. Indeed, if $n\sigma^2 < \phi(3-4n^{-2})$, then $\tilde{\sigma}^2$ is strongly consistent for zero, so that the collapsing phenomenon again occurs.

8. BAYESIAN ANALYSIS FOR THE ONE-WAY RANDOM-EFFECTS MODEL

For the model described in Section 2, the posterior distribution of θ_i conditional on ϕ and σ^2, is normal with mean

$$\theta_i^* = \frac{n\phi^{-1}y_{i\cdot} + \sigma^{-2}y_{\cdot\cdot}}{n\phi^{-1}+\sigma^{-2}}, \tag{24}$$

and variance $(n\phi^{-1}+\sigma^{-2})^{-1}(1+m^{-1}n^{-1}\phi/\sigma^2)$. O'Hagan's marginal modal estimate $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_m)^T$, would replace ϕ and σ^2 in (8.1), by the values maximizing their joint posterior density

$$\pi(\phi, \sigma^2|\mathbf{y}) \propto \pi(\phi,\sigma^2)\phi^{-\nu_{\mathrm{W}}/2}(\phi+n\sigma^2)^{-\nu_{\mathrm{B}}/2} \times \exp\left\{-\frac{1}{2}\phi^{-1}S_{\mathrm{B}}^2 - \frac{1}{2}(\phi+n\sigma^2)^{-1}S_{\mathrm{B}}^2\right\}, \tag{25}$$

with $\nu_{\mathrm{W}} = m(n-1), \nu_{\mathrm{B}} = m-1$, and $\pi(\phi,\sigma^2)$ denoting the corresponding prior density. When $\pi(\phi,\sigma^2) \propto 1$, (25) is maximized when $\phi = M_{\mathrm{W}}$ and $\sigma^2 = \max\{n^{-1}(M_{\mathrm{B}} - M_{\mathrm{W}}), 0\}$. Hence O'Hagan's marginal mode suggestion in this case would provide the Lindley–Stein estimators

$$\tilde{\theta}_i = \begin{cases} y_{i\cdot} - F^{-1}(y_{i\cdot} - y_{\cdot\cdot}) & \text{if } F \geq 1, \\ y_{\cdot\cdot} & \text{if } F < 1, \end{cases} \tag{26}$$

where F is the usual F statistic. Under the prior assumption for μ, σ^2, and ϕ, of Sections 5 and 7, the joint posterior density of $\theta_1, \ldots, \theta_m$ is

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \left[\nu_0\zeta_0 + S_{\mathrm{W}}^2 + n\sum_{i=1}^m (\theta_i - y_{\cdot})^2\right]^{-(\nu_0+mn)/2} \times \left[\nu\zeta + \sum_{i=1}^m (\theta_i - \theta_{\cdot})^2\right]^{-(\nu+m-1)/2}. \tag{27}$$

The Laplacian T procedure of Sections 3 and 4 may be implemented for linear transformation $\eta = \mathbf{a}^T\boldsymbol{\theta}$ by replacing $\boldsymbol{\theta}^*$ in (11) and (12) by the O'Hagan vector $\tilde{\boldsymbol{\theta}}$. Furthermore, $\mathbf{D}$ in (11) and (12) should replace ϕ and σ^2 in the matrix $\mathbf{D}$, satisfying

$$\mathbf{D}^{-1} = n\phi^{-1}\mathbf{I}_m + \sigma^{-2}[\mathbf{I}_m - m^{-1}\mathbf{e}_m\mathbf{e}_m^T],$$

by the marginal modes described previously where $\mathbf{e}_m$ denotes the $m \times 1$ unit vector. Then (11) and (12) provide an approximate vector $\boldsymbol{\theta}_\eta$ satisfying $\eta = \mathbf{a}^T\boldsymbol{\theta}_n$.

