Generalized Laplacian Approximations in Bayesian Inference

John S. J. Hsu

The Canadian Journal of Statistics / La Revue Canadienne de Statistique, Vol. 23, No. 4. (Dec., 1995), pp. 399-410.

Stable URL:
http://links.jstor.org/sici?sici=0319-5724%28199512%2923%3A4%3C399%3AGLAIBI%3E2.0.CO%3B2-B

The Canadian Journal of Statistics / La Revue Canadienne de Statistique is currently published by Statistical Society of Canada.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/ssc.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Thu Aug 31 19:11:22 2006

The Canadian Journal of Statistics
Vol. 23, No. 4, 1995, Pages 399–410
La Revue Canadienne de Statistique

Generalized Laplacian approximations in Bayesian inference

John S.J. HSU

University of California, Santa Barbara

Key words and phrases: Marginalization, contingency table, quasiindependence model, Fisher-Behrens problem.
AMS 1991 subject classifications: Primary 62F15; secondary 62H17, 62J10.

ABSTRACT

This paper presents a new Laplacian approximation to the posterior density of $\eta = g(\boldsymbol{\theta})$. It has a simpler analytical form than that described by Leonard *et al.* (1989). The approximation derived by Leonard *et al.* requires a conditional information matrix $\mathbf{R}_\eta$ to be positive definite for every fixed η. However, in many cases, not all $\mathbf{R}_\eta$ are positive definite. In such cases, the computations of their approximations fail, since the approximation cannot be normalized. However, the new approximation may be modified so that the corresponding conditional information matrix can be made positive definite for every fixed η. In addition, a Bayesian procedure for contingency-table model checking is provided. An example of cross-classification between the educational level of a wife and fertility-planning status of couples is used for explanation. Various Laplacian approximations are computed and compared in this example and in an example of public school expenditures in the context of Bayesian analysis of the multiparameter Fisher-Behrens problem.

RÉSUMÉ

Cet article présente une nouvelle approximation laplacienne de la densité a posteriori de $\eta = g(\boldsymbol{\theta})$. Sa forme analytique est plus simple que celle décrite par Leonard *et al.* (1989). L'approximation de Leonard *et al.* exige qu'une matrice d'information conditionnelle $\mathbf{R}_\eta$ soit définie positive pour tout η fixé. Dans de nombreux cas, cependant, les $\mathbf{R}_\eta$ ne sont pas toutes définies positives. Le calcul de leur approximation échoue donc, car cette approximation est alors impossible à normaliser. En revanche, l'approximation proposée ici peut être modifiée de manière à ce que la matrice d'information conditionnelle soit définie positive pour tout η fixé. Une procédure bayésienne de vérification de modèles pour tableaux de fréquence est également présentée. Un exemple de tri-croisé entre le niveau de scolarité de femmes mariées et les moyens de contrôle des naissances privilégiés par leur couple sert à illustrer la méthodologie. Plusieurs approximations laplaciennes sont calculées et comparées dans cet exemple, ainsi que dans le contexte d'une approche bayésienne du problème de Fisher-Behrens multiparamétrique appliquée à l'analyse des dépenses d'une école du secteur public.

1. INTRODUCTION

Consider an $n \times 1$ vector $\mathbf{y}^{\mathrm{T}} = (y_1, \ldots, y_n)$ of observations with joint probability mass or probability density function $p(\mathbf{y}|\boldsymbol{\theta})$ given an unknown $p \times 1$ vector of parameters $\boldsymbol{\theta}^{\mathrm{T}} = (\theta_1, \ldots, \theta_p)$. Suppose that the parameter of interest, $\eta = g(\boldsymbol{\theta})$, is a function of $\boldsymbol{\theta}$. The marginal posterior density of η given $\mathbf{y}$, which requires a p-dimensional integration, is desired for Bayesian inference. For each η, we define the parameter subspace $\Theta_\eta \subset \mathbb{R}^p$, where

$$\Theta_\eta = \{\boldsymbol{\theta} : g(\boldsymbol{\theta}) = \eta, \text{ for all } \boldsymbol{\theta} \in \mathbb{R}^p\}$$

is the target space for the integration. Hence, it is essential to make the approximation as precise as possible for the $\boldsymbol{\theta}$ in the target space Θ_η. Leonard (1982) presented the Laplacian approximation for the marginal posterior density of η in a special case, when the parameter of interest, $\eta = g(\boldsymbol{\theta}) = \theta_k$, is the kth coordinate of $\boldsymbol{\theta}$. Tierney and Kadane (1986) discussed the asymptotic properties of such approximation. Leonard *et al.* (1989) and Tierney *et al.* (1989) both derived the Laplacian approximations for more general η but in different degrees and with different approaches. Some other related approximations are given by Wong and Li (1992) and Leonard *et al.* (1994).

Let $\pi(\boldsymbol{\theta})$ be the prior density of $\boldsymbol{\theta}$. Leonard *et al.* (1989) approximated the posterior density

$$\pi(\boldsymbol{\theta}\,|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$$

via a Taylor series expansion of $L(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}\,|\mathbf{y})$ about $\boldsymbol{\theta}_\eta$, where $\boldsymbol{\theta}_\eta$ conditionally maximizes $\pi(\boldsymbol{\theta}\,|\mathbf{y})$ given that $\eta = g(\boldsymbol{\theta})$. Neglecting cubic and higher-order terms in the expansion provides the approximation $\pi^*(\boldsymbol{\theta}\,|\mathbf{y})$ to $\pi(\boldsymbol{\theta}\,|\mathbf{y})$, where

$$\log \pi^*(\boldsymbol{\theta}\,|\mathbf{y}) = \log \pi_M(\eta|\mathbf{y}) + \mathbf{l}_\eta^{\mathsf{T}}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta) - \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta)^{\mathsf{T}}\mathbf{R}_\eta(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta), \qquad (1.1)$$

in which

$$\pi_M(\eta|\mathbf{y}) = \sup_{\boldsymbol{\theta}\in\Theta_\eta} \pi(\boldsymbol{\theta}\,|\mathbf{y}),$$

