Statistica Sinica 4(1994), 127-142

THE LAPLACIAN T-APPROXIMATION IN BAYESIAN INFERENCE

Tom Leonard, John S. J. Hsu*, and Christian Ritter

University of Wisconsin-Madison

*and University of California at Santa Barbara**

Abstract: A remarkably accurate approximation is proposed for a marginal density, for finite sample situations where the tails of the posterior density are not accurately represented by a more standard Laplacian approximation. An approximation is developed for the posterior density of an arbitrary linear combination of the means, in the context of the Bayesian analysis of the multi-parameter Fisher-Behrens problem. Advantages of Laplacian methods for non-linear regression problems, when compared with sampling based methods, are discussed.

Key words and phrases: Marginal posterior density, Laplacian approximations, importance sampling, Behrens-Fisher problem, non-linear regression, Grid-based Gibbs sampler, Metropolis algorithm.

1. The Laplacian T-Approximation

A $p \times 1$ vector $\boldsymbol{\theta}$ is said to possess a multivariate t-distribution, with ν degrees of freedom, mean vector $\boldsymbol{\mu}$, and precision matrix $\boldsymbol{R}$, if $\boldsymbol{\theta}$ possesses density

$$t_{\boldsymbol{\theta}}(\nu, p, \boldsymbol{\mu}, \boldsymbol{R}) = C_{\nu,p}|\boldsymbol{R}|^{\frac{1}{2}}\Big[1 + \nu^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu})^T\boldsymbol{R}(\boldsymbol{\theta} - \boldsymbol{\mu})\Big]^{-\frac{1}{2}(\nu+p)} \quad (\boldsymbol{\theta} \in R^p), \quad (1.1)$$

where

$$C_{\nu,p} = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)\pi^{\frac{1}{2}p}\nu^{\frac{1}{2}p}}, \quad (1.2)$$

and R^p denotes p-dimensional Euclidean space.

Consider a posterior density $\pi_y(\eta, \boldsymbol{\xi})$ for unconstrained parameters η and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_q)^T$, and suppose that an approximation is required to the marginal posterior density

$$\pi_y(\eta) = \int_{R^q} \pi_y(\eta, \boldsymbol{\xi})d\boldsymbol{\xi} \quad (\eta \in R^1). \quad (1.3)$$

Following Leonard (1982), Leonard and Novick (1986), Tierney et al. (1989), and Hsu et al. (1991), let $\boldsymbol{\xi}_\eta$ denote some choice of conditional vector for $\boldsymbol{\xi}$, given η, and let

$$\boldsymbol{U}_\eta = -\frac{\partial^2 \log \pi_y(\eta, \boldsymbol{\xi}_\eta)}{\partial(\boldsymbol{\xi}_\eta \boldsymbol{\xi}_\eta^T)}, \tag{1.4}$$

denote the corresponding conditional information matrix. When $\boldsymbol{\xi}_\eta$ conditionally maximizes $\pi(\eta, \boldsymbol{\xi})$, given η, these authors develop the Laplacian approximation

$$\tilde{\pi}_y(\eta) \propto \pi_y(\eta, \boldsymbol{\xi}_\eta)/|\boldsymbol{U}_\eta|^{\frac{1}{2}} \quad (\eta \in R^1), \tag{1.5}$$

to $\pi_y(\eta)$. The approximation (1.5) is exact whenever $\pi_y(\eta, \boldsymbol{\xi})$ is either a $p = q+1$ dimensional multivariate t-density, or a multivariate normal density. It is also remarkably accurate in situations where a preliminary transformation has been chosen to ensure that the conditional distribution of $\boldsymbol{\xi}$, given η, is approximately multivariate normal. It can be justified via a second order Taylor Series approximation to $\log \pi_y(\eta, \boldsymbol{\xi})$.

Nevertheless, in a range of examples, in particular the Behrens-Fisher problem discussed in Section 3, and the hierarchical models analysed by Sun (1992), (1.5) inadequately approximates the tails of the marginal posterior density (1.3). Therefore we consider, instead, a Taylor Series expansion of $[\pi_y(\eta, \boldsymbol{\xi})]^{-\alpha}$ about an arbitrary $\boldsymbol{\xi} = \boldsymbol{\xi}_\eta$, where α is typically constant in η, but may depend on η, giving

$$[\pi_y(\eta, \boldsymbol{\xi})]^{-\alpha} = [\pi_y(\eta, \boldsymbol{\xi}_\eta)]^{-\alpha}\Big[1 - \alpha \boldsymbol{\ell}_\eta^T(\boldsymbol{\xi} - \boldsymbol{\xi}_\eta) + \frac{1}{2}\alpha(\boldsymbol{\xi} - \boldsymbol{\xi}_\eta)^T \boldsymbol{Q}_\eta(\boldsymbol{\xi} - \boldsymbol{\xi}_\eta) + \text{cubic and higher terms}\Big], \tag{1.6}$$

where

$$\boldsymbol{\ell}_\eta = \frac{\partial \log \pi_y(\eta, \boldsymbol{\xi}_\eta)}{\partial \boldsymbol{\xi}_\eta}, \tag{1.7}$$

$$\boldsymbol{Q}_\eta = \boldsymbol{U}_\eta + \alpha \boldsymbol{\ell}_\eta \boldsymbol{\ell}_\eta^T, \tag{1.8}$$

and $\boldsymbol{U}_\eta$ satisfies (1.4). Now, set

$$\alpha = \frac{2}{\nu + q}, \tag{1.9}$$

where ν is assumed positive. Then, neglecting cubic and higher terms in (1.6), and raising both sides to the power $-(\nu+q)/2$, yields, after some rearrangement, the approximation

$$\pi_y^*(\eta, \boldsymbol{\xi}) = \pi_y(\eta, \boldsymbol{\xi}_\eta)\lambda_\eta^{-\frac{1}{2}(\nu+q)}\Big[1 + \nu^{-1}(\boldsymbol{\xi} - \boldsymbol{\xi}_\eta^*)^T \boldsymbol{T}_\eta(\boldsymbol{\xi} - \boldsymbol{\xi}_\eta^*)\Big]^{-\frac{1}{2}(\nu+q)}, \tag{1.10}$$

where

$$\boldsymbol{\xi}_\eta^* = \boldsymbol{\xi}_\eta + \boldsymbol{Q}_\eta^{-1}\boldsymbol{\ell}_\eta, \tag{1.11}$$

$$\lambda_\eta = 1 - (\nu + q)^{-1}\boldsymbol{\ell}_\eta^T \boldsymbol{Q}_\eta^{-1}\boldsymbol{\ell}_\eta, \tag{1.12}$$

and

$$\boldsymbol{T}_\eta = \frac{\nu \boldsymbol{Q}_\eta}{(\nu + q)\lambda_\eta}. \tag{1.13}$$