The approximation (5) may not be directly applied. Note that the derivatives in (3) and (4) refer to the logarithm of the posterior density (27), and that when $\pi(\phi,\sigma^2) \propto 1$, the choices $\nu_0 = \nu = -2$ and $\zeta_0 = \zeta = 0$ should be applied to (27). The density in (27) will still remain proper, unless m is too small. Note also that the mode $\tilde{\eta}$ of the Laplacian approximation generally will not satisfy $\tilde{\eta} = \mathbf{a}^T\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ denotes the O'Hagan estimate, but instead will seek to approximate the mode of the marginal posterior density of η.

As a modification to the importance sampling procedure of Section 4, when $\pi(\phi,\sigma^2) \propto 1$, note that the posterior density of $\gamma_1 = \log\phi$ and $\gamma_2 = \log(\phi+n\sigma^2)$ is

$$\pi(\gamma_1,\gamma_2|\mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\nu_{\mathrm{W}}-2)\gamma_1 - \frac{1}{2}e^{-\gamma_1}S_{\mathrm{W}}^2\right\} \times \exp\left\{-\frac{1}{2}(\nu_{\mathrm{B}}-2)\gamma_2 - \frac{1}{2}e^{-\gamma_2}S_{\mathrm{B}}^2\right\}, \quad (-\infty < \gamma_1 \leq \gamma_2 < \infty).$$

This motivates a generalized bivariate t importance function, combined with a rejection of those simulated values for which $\gamma_1 > \gamma_2$. The importance function should possess, say, $\omega^* = 30$ degrees of freedom, together with mean vector $(\log[S_{\mathrm{W}}^2/(\nu_{\mathrm{W}}-2)], \log[S_{\mathrm{B}}^2/(\nu_{\mathrm{B}}-2)])$ and diagonal precision matrix with diagonal elements equal to $\omega^*(\nu_{\mathrm{W}}-2)/2(\omega^*+2)$ and $\omega^*(\nu_{\mathrm{B}}-2)/2(\omega^*+2)$. Similar constructions are available for the more complicated experimental designs reported by Sun (1992). This importance sampling procedure may be used to compute the exact unconditional posterior distribution of each θ_i, based on the normal distribution, conditional on ϕ and σ^2, described earlier.

9. NUMERICAL EXAMPLE

Consider the data reported by Hald (1952, p. 434) relating to the tensile strengths of $n = 12$ wires contained in each of $m = 9$ cables. The group means for the $m = 9$ cables are described in the first column of Table 1. (Hald has subtracted 340 from each observation.)

We applied the importance sampling/rejection method of Section 8 to calculate the posterior distributions of each group mean under uniform priors for ϕ and σ_1^2. The exact posterior density of the first group mean is described by the solid curve in Figure 1. We then applied the Laplacian T approximation of Section 3. The degrees of freedom ω appearing in the approximation were chosen by the matching method (i.e., comparison with the marginal based on a bivariate t approximation to the posterior distribution of $\log\phi$ and $\log\sigma^2$) described by Leonard et al. (1994). Any value of ω between 80 and 100 performed virtually equally well. The Laplacian approximation to the posterior density of the first group mean is reported by the dotted curve in Figure 1, with $\omega = 100$.

Note the excellent correspondence between the (algebraically explicit) Laplacian approximation and the exact

Table 1. Estimates of Group Means (Hald's Tensile Strength Data)

Group	Group mean	Posterior mean	(Standard deviation)	Laplacian mode	Lindley–Stein	Joint mode
1	−4.083	−3.796	(1.463)	−3.791	−3.714	−3.696
2	−7.000	−6.463	(1.488)	−6.461	−6.395	−6.275
3	−6.083	−5.625	(1.479)	−5.621	−5.492	−5.465
4	−2.667	−2.501	(1.456)	−2.497	−2.453	−2.442
5	1.917	1.690	(1.459)	−1.685	1.625	1.611
6	.833	.699	(1.455)	.696	.661	.652
7	.917	.776	(1.456)	.772	.735	.726
8	3.333	2.985	(1.468)	2.980	2.885	2.863
9	6.250	5.652	(1.497)	5.653	5.480	5.443

results. Results of similar accuracy were obtained for all $m = 9$ groups.