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

and $\mathbf{R}_\eta$, defined as

$$\mathbf{R}_\eta = -\left.\frac{\partial^2 \log \pi(\boldsymbol{\theta}\,|\mathbf{y})}{\partial\boldsymbol{\theta}\boldsymbol{\theta}^{\mathsf{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta},$$

is the posterior information matrix of $\pi(\boldsymbol{\theta}\,|\mathbf{y})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_\eta$. The function $\log \pi^*(\boldsymbol{\theta}\,|\mathbf{y})$ can also be expressed as

$$\log \pi^*(\boldsymbol{\theta}\,|\mathbf{y}) = \log \pi_M(\eta|\mathbf{y}) + \tfrac{1}{2}\,\mathbf{l}_\eta^{\mathsf{T}}\mathbf{R}_\eta^{-1}\mathbf{l}_\eta - \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta^*)^{\mathsf{T}}\mathbf{R}_\eta(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta^*) \qquad (1.2)$$

by completing the square in (1.1), where

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

The resulting posterior density $\pi^*(\eta|\mathbf{y})$ of η given $\mathbf{y}$ is then the density of $\eta = g(\boldsymbol{\theta})$, while $\boldsymbol{\theta}$ possesses the density $\pi^*(\boldsymbol{\theta}\,|\mathbf{y})$ defined in (1.2). That is,

$$\pi^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y}) \exp(\tfrac{1}{2}\,\mathbf{l}_\eta^{\mathsf{T}}\mathbf{R}_\eta^{-1}\mathbf{l}_\eta) \times \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\Omega_{\eta,\epsilon}} \exp\{-\tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta^*)^{\mathsf{T}}\mathbf{R}_\eta(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta^*)\}\, d\boldsymbol{\theta}, \qquad (1.3)$$

where the space

$$\Omega_{\eta,\epsilon} = \{\boldsymbol{\theta} : \eta \le g(\boldsymbol{\theta}) \le \eta + \epsilon\}$$

is defined for $\eta \in \mathbb{R}$ and $\epsilon > 0$. If $\mathbf{R}_\eta$ is positive definite, then $\pi^*(\eta|\mathbf{y})$ can be expressed as

$$\pi^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y})|\mathbf{R}_\eta|^{-\frac{1}{2}} \exp(\tfrac{1}{2}\,\mathbf{l}_\eta^{\mathsf{T}}\mathbf{R}_\eta^{-1}\mathbf{l}_\eta)\, f(\eta|\boldsymbol{\theta}_\eta^*, \mathbf{R}_\eta^{-1}), \qquad (1.4)$$

where the function $f(\eta|\boldsymbol{\mu}, \mathbf{C})$ is the density of $\eta = g(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ possesses a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. It has been shown in several numerical examples, by Leonard *et al.* (1989) and Hsu *et al.* (1991), that the approximation (1.4) possesses excellent numerical accuracy, when compared with the exact result.

In cases where the parameter of interest $\eta = g(\boldsymbol{\theta})$ is a twice differentiable function of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \Theta_\eta$, an alternative approximation is available. Note that for each η, the conditional maximum $\boldsymbol{\theta}_\eta$ satisfies

$$\left[\frac{\partial \log \pi(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta}} - \lambda_\eta \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta} = 0, \tag{1.5}$$

where λ_η is a Lagrange multiplier. Now consider a Taylor series expansion of

$$\bar{L}(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}|\mathbf{y}) - \lambda_\eta\{g(\boldsymbol{\theta}) - \eta\} = L(\boldsymbol{\theta}) - \lambda_\eta\{g(\boldsymbol{\theta}) - \eta\} \tag{1.6}$$

about $\boldsymbol{\theta} = \boldsymbol{\theta}_\eta$. The linear term will vanish, because of (1.5). Therefore, neglecting cubic and higher-order terms gives the second-order Taylor series approximation

$$\log \bar{\pi}^*(\boldsymbol{\theta}|\mathbf{y}) = \log \pi_M(\eta|\mathbf{y}) - \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta)^{\mathrm{T}}\bar{\mathbf{R}}_\eta(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta), \tag{1.7}$$

where

$$\bar{\mathbf{R}}_\eta = \mathbf{R}_\eta + \lambda_\eta \left.\frac{\partial^2 g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta}. \tag{1.8}$$

However, for $\boldsymbol{\theta} \in \Theta_\eta$, $\bar{L}(\boldsymbol{\theta})$ reduces to $\log \pi(\boldsymbol{\theta}|\mathbf{y})$. Consequently (1.7) also provides a second-order Taylor series approximation to $\log \pi(\boldsymbol{\theta}|\mathbf{y})$, for $\boldsymbol{\theta} \in \Theta_\eta$. Taking exponentials provides an approximation of a simpler form

$$\bar{\pi}^*(\boldsymbol{\theta}|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y}) \exp\{-\tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta)^{\mathrm{T}}\bar{\mathbf{R}}_\eta(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta)\}.$$

Consequently, whenever $\bar{\mathbf{R}}_\eta$ is positive definite, the marginal posterior density of $\eta = g(\boldsymbol{\theta})$ may be approximated by

$$\bar{\pi}^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y})|\bar{\mathbf{R}}_\eta|^{-\frac{1}{2}} f(\eta|\boldsymbol{\theta}_\eta, \bar{\mathbf{R}}_\eta^{-1}), \tag{1.9}$$

where the f contribution is defined as for (1.4). The construction in (1.6) has enabled us to greatly simplify the expression in (1.4), upon adjusting $\mathbf{R}_\eta$ to the matrix in (1.8).