If $\lambda_\eta > 0$, and $\boldsymbol{T}_\eta$ is positive definite, then the right hand side of (1.10) is proportional to a $t_{\boldsymbol{\xi}}(\nu, q, \boldsymbol{\xi}_\eta, \boldsymbol{T}_\eta)$ density. Therefore, integrating over $\boldsymbol{\xi} \in R^q$ yields the approximation

$$\pi_y^*(\eta) \propto \frac{\pi_y(\eta, \boldsymbol{\xi}_\eta)}{|\boldsymbol{Q}_\eta|^{\frac{1}{2}}} \lambda_\eta^{-\frac{1}{2}\nu} \frac{\Gamma(\nu/2)(\nu+q)^{\frac{1}{2}q}}{\Gamma[(\nu+q)/2]} \tag{1.14}$$

to the marginal posterior density $\pi_y(\eta)$ of η, where ν could depend upon η. In the special case where $\boldsymbol{\xi}_\eta$ is the conditional maximum of (1.10), with respect to $\boldsymbol{\xi}$, given η, the "Laplacian T-approximation" (1.14) reduces to

$$\pi_y^*(\eta) \propto \tilde{\pi}_y(\eta) \frac{\Gamma(\nu/2)(\nu+q)^{\frac{1}{2}q}}{\Gamma((\nu+q)/2)} \tag{1.15}$$

where $\tilde{\pi}_y(\eta)$ is the ordinary Laplacian approximation (1.5). Therefore, if ν is constant in η, the Laplacian T-approximation based upon the conditional mode $\boldsymbol{\xi}_\eta$, does not improve (1.5). A choice of ν, dependent upon η, would modify (1.5). However, our primary recommendation for improving the approximation to (1.3) is to seek alternative choices of $\boldsymbol{\xi}_\eta$ (see Section 3). The approximation (1.14) is then dependent upon the choice of ν, but is rather robust under modest changes to ν. The latter should be chosen to ensure that the conditional posterior density of $\boldsymbol{\xi}$, given η, is approximately a multivariate t-density with ν degrees of freedom; and this would be implied by the requirement that the posterior density $\pi_y(\boldsymbol{\theta})$ of $\boldsymbol{\theta} = (\eta, \boldsymbol{\xi})^T$ is approximately a multivariate t-density with ν degrees of freedom.

While our procedure does not give a general criterion for choosing $\boldsymbol{\xi}_\eta$, some natural choices are available for the Behrens-Fisher problem (see Section 3), and for Bayesian hierarchical models (see Sun (1992) and Sun, Guttman, and Leonard (1993)). As a general guideline, $\boldsymbol{\xi}_\eta$ might be viewed as some reasonable approximation to the conditional posterior mean vector of $\boldsymbol{\xi}$, given η, or as some reasonable measure of center of the posterior density of $\boldsymbol{\xi}$, given η. Mode vectors do not necessarily provide sensible measures of center. The choice of ν can vary, according to the particular parameter of interest under consideration.

2. Choosing the Degrees of Freedom

In cases where the exact second and fourth central posterior moments M_2 and M_4 of η exist, and are computable, ν can be evaluated by equating the exact kurtosis $K = M_4/M_2^2$ to the kurtosis of the approximation (1.14), obtained via the appropriate one-dimensional numerical integrations.

More generally, the posterior density $\pi_y(\boldsymbol{\theta}) = \pi_y(\eta, \boldsymbol{\xi})$ of $\boldsymbol{\theta} = (\eta, \boldsymbol{\xi})^T$ may be approximated by a $t_{\boldsymbol{\theta}}(\nu, p, \tilde{\boldsymbol{\theta}}, \boldsymbol{T})$ density, where $p = q + 1$, $\tilde{\boldsymbol{\theta}}$ denotes the unconditional posterior mode vector, and $\boldsymbol{T} = \nu \boldsymbol{Q}/(\nu + p)$, where $\boldsymbol{Q}$ is the posterior information matrix. Consequently, the marginal posterior density of η may be approximated by a $t_\eta(\nu, 1, \boldsymbol{a}^T\tilde{\boldsymbol{\theta}}, (\boldsymbol{a}^T\boldsymbol{T}^{-1}\boldsymbol{a})^{-1})$ density, where $\boldsymbol{a} = (1, 0, 0, \ldots, 0)^T$. This t-density may be compared graphically with the Laplacian T-approximation (1.14), for the same choice of ν. We recommend a choice of ν which ensures that these two alternative approximations, to the marginal posterior density of η, are as numerically close to each other as feasible. This also provides a general procedure for approximating a posterior density by a multivariate t-density, when a particular parameter is designated to be of interest. Methods based upon equating fourth derivatives do not appear to possess general applicability, owing to their restrictive forms, when differentiating the logarithm of a multivariate t-density. The proposed procedure is, however, designed to correctly evaluate ν, whenever the posterior density of $\boldsymbol{\theta}$ is exactly a t-density with ν degrees of freedom.