The observed group means are described by the entries in the first column of Table 1. The entries in the second column describe the exact posterior means of the treatment means, under our uniform prior for ϕ and σ^2. The corresponding posterior standard deviations are in brackets. The third column lists the unconditional posterior modes of the treatment means, under our Laplacian approximations to their marginal posterior densities. The fourth column presents the Lindley–Stein (O'Hagan) estimates (25). Finally, the fifth column lists the (no longer recommended) joint modal estimates (23).

Note that the posterior means shrink the corresponding sample means toward the grand mean, $y_{\cdot\cdot} = -.73$. Furthermore, both the Lindley–Stein (O'Hagan) estimates and the Laplacian modes are very close in numerical terms to the posterior means. But the joint modal estimates in the fifth column shrink the group means slightly further toward the grand mean.

The foregoing calculations were performed on a Sun 4/330 computer. About 1 second of CPU time was needed to calculate each Laplacian curve. In contrast, 1,333.4 seconds were needed to calculate the corresponding curve by importance sampling, using a half-million simulations to obtain accuracy to about three significant digits. But the $m = 9$ posterior means and variances were calculated in just 336.6 seconds, with a similar number of simulations yielding about four significant digit accuracy. The random number generator (RNOR) used was produced by the Computer Science Department of Washington State University.

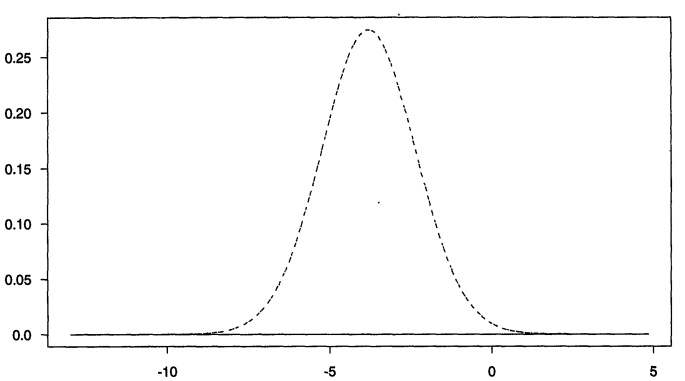

Figure 1. Posterior Density of First Group Mean. The dotted curve represents Laplacian T approximation ($\omega = 100$); the solid curve, exact (by importance sampling).

10. FREQUENCY SIMULATIONS

We simulated the risks of the various choices of estimators for $\theta_1, \ldots, \theta_m$ under five loss functions: average squared error loss, average absolute loss, maximum squared error loss, maximum absolute loss, average step loss, with unit step width. A large range of choices of m, n, ϕ, and $\theta_1, \ldots, \theta_m$ yielded qualitatively similar results to those described in Tables 2, 3, and 4, which take $m = 10$ and $n = 20$. In Table 2, it is supposed that $\theta_1 = 2, \theta_2 = 4, \ldots, \theta_{10} = 20$, whereas in Table 3, it is instead assumed that $\theta_1 = 1$, $\theta_2 = 2, \ldots, \theta_9 = 9$, and $\theta_{10} = 20$. Finally, in Table 4, the choices $\theta_1 = 1, \theta_2 = 6, \theta_3 = 7, \ldots, \theta_8 = 12, \theta_9 = 13$, and $\theta_{10} = 23$ are made.