The main result of Tierney *et al.* (1989) is to approximate the posterior density of $\pi(\eta|\mathbf{y})$ by

$$\bar{\pi}^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y})|\bar{\mathbf{R}}_\eta|^{-\frac{1}{2}}(\mathbf{b}_\eta^{\mathrm{T}}\bar{\mathbf{R}}_\eta^{-1}\mathbf{b}_\eta)^{-\frac{1}{2}}, \tag{1.10}$$

where

$$\mathbf{b}_\eta = \left.\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta}. \tag{1.11}$$

They derived Equation (1.10) by making a transformation $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta})$ such that $\boldsymbol{\phi}$ can be partitioned as $\boldsymbol{\phi}^{\mathrm{T}} = (g(\boldsymbol{\theta}), \boldsymbol{\phi}_2)$ and then applying a Laplacian approximation for the first coordinate $\eta = g(\boldsymbol{\theta})$. As an alternative derivation of Equation (1.10), we can further approximate the f contribution in Equation (1.9) by a normal density with mean $g(\boldsymbol{\theta}_\eta)$ and variance $\mathbf{b}_\eta^{\mathrm{T}}\bar{\mathbf{R}}_\eta^{-1}\mathbf{b}_\eta$; then Equation (1.10) follows. Therefore, Equation (1.9) provides a closer connection between (1.4) of Leonard *et al.* (1989) and (1.10) of Tierney *et al.* (1989). The approximation (1.10) is easier to use and possesses very good numerical accuracy if the function g is linear or nearly linear, whereas the approximations (1.4) and (1.9) possess better numerical accuracy for a more general functional form of g.

Note that a p-dimensional conditional maximization is needed for each η for all of the Laplacian approximations discussed above. This maximization procedure is generally not straightforward. However, it is far simpler than performing a p-dimensional numerical integration. Many different approaches may be used to solve conditional maximization problems. One approach is to fix a λ_η; then the conditional maximum $\boldsymbol{\theta}_\eta$ may be obtained by solving the equations in (1.5), utilizing numerical algorithms, such as standard Newton-Raphson procedures; $\eta = g(\boldsymbol{\theta}_\eta)$ may be calculated thereafter. The solution $\boldsymbol{\theta}_\eta$ is a local maximum if

(a) the quantity $\mathbf{z}^{\mathrm{T}}\bar{\mathbf{R}}_\eta\mathbf{z}$ is greater than zero for all nonzero $\mathbf{z}$ such that $\mathbf{z}^{\mathrm{T}}\mathbf{b}_\eta = 0$, where $\mathbf{b}_\eta$ is defined in (1.11), or

(b) the matrix $\mathbf{Q} = \mathbf{B}^{\mathrm{T}}\bar{\mathbf{R}}_\eta\mathbf{B}$ is positive definite, where $\mathbf{B}$ is any full-rank $p \times (p-1)$ matrix such that $\mathbf{b}_\eta^{\mathrm{T}}\mathbf{B} = 0$ and may be constructed in various ways. For example, $\mathbf{B}$ may be constructed as

$$\mathbf{B} = \begin{bmatrix} -b_{2\eta}/b_{1\eta} & -b_{3\eta}/b_{1\eta} & \cdots & -b_{p\eta}/b_{1\eta} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{1.12}$$

where $b_{i\eta}$ is the ith coordinate of $\mathbf{b}_\eta$ for $i = 1, \ldots, p$.

For the details of the above conditions, see for example Scales (1985). To find $\boldsymbol{\theta}_\eta$ for all necessary η, we may start with the posterior mode. The posterior mode may be obtained by setting $\lambda_\eta = 0$. In this case, because the function g is not involved, only an unconstrained maximization is needed, and the mode may be easily obtained. After the posterior mode has been found, we may use it as the initial values to approach the conditional maximum for a $\lambda_\eta > 0$, which is near zero. The posterior mode provides a good guess for this conditional maximum, since λ_η is near zero. We may move λ_η incrementally further away from its last point and use that previous maximum as the initial value for the current conditional maximization. The same process may be used for $\lambda_\eta < 0$. The process may be continued until all needed η's are found.

The following example presents a Bayesian procedure for contingency-table model checking. The Laplacian approximations discussed above will be compared and discussed.

2. AN EXAMPLE BASED ON A QUASIINDEPENDENCE MODEL

Following Leonard and Novick (1986) and Leonard *et al.* (1989), consider an $r \times s$ contingency table with cell counts y_{ij}, where $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$. Suppose that y_{ij} are independent and each has a Poisson distribution with means $\theta_{ij} > 0$. Also suppose that the θ_{ij} are *a priori* independent and gamma-distributed with respective means α_{ij}/β and variances α_{ij}/β^2. Therefore, the posteriors of θ_{ij} are independent and gamma-distributed with respective means $(\alpha_{ij}+y_{ij})/(\beta+1)$ and variances $(\alpha_{ij}+y_{ij})/(\beta+1)^2$. Hence the $\gamma_{ij} = \log\theta_{ij}$ possess joint posterior density

$$\pi(\boldsymbol{\gamma}|\mathbf{y}) \propto \exp\left(\sum_{i=1}^{r}\sum_{j=1}^{s}(y_{ij}+\alpha_{ij})\gamma_{ij} - (\beta+1)\sum_{i=1}^{r}\sum_{j=1}^{s}e^{\gamma_{ij}}\right),$$

which approaches the likelihood of $\boldsymbol{\gamma}$ as β and all of the α_{ij} tend to 0. This limiting situation will be assumed in the example below, and we will therefore analyze the table based on the likelihood function. Goodman (1964) described the following full-rank log-linear interaction model:

$$\log \theta_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \qquad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, s, \tag{2.1}$$

where μ is the overall effect, λ_i^A the ith row effect, λ_j^B the jth column effect, and λ_{ij}^{AB} the interaction effect for cell (i,j). Under regular side conditions, λ_{ij}^{AB} can be represented as a linear function of $\gamma_{11}, \gamma_{12}, \ldots, \gamma_{rs}$ as

$$\lambda_{ij}^{AB} = \gamma_{ij} - \gamma_{i\cdot} - \gamma_{\cdot j} + \gamma_{\cdot\cdot}$$