3. The Multi-Parameter Fisher-Behrens Problem

Given $(\theta_1, \phi_1), \ldots, (\theta_p, \phi_p)$, consider observations y_{ij} which are independent, with respective means θ_i and variances ϕ_i, $(i = 1, \ldots, m; j = 1, \ldots, n_i)$. Take the θ_i and $\log \phi_i$ to be a priori independent and uniformly distributed over R^1. Then the posterior density of $\theta_1, \ldots, \theta_n$ is

$$\pi(\boldsymbol{\theta}|y) \propto \prod_i \left[U_i(\theta_i)\right]^{-\frac{1}{2}n_i}, \tag{3.1}$$

where

$$U_i(\theta_i) = S_i^2 + n_i(\theta_i - \bar{y}_i)^2, \tag{3.2}$$

and $\bar{y}_i$ and S_i^2 denote the sample mean, and within group sum of squares, for the ith group.

The methodology of Section 1, may be used to approximate the posterior density of an arbitrary linear transformation $\eta = \sum a_i\theta_i$ of $\theta_1, \ldots, \theta_p$. Let $\xi_1 = \theta_1, \ldots, \xi_q = \theta_q$, with $q = p - 1$. Following O'Hagan (1976), we note that, when the joint posterior density is the product of multivariate t-densities, and the parameters are not independent, the joint posterior mode vector need not reasonably characterize the posterior distribution. This aspect is investigated in detail by Sun (1992) in the context of the "Lindley-Smith collapsing phenomenon". In the current context, the joint posterior modes of $\xi_1, \ldots, \xi_q$, given η, may not reasonably represent the location of the conditional posterior density of $\xi_1, \ldots, \xi_q$, given η. Therefore, following a suggestion by O'Hagan, we consider instead, the joint posterior modes $\xi_1^{(\eta)}, \ldots, \xi_q^{(\eta)}$, of $\xi_1, \ldots, \xi_q$, given $\phi_1, \ldots, \phi_p$, and η, and then replace ϕ_i by $\hat{\phi}_i = n_i^{-1} S_i^2$, for $i = 1, \ldots, p$. These conditional modes are denoted by

$$\xi_i^{(\eta)} = \bar{y}_i + n_i^{-1} \hat{\phi}_i a_i (\eta - \hat{\eta}) \Big/ \sum_k n_k^{-1} a_k^2 \hat{\phi}_k \quad (i = 1, \ldots, q). \tag{3.3}$$

The matrix (1.4) can be obtained in algebraic form from the joint posterior density of η and $\xi_1, \ldots, \xi_q$, unconditionally upon $\phi_1, \ldots, \phi_q$. It will not remain positive definite for all values of η and $\boldsymbol{\xi}_\eta = (\xi_1^{(\eta)}, \ldots, \xi_q^{(\eta)})^T$, in particular for values of η lying in the tails of the posterior density. However, when calculating the Laplacian T-approximation (1.14), the $\xi_i^{(\eta)}$ can be replaced by the corresponding

$$\tilde{\xi}_i^{(\eta)} = \begin{cases} \xi_i^{(\eta)} & \text{for } a < \eta < b \\ \xi_i^{(a)} & \text{for } \eta \le a \\ \xi_i^{(b)} & \text{for } \eta \ge b, \end{cases} \tag{3.4}$$

where a and b are chosen to ensure that $\boldsymbol{U}_\eta$ in (1.4) is replaced by a matrix $\tilde{\boldsymbol{U}}_\eta$ which always remains positive definite. The corresponding Laplacian T-approximation (1.14) is continuous, and well-defined, even in the tails. It will be different from the ordinary Laplacian approximation (1.5), owing to the choices (3.3) and (3.4).

Consider the Poland China Pig data, reported by Scheffé (1959, p.87). The data comprise the birthweights in $p = 8$ litters, with 10, 8, 10, 8, 6, 4, 6, and 4 pigs in the different litters. The group sample means were respectively 2.84, 2.66, 3.18, 2.98, 2.37, 2.90, 1.98, and 2.35, and the maximum likelihood estimates of the sample variances were respectively 0.818, 0.435, 0.068, 0.089, 0.122, 0.060, 0.328 and 0.293. A parameter of particular interest is the contrast

$$\eta = \frac{1}{3}(\theta_1 + \theta_3 + \theta_4) - \frac{1}{5}(\theta_2 + \theta_5 + \theta_6 + \theta_7 + \theta_8) \tag{3.5}$$

since this relates to the difference in average weights between the offspring of two particular boars. However, results of similar accuracy were obtained for many choices of contrast.

Curve (a) of Fig. 1 describes the Laplacian T-approximation (1.14) to the posterior density of η, with $\nu = 13.5$, the elements of $\boldsymbol{\xi}_\eta$ replaced by the appropriate $\tilde{\xi}_i^\eta$ in (3.3), $a = 0.1103$, and $b = 1.2020$. The choice $\nu = 13.5$ was obtained by the graphical method of Section 2, and curve (a) of Fig. 1 is virtually indistinguishable from a $t_\eta(13.5, 1, 0.546, 32.161)$ density. The kurtosis $K = M_4/M_2^2$ corresponding to density (a) is $K = 0.355$. Curve (b) of Fig. 1 describes a similar Laplacian T-approximation, but with $\nu = 6$. The value $\nu = 6$ was obtained by the method described in the first paragraph of Section 2, and evaluates the exact kurtosis $K = 0.294$ correctly, to three decimal places. Curve (c) is the Laplacian approximation (1.5), and is also virtually indistinguishable from an ordinary Laplacian approximation based upon the joint posterior density of the θ_i and $\log \phi_i$. Histogram (d) represents the exact posterior density, based upon 500,000 Monte Carlo simulations (consuming 23 mins. of CPU time on a Sunsparc station, where, on each simulation, values were generated from the eight independent t-distributions appearing in the posterior distribution (3.1). Hence, straightforward Monte Carlo, rather than more complicated importance sampling, (e.g., Leonard, Hsu, and Tsui (1989)) was employed.