The risks for the sample means, Laplacian modes (under uniform priors for ϕ and σ^2), Lindley–Stein estimators, and joint posterior modes of $\theta_1, \ldots, \theta_m$ (also under uniform priors for ϕ and σ^2) are described in the last four columns of Table 2 for the five choices of loss function. The joint modes are described in (23); these maximize the joint posterior density of $\theta_1, \ldots, \theta_m$, unconditional on θ and σ^2. Note the following results:

1. The Laplacian modes and Lindley–Stein estimators perform better than the sample means under all five loss functions.
2. There is little to choose from in terms of frequency performance between the Laplacian modes and the Lindley–Stein estimators. But it is anticipated that the exact posterior means might perform slightly better than the Lindley–Stein estimators, if their risks could indeed be simulated.
3. Though the joint modal estimators perform almost as well as Lindley–Stein for some parameter values, they perform worse than even the sample means for other parameter values (e.g., note the results for $\phi = 120$).

More substantial differences are noticeable in Tables 3 and 4. Note, for example, the inferior performance of the joint modes (under uniform priors for ϕ and σ^2) when $\phi = 120$. Sun (1992) further investigated frequency properties of the joint modes, but under proper inverted chi-squared priors for ϕ and σ^2. He showed that for some choices of the prior parameters appearing in the inverted chi-squared distributions, the MSE properties of the joint modes can be greatly improved, in comparison to our results under uniform priors for ϕ and σ^2. But for a very wide range of choices of the prior parameters, the joint modes do not perform so well.

Table 2. Risks Under Alternative Loss Functions (First Set of Treatment Means)

Loss function	Sample means	Laplacian modes	Lindley–Stein	Joint modes
		$\phi = 40$		
(a)	1.983	1.927	1.932	1.935
(b)	1.122	1.106	1.108	1.109
(c)	7.558	7.368	7.385	7.396
(d)	2.652	2.619	2.622	2.624
(e)	.722	.718	.719	.719
		$\phi = 80$		
(a)	3.966	3.757	3.768	3.820
(b)	1.587	1.545	1.548	1.559
(c)	15.117	14.337	14.354	14.500
(d)	3.750	3.655	3.658	3.677
(e)	.801	.796	.798	.799
		$\phi = 120$		
(a)	5.946	5.541	.511	6.318
(b)	1.944	1.877	1.874	1.976
(c)	22.668	21.022	20.839	22.828
(d)	4.592	4.426	4.414	4.565
(e)	.837	.833	.830	.835

The simulations presented in this article were performed using the random number generator described in Section 9. The entries in Tables 2, 3, and 4 are each based on 5,000 simulations, yielding good accuracy to about three significant digits.

11. FURTHER DISCUSSION

Note that for the one-way random-effects model of Section 8, as $m \to \infty$, with $y_{1\cdot}, y_{2\cdot}, \ldots$ fixed and convergent to a limit $y_{\cdot\cdot}$ and M_W, M_B, and n fixed, $\theta_1, \theta_2, \ldots$ converge (in posterior distribution and irrespective of our prior assumptions for ϕ and σ^2) to independent normal variates. The limiting posterior means are simply the Lindley–Stein estimates in (26), and the limiting common posterior variances is $n^{-1}F^{-1}\max(M_B - M_W, 0)$.

The simulation results of Section 10 suggest that Stein-type shrinkage estimators may have good risk properties under the loss function (22), as well as under average squared error loss. This is desirable, because it seems important to control all individual losses, as well as their average. A Bayesian property is now developed.

Lemma 2. Suppose that $\theta_1, \ldots, \theta_m$ are a posteriori independent and normally distributed with means $\theta_1^*, \ldots, \theta_m^*$ and variances $v_1, \ldots, v_m$. Then $\theta_1^*, \ldots, \theta_m^*$ are Bayes (or generalized Bayes) estimates under any loss function of the form

$$L(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \max_{j=1,\ldots,m} \Lambda\{|\tilde{\theta}_j - \theta_j|\}, \tag{28}$$

where $\Lambda\{\cdot\}$ is a strictly increasing, with $\Lambda\{0\} = 0$.