(with a dot denoting averaging with respect to the subscript). The concept of quasi-independence model was initiated by Goodman (1968). Commonly used models for analyzing contingency tables such as the extreme-ends model, four-corners model, and main-diagonal model are examples of the quasiindependence model. See Fienberg (1987), Fingleton (1984) and Agresti (1990) for details. Let S be the subset of $I = \{(i,j) : i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, s\}$ such that the row and column classifications are quasiindependent in S. Therefore, the quasiindependence model can be written as

$$\log \theta_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}\delta_{ij}, \qquad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, s, \tag{2.2}$$

where

$$\delta_{ij} = \begin{cases} 0 & \text{if} \quad (i,j) \in S, \\ 1 & \text{otherwise.} \end{cases}$$

The parameter

$$\eta = \frac{1}{m} \sum_{(i,j) \in S} (\gamma_{ij} - \gamma_{i\cdot} - \gamma_{\cdot j} + \gamma_{\cdot\cdot})^2, \tag{2.3}$$

where m is the number of elements in S, provides a measure of association for investigating whether the row and column classifications are quasiindependent for the subset S, by comparing the models (2.1) and (2.2). Note that the independence model also belongs to (2.2), where S includes all of the cells in the table. The inference of η provides an alternative to a goodness-of-fit test for the quasiindependence model. The posterior density of η given $\mathbf{y}$ can be fully described using Laplacian approximations. A Bayesian inference of the quasiindependence model is therefore permitted.

The data on the cross-classification between educational level of wife and fertility-planning status of couples discussed by Goodman and Kruskal (1954) are described in Table 1. Intuitively, wives with the highest education should tend to plan more effectively and those with the least education to plan less effectively. Therefore, a four-corners model may be suitable for the analysis. The subset S here is the set which includes all cells but the four corner cells $(1,1)$, $(1,4)$, $(3,1)$ and $(3,4)$. Curves (a), (b) and (c) in Figure 1 represent the Laplacian approximations (1.4), (1.9) and (1.10) respectively. The histogram (e) represents the exact results based upon 100,000 simulations. In each simulation, twelve θ_{ij} were randomly chosen from independent gamma distributions, each with mean y_{ij} and variance y_{ij}, and then a η was calculated as formulated in (2.3). It is not surprising that curve (a) of the approximation (1.4) is the most accurate one and curve (b) of (1.9) is more accurate than curve (c) of (1.10). While (1.4) is derived from $L(\boldsymbol{\theta})$, the original log-posterior density, (1.9) is derived from $\bar{L}(\boldsymbol{\theta})$ [$L(\boldsymbol{\theta})$ plus an extra term], and (1.10)

TABLE 1: Cross-classification between educational level of wife and fertility-planning status of couples.[a]

	B_1	B_2	B_3	B_4
A_1	102	35	68	34
A_2	191	80	215	122
A_3	110	90	168	223

[a] Key: A_1: 1 year of college or more; A_2; 3 or 4 years high school; A_3: less than 3 years high school. B_1 through B_4 represent the range of effective planning of number and spacing of children, with B_1 as the most effective, and B_4 as the least effective.

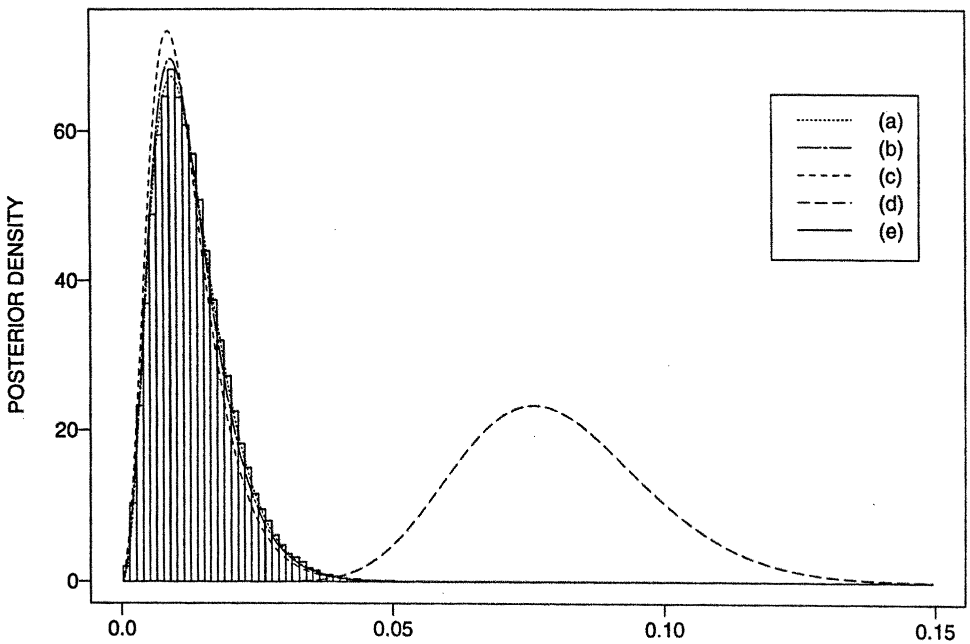

FIGURE 1: Marginal posterior density of η. (a), (b), (c), (e) Laplacian approximations (1.4), (1.9) and (1.10) and histogram (simulated from true distribution), respectively, when S includes all but the four corner cells in Table 1. (d) Laplacian approximation (1.4), when S includes all of the cells in Table 1.

is a further approximation to (1.9). Curve (d) represents the approximation (1.4) when the four corner cells were also included in S, and shows that the independence model is refuted and the four-corners (quasiindependence) model is not refuted.

Note that both $\mathbf{R}_\eta$ in (1.4) and $\tilde{\mathbf{R}}_\eta$ in (1.9) and (1.10) were positive definite for every η considered in this example. However, it is not always the case that those matrices are positive definite. If those matrices are not positive definite for some η, the corresponding approximation is not defined at that η, and the posterior density of η given $\mathbf{y}$ is not obtainable, due to the failure to calculate the proportionality constant. A more general approximation to remedy this deficiency is therefore necessary.