We conclude that the Laplacian T-approximation gives greater accuracy in the tails, in the current situation, when compared with the ordinary Laplacian approximation, and that it would only be necessary to obtain a sensible evaluation of ν in order to establish this superiority. The exact, computer simulated probability, that $\eta < 0$, is 0.00530, and this may be compared with the approximations 0.00475, 0.00382, and 0.00118, which respectively correspond to curves (a), (b), and (c) of Fig. 1. These results indicate that ν should not be taken too small (e.g., $\nu = 6$), or too large (e.g., $\nu = \infty$), but that the choice $\nu = 13.5$ is adequate. The tails of the approximations (a) and (b) are slightly too thin outside the interval $(-0.1103, 1.2020)$, but in the left tail this causes differences from the exact result of an area no greater than 0.00168. Note that the usual t-test, for $\eta = 0$, assuming equal variances, yields a one-sided significance probability of about 0.0008, and does not therefore approximate our results well in the unequal variance case.

4. Non-Linear Transformations

Consider now a vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_q)^T$ of unknown parameters, which possesses posterior density $\pi_y(\boldsymbol{\xi})$, and consider approximations to the marginal posterior density

$$\pi_y(\eta) = \lim_{\epsilon \to 0} \int_D \pi_y(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{4.1}$$

of a specified non-linear transformation $\eta = g(\boldsymbol{\xi})$ of $\boldsymbol{\xi}$, where $D = \{\boldsymbol{\xi} : |\eta - g(\boldsymbol{\xi})| < \epsilon\}$. By a parallel development to the theory of Section 1, $\pi_y(\eta)$ is approximated by

$$\pi_y^*(\eta) \propto C_{\nu,q}^{-1} |\boldsymbol{T}_\eta|^{-\frac{1}{2}} \pi_y(\boldsymbol{\xi}_\eta) \lambda_\eta^{-\frac{1}{2}(\nu+q)} f_\eta(\nu, \boldsymbol{\xi}_\eta^*, \boldsymbol{T}_\eta), \tag{4.2}$$

where $C_{\nu,q}$, $\boldsymbol{\xi}_\eta^*$, λ_η, and $\boldsymbol{T}_\eta$ respectively satisfy (1.2), (1.11), (1.12), and (1.13), but with $\pi_y(\eta, \boldsymbol{\xi}_\eta)$ in the expressions for $\boldsymbol{U}_\eta$ and $\boldsymbol{\ell}_\eta$ in (1.4) and (1.7), replaced by $\pi_y(\boldsymbol{\xi}_\eta)$. Also, $\boldsymbol{\xi}_\eta$ in (1.11) denotes an arbitrary conditional vector for $\boldsymbol{\xi}$, given η, and $f_\eta(\nu, \boldsymbol{\xi}^*, \boldsymbol{T})$ is the density of $\eta = g(\boldsymbol{\xi})$ when $\boldsymbol{\xi}$ possesses a $t(\nu, q, \boldsymbol{\xi}^*, \boldsymbol{T})$ density.

Some further approximations or simulations may be needed for the f_η contribution to (4.2). As $\nu \to \infty$, (4.2) reduces to the LHT approximation, which was shown in several numerical examples, by Leonard et al. (1989), and Hsu et al. (1991), to possess excellent numerical accuracy, when compared with the exact simulated result. The refinement (4.2) will affect the tails of the approximation. The degrees of freedom ν can be evaluated by comparing (4.1) with the density of η, under a multivariate t-approximation of the type introduced in Section 2, to the posterior density of $\boldsymbol{\xi}$.

In the special case when $\eta = g(\boldsymbol{\xi}) = \boldsymbol{a}^T\boldsymbol{\xi}$ denotes a linear transformation of $\boldsymbol{\xi}$, the f_η contribution to (4.2) satisfies

$$f_\eta(\nu, \boldsymbol{\xi}_\eta^*, \boldsymbol{T}_\eta) = t_\eta\left(\nu, 1, \boldsymbol{a}^T\boldsymbol{\xi}_\eta^*, (\boldsymbol{a}^T\boldsymbol{T}_\eta^{-1}\boldsymbol{a})^{-1}\right) \tag{4.3}$$

yielding an alternative representation of the Laplacian T-approximation (1.14), and generalizing an algebraic rearrangement of (1.5), recommended by Tierney et al. (1989).

5. Laplacian and Sampling Based Methods in Non-Linear Regression

Consider non-linear regression models of the form

$$y_i = f(x_i, \boldsymbol{\theta}) + \epsilon_i \quad (i = 1, \ldots, n) \tag{5.1}$$

where the ϵ_i are independent, and normally distributed with zero means and common unknown variance σ^2, f is a specified function of x_i and $\boldsymbol{\theta}$, $\boldsymbol{\theta}$ is an unknown $p \times 1$ vector of parameters, and $x_1, \ldots, x_n$ are specified constants. If $\boldsymbol{\theta}$ is a priori uniformly distributed over R^1, and independent of $\log \sigma^2$ which is uniformly distributed over R^1, then the posterior density of $\boldsymbol{\theta}$, if this exists, is

$$\pi_y(\boldsymbol{\theta}) \propto [\rho(\boldsymbol{\theta})]^{-\frac{1}{2}(\omega+p)}, \tag{5.2}$$

where

$$\rho(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - f(x_i, \boldsymbol{\theta}))^2, \tag{5.3}$$

and $\omega = n - p$. If instead σ^2 is uniformly distributed over $(0, \infty)$, then $\omega = n - p - 2$. If the posterior density exists then the approximation (1.5) often works well, and may be applied to compute the marginal posterior density of any element of $\boldsymbol{\theta}$ e.g., $\eta = \theta_1$. The conditional maximum $\boldsymbol{\theta}_\eta$ may be computed by reference to the profile likelihood methods of Bates and Watts (1988, p.205). The underlying algorithms are currently in the statistical software package S. The algorithms can break down, e.g., the conditional optimization procedures may fail in the tails of the posterior density, or the conditional Hessian $\boldsymbol{U}_\eta$ in (1.4) may become close to singular. In such circumstances it is better to refer to sampling based methods, e.g., importance sampling, (e.g., Leonard et al. (1989) and their bibliography, Tanner and Wong (1987), Leonard and Hsu (1992)), the Gibbs sampler (Carlin and Gelfand (1990), Gelfand and Smith (1990)), or the Metropolis algorithm (Metropolis et al. (1953), Müller (1991)). However, if the Bates-Watts algorithm succeeds, then the ordinary Laplacian approximation (1.5) can be extremely accurate, and can be much less computer intensive than sampling based methods. Leonard and Hsu (1992) consider more complex examples where the profile likelihood methods become overcomplicated, and it is necessary to devise special sampling methods. However, in the current context, (1.5) frequently works well without the modification (1.14), so that further choices of $\boldsymbol{\xi}_\eta$ and ν are unnecessary.