Table 3. Risks Under Alternative Loss Functions (Second Set of Treatment Means)

Loss function	Sample means	Laplacian modes	Lindley–Stein	Joint modes
		$\phi = 40$		
(a)	1.983	1.912	1.920	1.927
(b)	1.122	1.101	1.102	1.103
(c)	7.559	7.349	7.426	7.467
(d)	2.652	2.615	2.628	2.635
(e)	.722	.715	.716	.716
		$\phi = 80$		
(a)	3.966	3.683	3.723	3.886
(b)	1.587	1.524	1.527	1.545
(c)	15.117	14.368	14.759	16.009
(d)	3.750	3.653	3.697	3.803
(e)	.801	.791	.791	.791
		$\phi = 120$		
(a)	5.942	5.310	5.367	7.774
(b)	1.943	1.826	1.828	2.023
(c)	22.588	20.914	21.681	39.764
(d)	4.584	4.403	4.480	5.500
(e)	.837	.824	.822	.830

Table 4. Risks Under Alternative Loss Functions (Third Set of Treatment Means)

Loss function	Sample means	Laplacian modes	Lindley–Stein	Joint modes
		$\phi = 40$		
(a)	1.983	1.920	1.926	1.931
(b)	1.122	1.104	1.105	1.106
(c)	7.559	7.361	7.411	7.435
(d)	2.652	2.618	2.626	2.630
(e)	.723	.718	.718	.718
		$\phi = 80$		
(a)	3.966	3.719	3.746	3.833
(b)	1.587	1.533	1.536	1.548
(c)	15.117	14.394	14.615	15.146
(d)	3.750	3.659	3.684	3.739
(e)	.801	.795	.794	.794
		$\phi = 120$		
(a)	5.940	5.403	5.437	6.952
(b)	1.943	1.842	1.846	1.970
(c)	22.630	21.218	21.501	30.718
(d)	4.588	4.434	4.468	5.019
(e)	.837	.826	.826	.829

The proof of Lemma 2 is provided in the Appendix. Lemma 2 strongly justifies using either the exact posterior means of the θ_j or the Lindley–Stein estimators (26) when m is large under a wide variety of choices of Λ.

12. CONCLUSIONS

Importance sampling, Laplacian approximations, and the Gibbs sampler permit the consideration of models and prior assumptions of high complexity, because the computations can now be efficiently handled. The precise analyses of simple models is also facilitated. Thus the horizons for Bayesian research have been substantially broadened by the advent of these techniques, which replace former approaches based on joint modal estimators.

We have demonstrated that when uniform priors are assumed for variance components, the Bayesian estimates of the first-stage parameters can have excellent frequency properties. If prior knowledge is available, then these estimates can be further improved in a subjective sense via inverted chi-squared distributions for the variance components. Interval estimates and posterior probabilities are also available. For example, when investigating a hypothesis of the form H_0: $\eta = \eta_0$, we can compute the "Bayesian significance probability," $\varepsilon_0 = \rho(\eta \leq \eta_0)$, from our marginal posterior distribution for η.

These methodologies are also applicable to nonlinear variance components models; for example, the Bayesian graduation model introduced by Hickman and Miller (1981) and the genetic trait model considered by Foulley et al. (1990). The latter took observed counts $y_1, \ldots, y_m$ to be independent and Poisson distributed given their respective means $\theta_1, \ldots, \theta_m$. It also is assumed that the $\gamma_i = \log \theta_i$ satisfy

$$\gamma_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\alpha} \qquad (i = 1, \ldots, m),$$

where the $\mathbf{x}_i$ and $\mathbf{z}_i$ are specified $p \times 1$ and $q \times 1$ design vectors, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown "fixed effects." Moreover, $\boldsymbol{\alpha}$ is a vector of "random effects" that is taken to possess a multivariate normal distribution with zero mean vector and covariance matrix $\sigma^2 \mathbf{A}$. Then σ^2 is the unknown additive genetic variance, and $\mathbf{A}$ is a specified matrix of additive relationships.