3. A MORE GENERAL RESULT

For a given ρ we define

$$\begin{aligned}\hat{L}(\boldsymbol{\theta}) &= \log \pi(\boldsymbol{\theta}\,|\mathbf{y}) - \lambda_\eta\{g(\boldsymbol{\theta}) - \eta\} - \tfrac{1}{2}\rho\{g(\boldsymbol{\theta}) - \eta\}^2 \\ &= \bar{L} - \tfrac{1}{2}\rho\{g(\boldsymbol{\theta}) - \eta\}^2.\end{aligned} \tag{3.1}$$

Analogous to the derivation of the approximations in Section 1, we take a Taylor series expansion of $\hat{L}(\boldsymbol{\theta})$ about the conditional maximum $\boldsymbol{\theta}_\eta$ which satisfies (1.5). Neglecting cubic and higher-order terms in the expansion provides the approximation $\hat{\pi}^*(\boldsymbol{\theta}\,|\mathbf{y})$ to $\pi(\boldsymbol{\theta}\,|\mathbf{y})$, where

$$\begin{aligned}\log \hat{\pi}^*(\boldsymbol{\theta}\,|\mathbf{y}) &= \log \pi_M(\eta|\mathbf{y}) - \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta)^{\mathrm{T}}\hat{\mathbf{R}}_{\eta,\rho}(\boldsymbol{\theta} - \boldsymbol{\theta}_\eta), \\ \hat{\mathbf{R}}_{\eta,\rho} &= -\frac{\partial^2\hat{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta} = \bar{\mathbf{R}}_\eta + \rho\mathbf{b}_\eta\mathbf{b}_\eta^{\mathrm{T}},\end{aligned} \tag{3.2}$$

and ρ is sufficiently large to make the matrices $\hat{\mathbf{R}}_{\eta,\rho}$ positive definite. Note that the first-order term in the expansion again vanished. That is,

$$\frac{\partial\hat{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta} = \left[\frac{\partial \log \pi(\boldsymbol{\theta}\,|\mathbf{y})}{\partial\boldsymbol{\theta}} - \lambda_\eta\frac{\partial g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \rho\{g(\boldsymbol{\theta}) - \eta\}\frac{\partial g(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta} = 0,$$

because of Equation (1.5) and the constraint $g(\boldsymbol{\theta}_\eta) = \eta$. Also note that, for $\boldsymbol{\theta} \in \Theta_\eta$, $\hat{L}(\theta)$ reduces to $\log \pi(\boldsymbol{\theta}\,|\mathbf{y})$. Consequently (3.2) also provides a second-order Taylor series approximation to $\log \pi(\boldsymbol{\theta}\,|\mathbf{y})$ for $\boldsymbol{\theta} \in \Theta_\eta$. The existence of such ρ is guaranteed for each given η, and will be discussed in the theorem below. The resulting approximation to the posterior density of η given $\mathbf{y}$ is therefore

$$\hat{\pi}^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y})|\hat{\mathbf{R}}_{\eta,\rho}|^{-\frac{1}{2}}f(\eta|\boldsymbol{\theta}_\eta, \hat{\mathbf{R}}_{\eta,\rho}^{-1}), \tag{3.3}$$

where the f contribution is defined as for (1.4).

The following lemma, taken from Avriel (1976), provides a useful tool to prove the theorem below.

LEMMA 3.1. *Let u and v be continuous real functions on a compact set $K \subset \mathbb{R}^p$ such that $v(\mathbf{z}) \geq 0$ for all $\mathbf{z} \in K$. Then $u(\mathbf{z}) > 0$ whenever $v(\mathbf{z}) = 0$ if and only if there exists a number c^* such that for all $c \geq c^*$ one has*

$$u(\mathbf{z}) + cv(\mathbf{z}) > 0 \qquad \textit{forall} \quad \mathbf{z} \in K.$$

THEOREM 3.1. *Suppose that the posterior density $\pi(\boldsymbol{\theta}\,|\mathbf{y})$ and the parameter of interest $\eta = g(\boldsymbol{\theta})$ are both twice differentiable near the conditional maximum $\boldsymbol{\theta}_\eta$ which maximizes $L(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}\,|\mathbf{y})$ subject to $g(\boldsymbol{\theta}) = \eta$. We further assume that $\boldsymbol{\theta}_\eta$ satisfies Equation (1.5). Then there exists a ρ^* such that $\hat{\mathbf{R}}_{\eta,\rho}$ is a positive definite matrix for all $\rho \geq \rho^*$.*

Proof. Suppose that $\boldsymbol{\theta}_\eta$ conditionally maximizes $\log \pi(\boldsymbol{\theta}\,|\mathbf{y})$ subject to $g(\boldsymbol{\theta}) = \eta$ and satisfies (1.5). Then for all nonzero $\mathbf{z}$ in $\mathbb{R}^p$, the $\bar{\mathbf{R}}_\eta$ defined in (1.8) satisfies the following condition (a) described in Section 1:

$$\mathbf{z}^{\mathrm{T}}\bar{\mathbf{R}}_\eta\mathbf{z} > 0 \qquad \text{for all nonzero } \mathbf{z} \text{ such that } \mathbf{z}^{\mathrm{T}}\mathbf{b}_\eta = 0.$$

Now for all nonzero $\mathbf{z} \in \mathbb{R}^p$, we consider

$$\mathbf{z}^{\mathsf{T}}\hat{\mathbf{R}}_{\eta,\rho}\mathbf{z} = \mathbf{z}^{\mathsf{T}}\bar{\mathbf{R}}_{\eta}\mathbf{z} + \rho\mathbf{z}^{\mathsf{T}}\mathbf{b}_{\eta}\mathbf{b}_{\eta}^{\mathsf{T}}\mathbf{z}.$$

The quantity $\mathbf{z}^{\mathsf{T}}\mathbf{b}_{\eta}\mathbf{b}_{\eta}^{\mathsf{T}}\mathbf{z}$ is nonnegative, and the term $\mathbf{z}^{\mathsf{T}}\bar{\mathbf{R}}_{\eta}\mathbf{z}$ is positive whenever $\mathbf{z}^{\mathsf{T}}\mathbf{b}_{\eta} = 0$ and hence $\mathbf{z}^{\mathsf{T}}\mathbf{b}_{\eta}\mathbf{b}_{\eta}^{\mathsf{T}}\mathbf{z} = 0$. The theorem follows from the lemma on letting $K = \{\mathbf{z} : \mathbf{z} \in \mathbb{R}^p, \mathbf{z}^{\mathsf{T}}\mathbf{z} = 1\}$, $u(\mathbf{z}) = \mathbf{z}^{\mathsf{T}}\bar{\mathbf{R}}_{\eta}\mathbf{z}$ and $v(\mathbf{z}) = \mathbf{z}^{\mathsf{T}}\mathbf{b}_{\eta}\mathbf{b}_{\eta}^{\mathsf{T}}\mathbf{z}$ for all unit vectors $\mathbf{z}$ and hence for all nonzero vectors $\mathbf{z}$. □

The existence of ρ^* is therefore guaranteed. We make the following comments regarding the use of the approximations (1.4), (1.9), (1.10) and (3.3).