As an illustrative example (see Ratkowsky (1983, p.58)), consider the choice of regression function

$$f(x_i, \boldsymbol{\theta}) = \frac{\theta_1}{1 + \theta_2 x_i + \theta_3 x_i^2}. \tag{5.4}$$

The parameter θ_1 refers to a genetic potential, while θ_2 and θ_3 relate to the loss in yield y_i due to the value of a density x_i. Owing to unusual shapes of the tails of the likelihood function, the posterior density of θ_1, θ_2, and θ_3 does not properly exist under the prior assumptions discussed above. Therefore we assume instead that our uniform prior distributions for θ_1, θ_2, and θ_3, are truncated to the finite regions $0 \le \theta_1 \le 2000$, $0 \le \theta_2 \le 1$, and $-0.001 \le \theta_3 \le 0.001$, and that σ^2 is uniformly distributed over $(0, \infty)$. This ensures that the posterior density captures the domain of reasonable parameter values, while reaching out far enough to expose some of the tail behavior of the likelihood. This problem provides a possible counterexample to importance sampling, since it is difficult to find an approximation to the posterior density, sufficiently accurate in the tails, from which it is possible to simulate trial vectors $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$.

For the MG data ($n = 42$) set considered by Ratkowsky, the corresponding Laplacian approximations, are described by the solid curves in Figs. 2, 3, and 4. Each of the three approximations involved 800 evaluations of the posterior density, and the approximations based on these curves are accurate, to about 3 decimal places, to the exact curves obtained by numerical integration. They are also close to the less accurate histograms in Figs. 2, 3, and 4, which involved heavy computer simulation. Note that the Laplacian approximations ignore the constraints on the parameter space imposed by the prior, but are still virtually identical to the exact results based upon the truncated posterior.

The histograms are based upon a combination of the Grid-based Gibbs sampler (Ritter and Tanner (1992)), and the version of the Metropolis algorithm suggested by Müller (1991). 100 parallel chains were generated by several initial iterations of the Grid-based Gibbs sampler, followed by 800 iterations of the Metropolis algorithm. After 600 iterations, the quantile plots indicated convergence, and a Monte Carlo sample of 1000 observations was extracted by combining every 20th iteration for the last 200 iterations. A total of 100,000 evaluations of the posterior density were therefore required. While the computer efficiency of this sampling based method could be enhanced by reference to suggestions made by Tierney (1991), the Laplacian approximation is clearly much more efficient in terms of numbers of evaluations required of the posterior density. While the histograms in Figs. 2, 3, and 4 could be replaced by smooth curves, e.g., using a procedure recommended by Gelfand and Smith (1990), which averages conditional densities, these curves would not approximate the true curves as well as the Laplacian approximation, unless a very large Monte Carlo sample were used. Other examples are reported by Ritter, Bisgaard, and Bates (1991) and Ritter (1992). Overall, we conclude that the Laplacian and Laplacian T-approximations yield distinct computational advantages, when compared with sampling based methods, for a broad range of statistical problems. The approximation error is, for example, typically negligible, when compared with possible inexactness in the specification of the prior distribution. The Laplacian methods do however need to be reformulated for every particular parameter of interest, while sampling based methods can consider many parameters of interest simultaneously.

6. Further Work-Hierarchical Bayes Models

Sun (1992) applies the Laplacian T-approximation to Hierarchical Bayes models, in particular the multi-way analysis of variance random effects model with interaction effects discussed by Box and Tiao (1973, p.292). The conditional mode procedures introduced in Section 3, lead to algebraically explicit approximations, in the equally replicated two-way case, to the marginal posterior density of any marginal or interaction effect of interest, and hence effectively solve a problem which creates severe difficulties for sampling based methods. Sun validates these further applications of the Laplacian T-approximation by demonstrating that the corresponding approximate Bayesian intervals possess excellent frequency properties, under particular choices of prior distribution. These results are also reported by Sun, Guttman, and Leonard (1993).

Acknowledgements

The large Bayesian literature, during the last decade on approximate marginalization via conditional maximization, was initiated by discussions between the first co-author, and George C. Tiao, in 1981, and the derivation in Section 1 generalizes a suggestion made by Tiao. The authors are also indebted to Luke Tierney, for a variety of helpful comments and Richard Johnson, Douglas Bates, Irwin Guttman, Martin Tanner, Bradley Carlin, Li Sun, and Kam-Wah Tsui, and two referees for their encouragement.