If none of the observed y_i are small, then we may, as a first attempt, take the likelihood of γ_i to be approximately normal, with location $\log y_i$ and dispersion y_i^{-1}. Then all of the techniques herein are also available for drawing inferences about elements of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, unconditional on σ^2. More elaborate importance sampling and Laplacian methods are available that refer to the exact likelihoods of the γ_i (see, e.g., Tempelman and Gianola 1993). When compared to importance sampling, the Laplacian T approximation frequently has comparable accuracy and yields considerable savings in terms of CPU time. Nonetheless, it is often useful to refer to computer simulations to check the accuracy of the Laplacian approximation.

APPENDIX: PROOF OF LEMMA 2

The posterior cdf of θ_j is $\Phi[(\theta_j - \theta_j^*)/\sqrt{v_j}]$. Consequently,

$$p(\Lambda\{|\tilde{\theta}_j - \theta_j|\} \leq \lambda | \mathbf{y}) = \Phi\left(\frac{\tilde{\theta}_j + \varepsilon - \theta_j^*}{\sqrt{v_j}}\right) - \Phi\left(\frac{\tilde{\theta}_j - \varepsilon - \theta_j^*}{\sqrt{v_j}}\right), \quad \text{(A.1)}$$

where $\varepsilon = \Lambda^{-1}(\lambda)$. Furthermore, the posterior expectation of the loss function in (28) is

$$q_{\mathbf{y}}(\tilde{\boldsymbol{\theta}}) = \int_0^\infty \left[1 - \prod_{j=1}^m \Omega_j\right] d\lambda. \quad \text{(A.2)}$$

The expression in (A.1) is maximized, for any $\varepsilon > 0$, when $\tilde{\theta}_j = \theta_j^*$. Consequently, (A.2) is minimized when $\tilde{\theta}_j = \theta_j^*$ for $j = 1, \ldots, n$, thus completing the proof.

[Received August 1993. Revised August 1995.]

REFERENCES

Box, G. E. P., and Tiao, G. C. (1968), "Bayesian Estimation of Means for the Random Effects Model," *Journal of the American Statistical Association*, 63, 174–181.

Brennan, T. (1994), "Generalized Minimum Variance Deconvolution," unpublished doctoral thesis, University of Wisconsin-Madison, Dept. of Electrical and Computer Engineering.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Efron, B., and Morris, C. (1973), "Combining Possibly Related Estimations" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 35, 379–421.

Fong, D. W. H. (1992), "A Bayesian Approach to the Estimation of the Largest Normal Mean," *Journal of Statistical Computation and Simulation*, 40, 119–133.

Fortney, W. G., and Miller, R. B. (1987), "Bayesian Analysis in Random Coefficient m-Group Regression," *Statistics and Probability Letters*, 5, 135–142.

Foulley, J. L., Gianola, D., and Im, S. (1990), "Genetic Evaluation for Discrete Polygenic Traits in Animal Breeding," in *Advances in Statistical Methods for Genetic Improvement of Livestock*, eds. D. Gianola and K. Hammond, Heidelberg: Springer-Verlag, pp. 361–409.

Gelfand, A. F., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

George, E., Makov, U., and Smith, A. F. M. (1994), "Fully Bayesian Hierarchical Analysis for Exponential Families Via Monte Carlo Computation," in *Aspects of Uncertainty*, eds. P. R. Freeman and A. F. M. Smith, New York: John Wiley, pp. 181–196.

Geweke, G. (1988), "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference," *Journal of Econometrics*, 38, 73–89.

——— (1989), "Exact Predictive Density for Linear Models with Arch Distributions," *Journal of Econometrics*, 40, 63–86.

Hald, A. (1952), *Statistical Theory with Engineering Applications*, New York: John Wiley.

Hickman, J. C., and Miller, R. B. (1981), "Bivariate Bayesian Graduation and Forecasting," *Scandinavian Actuarial Journal*, 3, 129–150.

Hill, B. M. (1965), "Inference About Variance Components in the One-Way Model," *Journal of the American Statistical Association*, 60, 806–825.