(a) If $\eta = g(\boldsymbol{\theta})$ is a linear function of $\theta_1, \theta_2, \ldots, \theta_p$, then $\mathbf{R}_{\eta}$ and $\bar{\mathbf{R}}_{\eta}$ are identical, and so are the approximations (1.4), (1.9) and (1.10) if $\mathbf{R}_{\eta}$ and $\bar{\mathbf{R}}_{\eta}$ are positive definite. If $\mathbf{R}_{\eta}$ is not positive definite for some η, Equation (1.4) cannot be applied directly. However, the integral in (1.3) does exist in this case, and the approximation can be obtained by making a transformation $(\theta_1, \theta_2, \ldots, \theta_p) \longrightarrow (\eta, \theta_2, \ldots, \theta_p)$ in (1.3) before integration. The same treatment may be used for (1.9). The modified versions of (1.3) and (1.9) are analytically identical to the approximation (1.10). The approximation (3.3) is not actually needed when $g(\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$.

(b) The following equations are equivalent:

$$\pi(\eta|\mathbf{y}) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\Omega_{\eta,\epsilon}} \exp\{\log \pi(\boldsymbol{\theta}|\mathbf{y})\}\, d\boldsymbol{\theta} \tag{3.4}$$

$$= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\Omega_{\eta,\epsilon}} \exp[\log \pi(\boldsymbol{\theta}|\mathbf{y}) - \lambda_{\eta}\{g(\boldsymbol{\theta}) - \eta\}]\, d\boldsymbol{\theta} \tag{3.5}$$

$$= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\int_{\Omega_{\eta,\epsilon}} \exp[\log \pi(\boldsymbol{\theta}|\mathbf{y}) - \lambda_{\eta}\{g(\boldsymbol{\theta}) - \eta\} - \tfrac{1}{2}\rho\{g(\boldsymbol{\theta}) - \eta\}^2]\, d\boldsymbol{\theta}, \tag{3.6}$$

where $\Omega_{\eta,\epsilon}$ is defined as for (1.3) and ρ is a fixed constant. The approximations (1.4), (1.9) and (3.3) are the results when we neglect cubic and higher-order terms of the Taylor series expansions of the exponents $L(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}|\mathbf{y})$, $\bar{L}(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}|\mathbf{y}) - \lambda_{\eta}\{g(\boldsymbol{\theta}) - \eta\}$ and $\hat{L}(\boldsymbol{\theta}) = \log \pi(\boldsymbol{\theta}|\mathbf{y}) - \lambda_{\eta}\{g(\boldsymbol{\theta}) - \eta\} - \frac{1}{2}\rho\{g(\boldsymbol{\theta}) - \eta\}^2$ in (3.4), (3.5) and (3.6) respectively. Therefore, (1.4), (1.9) and (3.3) all have similar accuracy. Note that $L(\boldsymbol{\theta})$, $\bar{L}(\boldsymbol{\theta})$ and $\hat{L}(\boldsymbol{\theta})$ are identical for all $\boldsymbol{\theta}$ in the target space Θ_{η} but not identical in $\Omega_{\eta,\epsilon}$. Obviously, (1.4) is the most accurate approximation, (1.9) is more accurate than further approximated (1.10) (see also the example in Section 2), and the ρ in (3.3) should be chosen as small as possible.

(c) There is no simple analytical form for searching for the smallest ρ. However, in practice we may compute the conditional maximum $\boldsymbol{\theta}_{\eta}$ for every fixed η in a reasonably large range $[\eta_L, \eta_U]$ then simply try a ρ and increase it until $\hat{\mathbf{R}}_{\eta,\rho}$ is positive definite for every $\eta \in [\eta_L, \eta_U]$. The computation is straightforward and consumes only a little extra time.

4. AN EXAMPLE OF THE MULTIPARAMETER FISHER-BEHRENS PROBLEM

The public school expenditures, per pupil, per state in five regions (Northeast, Southeast, South Central, North Central, and Mountain Pacific) of the United States in 1977 were reported by Snedecor and Cochran (1989). The number of states, mean and variance for each of the five regions are summarized in Table 2.

TABLE 2: Public school expenditure data (in \$1000).

	Northeast	Southeast	South Central	North Central	Mountain Pacific
Number of states	10	7	9	11	11
Mean	1.763	1.330	1.179	1.563	1.507
Variance	0.1240	0.0335	0.0057	0.0448	0.0404

Consider the multiparameter Fisher-Behrens problem discussed by Leonard *et al.* (1994). Let y_{ij} denote the public school expenditure per pupil of the jth state within the ith region. Given $\theta_1, \ldots, \theta_5$ and $\phi_1, \ldots, \phi_5$, y_{ij} are independent and normally distributed, with respective means θ_i and variances ϕ_i ($i = 1, \ldots, 5, j = 1, \ldots, n_i$), where n_i is the number of states in the ith region. Assume that the θ_i and log variance $\alpha_i = \log \phi_i$ are *a priori* independent, and each is uniformly distributed over the real line. Then the posterior density of $\boldsymbol{\theta}$ given $\mathbf{y}$ is the product of five independent t-densities of form

$$\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) \propto \prod_{i=1}^{5} \{S_i^2 + n_i(\theta_i - \bar{y}_i)^2\}^{-n_i/2},$$

where $\bar{y}$ is the sample mean, and S_i^2 is the sum of squares within region i. We consider the Lagrangian

$$\bar{L} = \sum_{i=1}^{5} -\frac{n_i}{2} \log\{S_i^2 + n_i(\theta_i - \bar{y}_i)^2\} - \lambda_\eta\{g(\boldsymbol{\theta}) - \eta\}.$$