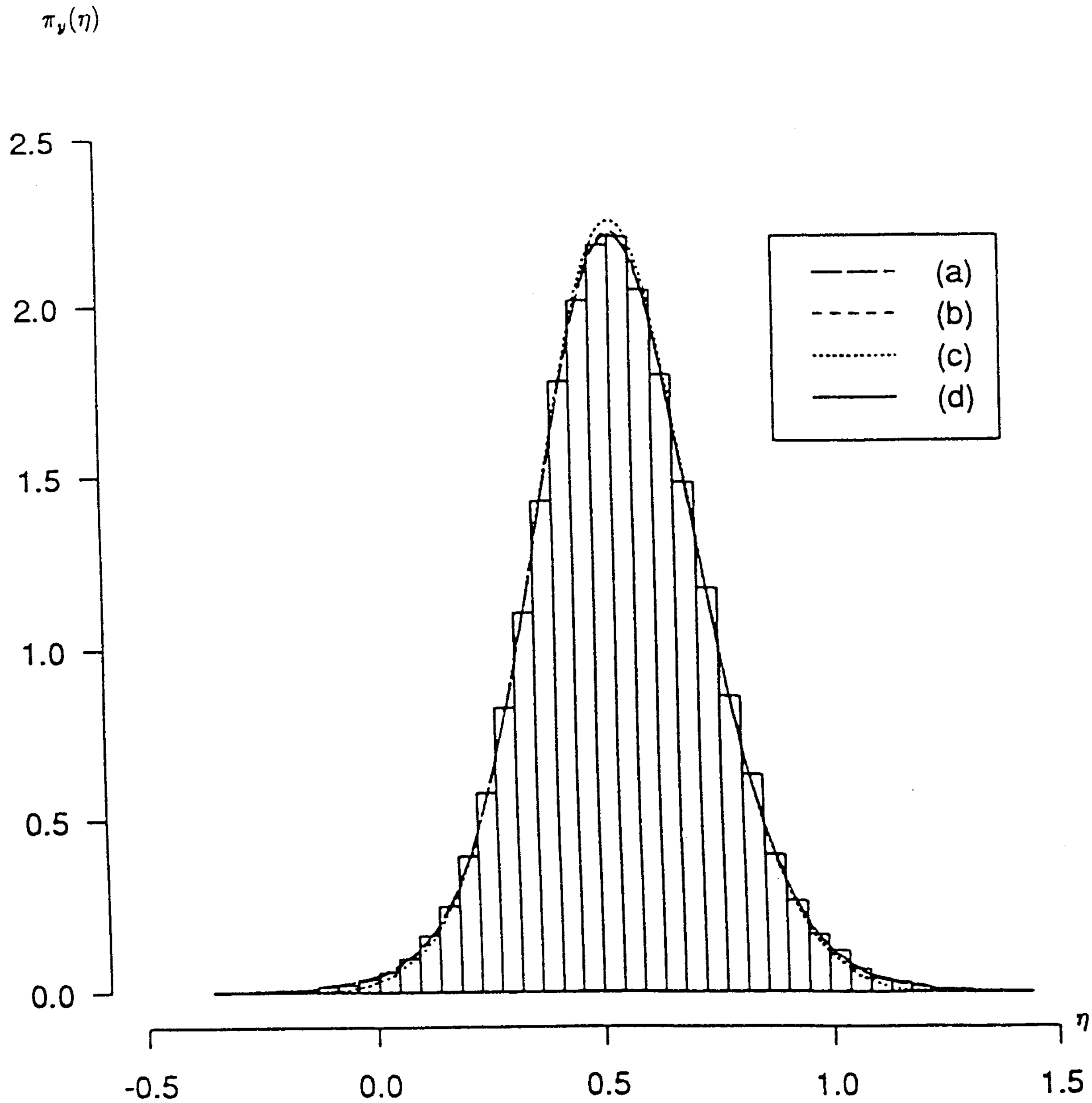

Figure 1. Marginal posterior density of linear contrast
(a) Laplacian T-approximation ($\nu = 13.5$)
(b) Laplacian T-approximation ($\nu = 6$)
(c) Ordinary Laplacian approximation
(d) Histogram, simulated by Monte Carlo