Kass, R. E., and Steffey, D. (1989), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models," *Journal of the American Statistical Association*, 84, 717–726.

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the American Society of Mechanical Engineering*, Ser. D, 82, 35–45.

Laird, N. M. (1978), "Empirical Bayes Methods for Two-Way Contingency Tables," *Biometrika*, 65, 581–590.

Leonard, T. (1972), "Bayesian Methods for Binomial Data," *Biometrika*, 59, 581–589.

——— (1976), "Some Alternative Approaches to Multiparameter Estimations," *Biometrika*, 63, 69–75.

——— (1982), "Comment on "A Simple Predictive Density Function," by M. Lejeune and G. D. Faulkenberry," *Journal of the American Statistical Association*, 77, 657–658.

Leonard, T., and Hsu, J. (1992), "Bayesian Inference for a Covariance Matrix," *The Annals of Statistics*, 20, 1669–1696.

Leonard, T., Hsu, J., and Ritter, C. (1994), "The Laplacian T-Approximation in Bayesian Inference," *Statistica Sinica*, 4, 127–142.

Leonard, T., Hsu, J., and Tsui, K.-W. (1989), "Bayesian Marginal Inference," *Journal of the American Statistical Association*, 84, 1051–1058.

Lindley, D. V. (1962), Discussion of "Confidence Sets for the Mean of a Multivariate Normal Distribution," by C. Stein, *Journal of the Royal Statistical Society*, Ser. B, 24, 265–296.

——— (1971), "The Estimation of Many Parameters," in *Foundations of Statistical Inference*, eds. V. P. Godanbe and D. A. Sprott, Toronto: Holt, Rinehart and Winston, pp. 435–455.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Lord, F. M., and Novick, M. (1967), *Statistical Theories of Mental Test Scores*, Reading, MA: Addison-Wesley.

O'Hagan, A. (1976), "On Posterior Joint and Marginal Modes," *Biometrika*, 63, 329–333.

Miller, R. B., and Fortney, W. E. (1984), "Industry-Wide Expense Standards Using Random Coefficient Regression," *Insurance Mathematics and Economics*, 3, 19–33.

Miller, R. B., and Hickman, J. C. (1975), "Insurance, Credibility and Bayesian Estimation. Credibility," in *Theory and Applications, Proceedings of the Berkeley Actuarial Research Conference*, ed. P. M. Kahn, New York: Academic Press, pp. 249–279.

Reinsel, G. (1985), "Mean Squared Error Properties of Empirical Bayes Estimators in a Multivariate Random-Effects General Linear Model," *Journal of the American Statistical Association*, 80, 646–650.

Reinsel, G., and Tiao, G. (1987), "Impact of Chlorofluoromethanes on Stratospheric Ozone: A Statistical Analysis of Ozone Data for Trends," *Journal of the American Statistical Association*, 82, 20–30.

Scott, A., and Smith, T. M. F. (1969), "Estimation in Multistage Surveys," *Journal of the American Statistical Association*, 64, 830–840.

Searle, S. R., Casella, G., and McCullogh, C. E. (1992), *Variance Components*, New York: John Wiley.

Stein, C. (1956), "Inadmissibility of the Usual Estimator of the Mean of a Multivariate Normal Distribution," in *Proceedings of the Third Berkeley Symposium*, 1, 197–206.

Strawderman, W. (1971), "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," *The Annals of Statistics*, 42, 385–388.

Sun, L. (1992), "Bayesian Estimation Procedures for One- and Two-Way Hierarchical Models," unpublished doctoral thesis, University of Toronto.

Tempelman, R. J., and Gianola, D. (1993), "Marginal Maximum Likelihood Estimation of Variance Components in Poisson Mixed Models, Using Monte Carlo Integration," *Journal of Genetic Selection and Evolution*, 25, 305–319.

Tierney, L., and Kadane, J. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Zellner, A., and Rossi, P. E. (1984), "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometrics*, 25, 365–393.