For each given λ_η, the conditional maximum $\boldsymbol{\theta}_\eta$ can be obtained by solving the equations

$$\frac{\partial \bar{L}}{\partial \theta_i} = -\frac{n_i^2(\theta_i - \bar{y}_i)}{S_i^2 + n_i(\theta_i - \bar{y}_i)^2} - \lambda_\eta \frac{\partial g}{\partial \theta_i} = 0, \qquad \text{for} \quad i = 1, \ldots, 5, \tag{4.1}$$

and $\eta = g(\boldsymbol{\theta}_\eta)$ can be calculated thereafter. The solution $\boldsymbol{\theta}_\eta$ is a local maximum if the matrix $\mathbf{Q} = \mathbf{B}^\mathsf{T}\bar{\mathbf{R}}_\eta\mathbf{B}$ is positive definite, where $\mathbf{B}$ is defined in (1.12). However, it is only guaranteed that $\mathbf{Q}$ is positive definite, or equivalently, $\mathbf{z}^\mathsf{T}\bar{\mathbf{R}}_\eta\mathbf{z} > 0$ for all nonzero $\mathbf{z}$ such that $\mathbf{z}^\mathsf{T}\mathbf{b}_\eta = 0$, and not for all nonzero $\mathbf{z} \in \mathbb{R}^5$. Therefore, it is not guaranteed that $\bar{\mathbf{R}}_\eta$ or $\mathbf{R}_\eta$ is positive definite. As an example, we consider a linear function $\eta = g(\boldsymbol{\theta}) = \sum_{i=1}^{5} a_i\theta_i$ for given constants $a_1, \ldots, a_5$. The ith diagonal element of $\bar{\mathbf{R}}_\eta$, which is equivalent to $\mathbf{R}_\eta$, is calculated as

$$r_{ii} = \frac{n_i^2\{S_i^2 - n_i(\theta_i - \bar{y}_i)^2\}}{\{S_i^2 + n_i(\theta_i - \bar{y}_i)^2\}^2}$$

and is nonpositive whenever

$$\frac{S_i^2}{n_i} \leq (\theta_{i\eta} - \bar{y}_i)^2. \tag{4.2}$$

Equation (4.2) holds when η is sufficiently far from $\hat{\eta} = \sum_{i=1}^{5} a_i\bar{y}_i$. In such a case, $\theta_{i\eta}$ is much greater than $\bar{y}_i$ for some i, while S_i^2/n_i remains constant for every η.

Two parameters of interest are considered in this paper. They are:

(a) $\eta_a = \theta_1 - \frac{1}{4}(\theta_2 + \theta_2 + \theta_3 + \theta_4)$, a linear function of $\boldsymbol{\theta}$, which represents the difference in public school expenditures between the states in the Northeast region and states in other regions;

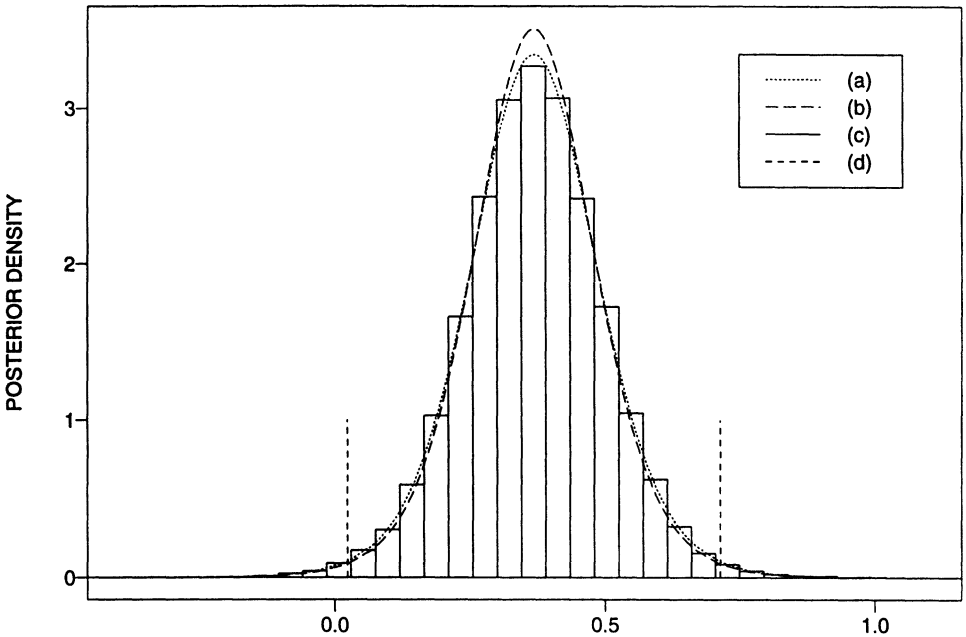

FIGURE 2: Marginal posterior density of η_a. (a) Laplacian approximations (1.10) and modified versions of approximations (1.4) and (1.9). (b) Single t-density with first four moments all correct. (c) Histogram, simulated from true distribution. (d) Lines indicating the lower and upper bounds to which the Laplacian approximations (1.4) and (1.9) can reach.

(b) $\eta_b = \sum_{i=1}^{5}(\theta_i - \bar{\theta})^2$, a nonlinear function of $\boldsymbol{\theta}$, which represents the corrected sum of squares in the context of one-way analysis of variance, where $\bar{\theta} = \frac{1}{5}(\theta_1+\theta_2+\theta_3+\theta_4+\theta_5)$.

Note that in Equation (4.1), for each given λ_η, five quadratics need to be solved for $g(\boldsymbol{\theta}) = \eta_a$ in (a) and five cubics for $g(\boldsymbol{\theta}) = \eta_b$ in (b). There is more than one solution to Equation (4.1). Caution should be taken so that the conditional maxima are used rather than any other solutions.