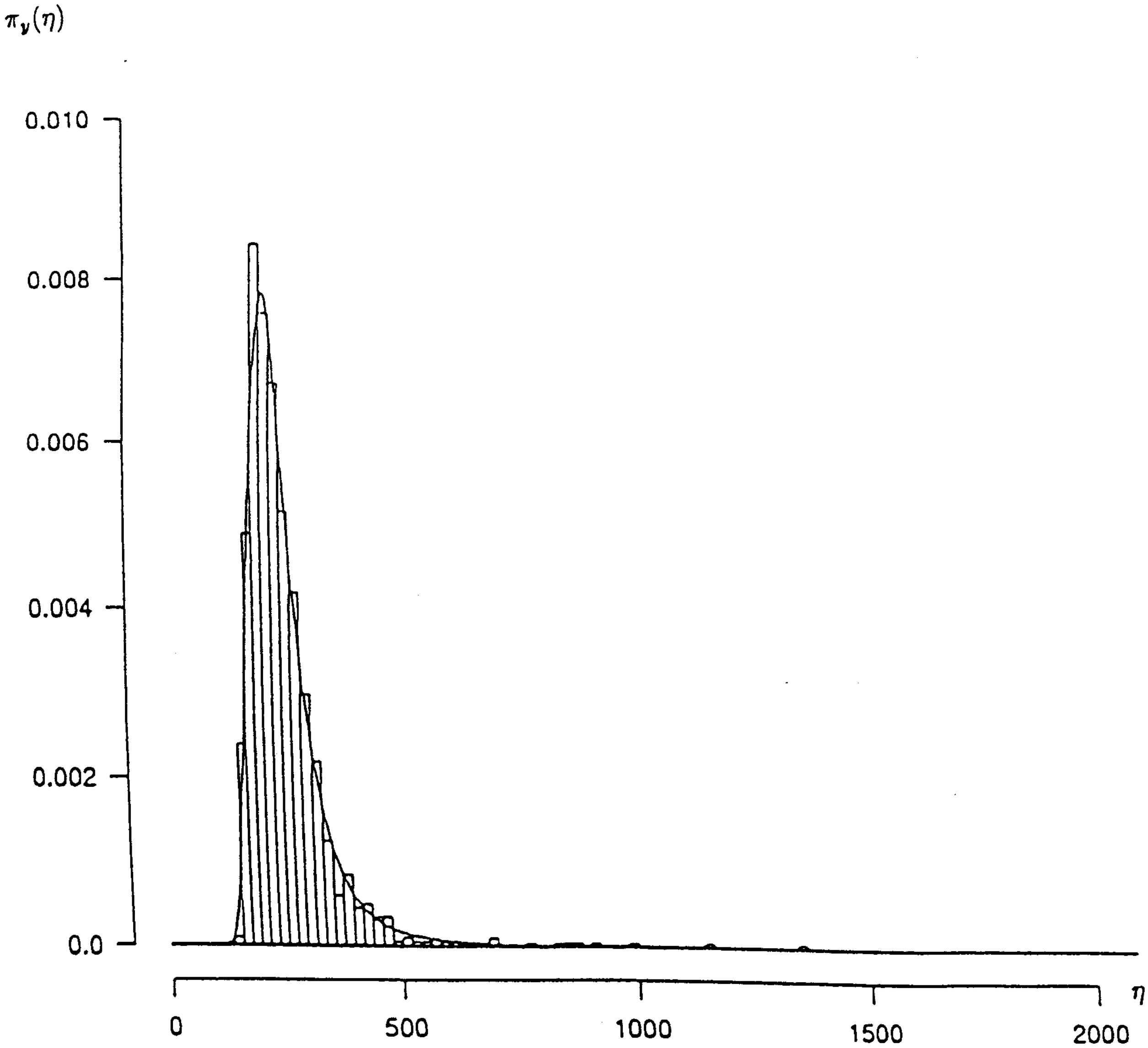

Figure 2. Marginal posterior density of genetic potential
Solid curve: Laplacian T-approximation (virtually exact)
Histogram: Simulated by Gibbs sampler/Metropolis algorithm

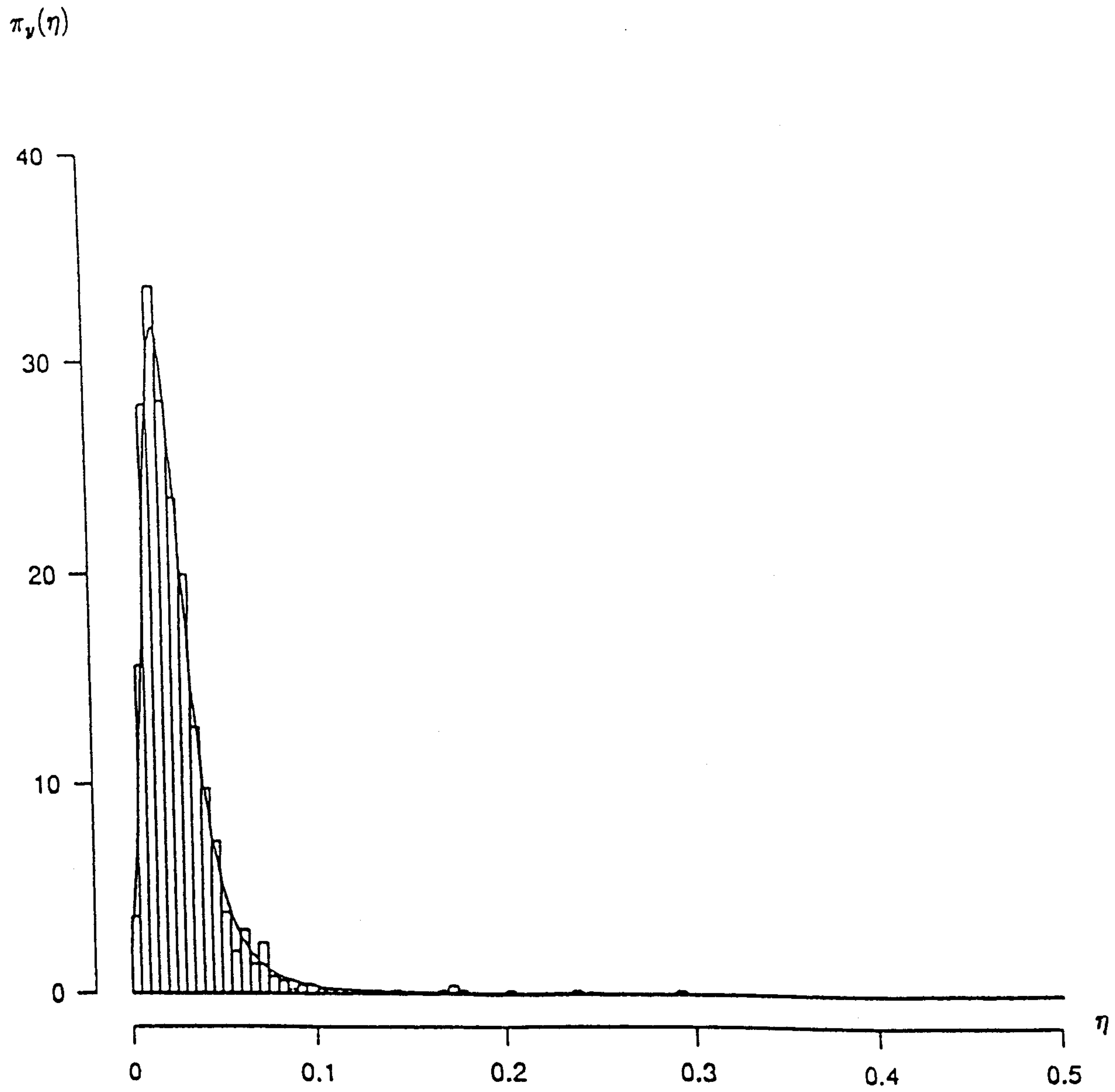

Figure 3. Marginal posterior density of linear coefficient
Solid curve: Laplacian T-approximation (virtually exact)
Histogram: Simulated by Gibbs sampler/Metropolis algorithm