Curve (a) of Figure 2 describes the Laplacian approximation (1.10) and the modified versions of the approximations (1.4) and (1.9) [see comment (a) in Section 3] to the posterior density of η_a given $\mathbf{y}$. Curve (b) is a t-density with all the first four moments correct. Histogram (c) represents the exact posterior density based on 500,000 simulations. In each simulation, a $\eta_a = \theta_1 - (\theta_2 + \theta_3 + \theta_4 + \theta_5)/4$ was calculated based on a set of $\boldsymbol{\theta}$ generated from five independent t-distributions. Two lines (d) indicate the lower and upper bounds (0.023 and 0.714) which the unmodified Laplacian approximations (1.4) can reach. The approximated probability that $\eta_a < 0$ obtained for the approximation corresponding to curve (a) is 0.0054, which is remarkably close to the exact, computer-simulated probability, 0.0053. This minute probability indicates that the public school expenditures per pupil per state in the Northeast region were significantly higher than those in other states.

Curve (a) in Figure 3 describes our generalized Laplacian density (3.3) to the posterior density of η_b given $\mathbf{y}$ with searched smallest $\rho = 80$ [see comment (c) in Section 3]. Histogram (b) represents the true posterior density of η_b based on 500,000 simulations, each simulated from five independent t-distributions. Curve (a) approximates the true histogram extremely well, being centered about 0.2 and spread from 0.0 to 0.6; it

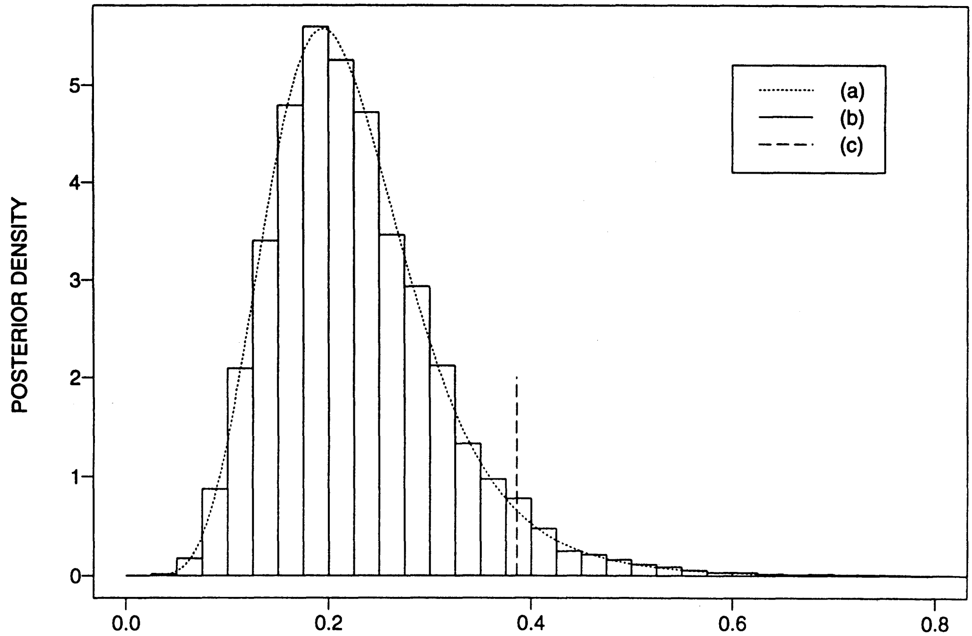

FIGURE 3: Marginal posterior density of η_b. (a) Generalized Laplacian approximation (3.3). (b) Histogram, simulated from true distribution. (c) Line indicating the upper bound to which the Laplacian approximation (1.4) can reach.

indicates that the public school expenditures per pupil per state were not the same for all of the five regions. Line (c) indicates the upper bound (about $\eta_b = 0.386$) below which the matrix $\bar{\mathbf{R}}_{\eta_b}$ is still positive definite and to which the approximation (1.4) can reach.

ACKNOWLEDGEMENTS

The author would like to thank Tom Leonard and a referee for helpful comments.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis.* Wiley, New York.

Avriel, M. (1976). *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Englewood Cliffs, N.J.

Fienberg, S.E. (1987). *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, Mass.

Fingleton, B. (1984). *Models of Category Counts.* Cambridge Univ. Press, Cambridge.

Goodman, L.A. (1964). Interactions in multidimensional contingency tables. *Ann. Math. Statist.*, 35, 632–647.

Goodman, L.A. (1968). The analysis of cross-classified data, independence, quasiindependence, and interaction in contingency tables with or without missing values. *J. Amer. Statist. Assoc.*, 63, 1091–1131.

Goodman, L.A., and Kruskal, W.H. (1954). Measures of association for cross classifications. *J. Amer. Statist. Assoc.*, 49, 732–764.

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

Leonard, T. (1982). Comment on "A simple predictive density function". *J. Amer. Statist. Assoc.*, 77, 657–658.

Leonard, T., Hsu, J.S.J., and Ritter, C. (1994). The Laplacian *t*-approximation in Bayesian inference. *Statist. Sinica*, 4, 127–142.

Leonard, T., Hsu, J.S.J., and Tsui, K. (1989). Bayesian marginal inference. *J. Amer. Statist. Assoc.*, 84, 1051–1057.

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

Scales, L.E. (1985). *Introduction to Non-linear Optimization.* Macmillan, London.

Snedecor, G.W., and Cochran, W.G. (1989). *Statistical Methods.* Eighth Edition. Iowa State University Press, Ames.

Tierney, L., and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities, *J. Amer. Statist. Assoc.*, 82, 528–549.

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

Wong, W.H., and Li, B. (1992). Laplace expansion for posterior densities of nonlinear functions of parameters. *Biometrika*, 79, 393–398.

Received 29 September 1994
Revised 12 July 1995
Accepted 28 July 1995

Department of Statistics and Applied Probability
University of California
Santa Barbara, California
93106, U.S.A.