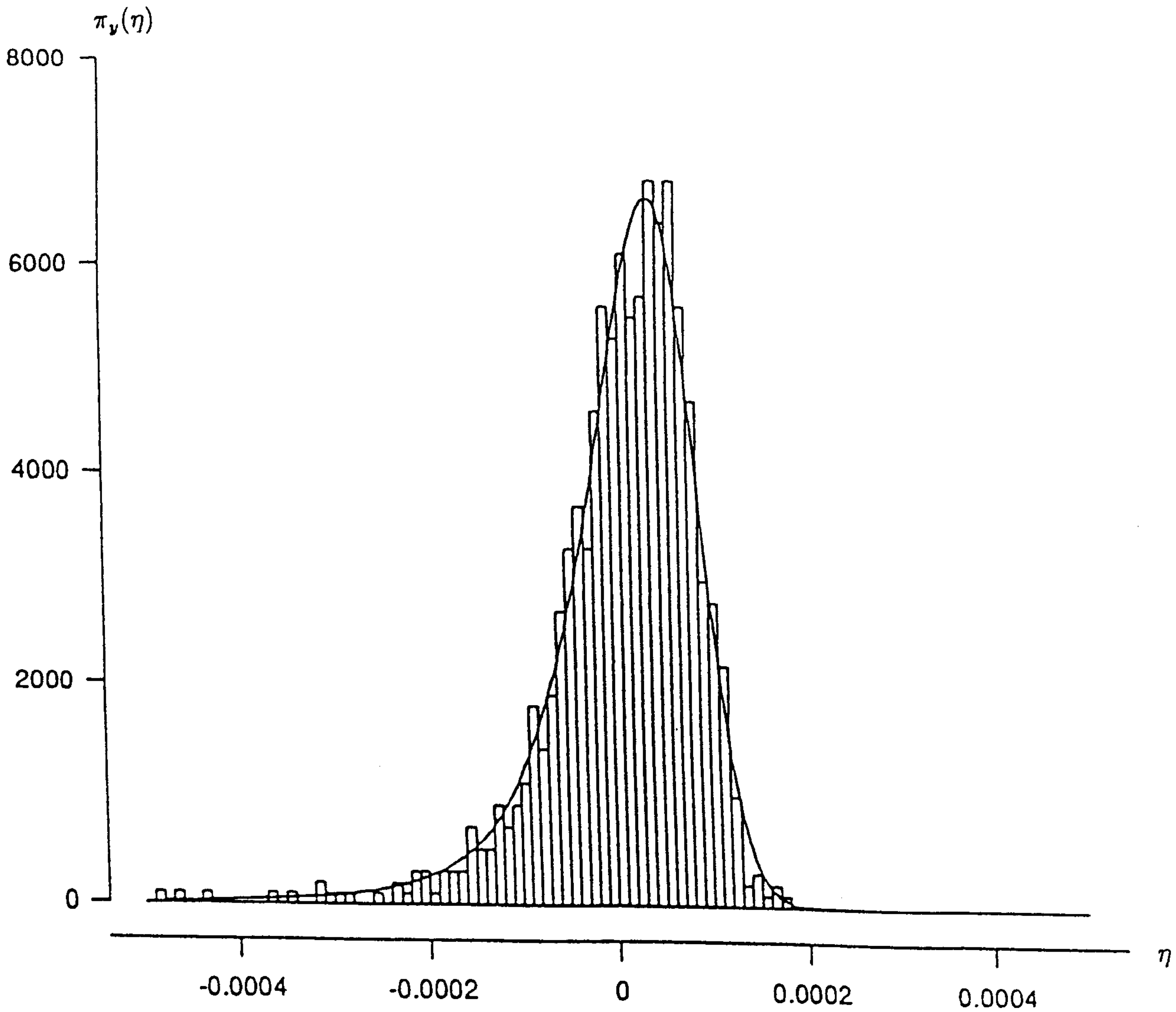

Figure 4. Marginal posterior density of quadratic coefficient
Solid curve: Laplacian T-approximation (virtually exact)
Histogram: Simulated by Gibbs sampler/Metropolis algorithm

References

Bates, D. M. and Watts, D. G. (1988). *Non-Linear Regression Analysis and Its Applications.* John Wiley, New York.

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

Carlin, B. P. and Gelfand, A. E. (1990). Approaches for empirical Bayes confidence intervals. *J. Amer. Statist. Assoc.* **85**, 105–114.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Hsu, J. S. J., Leonard, T. and Tsui, K.-W. (1991). Statistical inference for multiple choice tests. *Psychometrika* **56**, 327–348.

Leonard, T. (1982). Comment on the paper by M. Lejeune and G. D. Faulkenberry. *J. Amer. Statist. Assoc.* **77**, 657–658.

Leonard, T. and Novick, M. R. (1986). Bayesian full rank marginalization for two-way contingency tables. *J. Educational Statist.* **11**, 33–56.

Leonard, T., Hsu, J. S. J. and Tsui, K.-W. (1989). Bayesian marginal inference. *J. Amer. Statist. Assoc.* **84**, 1051–1058.

Leonard, T. and Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **20**, 1669–1696.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091.

Müller, P. (1991). A generic approach to posterior integration and Gibbs sampling. Technical Report #91–09, Department of Statistics, Purdue University.

O'Hagan, A. (1976). On posterior joint and marginal modes. *Biometrika* **63**, 329–333.

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling.* Marcel Dekker, New York.

Ritter, C. and Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the griddy Gibbs sampler. *J. Amer. Statist. Assoc.* **87**, 861–868.

Ritter, C., Bisgaard, S. and Bates, D. (1991). A comparison of approaches to inference for nonlinear models. *Proceedings of the 23rd Symposium of the Interface.* Interface Foundation of North America, Fairfax, VA, pp. 148–155.

Ritter, C. (1992). Modern inference in non linear least squares regression. Unpublished Ph.D. thesis. Department of Statistics, University of Wisconsin-Madison.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley, New York.

Sun, L. (1992). Bayesian estimation procedures for one and two-way hierarchical models. Unpublished Ph.D. thesis. Department of Statistics, University of Toronto.

Sun, L., Guttman, I. and Leonard, T. (1993). Bayesian inference for the one-way hierarchical model. Unpublished manuscript. Department of Statistics, University of Toronto.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* **82**, 528–550.

Tierney, L. (1991). Markov chains for exploring posterior distributions. Technical Report No. 560, School of Statistics, University of Minnesota.

Tierney, L., Kass, R. E. and Kadane, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76**, 425–433. Corrections (1991) **78**, 233–234.

Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.
Department of Statistics and Applied Probability, University of California, Santa Barbara, U.S.A.

(Received September 1991; accepted July 1993)