UNIVERSIDADE DE SÃO PAULO

Approximate bayesian analysis for non-normal hierarchical classification model

JORGE ALBERTO ACHCAR

Nº 11

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Approximate bayesian analysis for non-normal hierarchical classification model

JORGE ALBERTO ACHCAR

Nº 11

N O T A S D O I C M S C
Série Estatística

São Carlos
jun. / 1994

Approximate Bayesian Analysis for Non-Normal Hierarchical Classification Models

Jorge Alberto Achcar
Universidade de São Paulo
ICMSC, C. Postal 668
13560-970, São Carlos, S.P., Brazil

Abstract

In this paper, we explore the use of Laplace's method (see for example, Tierney and Kadane, 1986, JASA, 81, 82-86) to find approximate posterior summaries of interest for hierarchical classification models with two variance components assuming a mixture of normal densities (see Tiao and Ali, 1971, Technometrics, 13(3), 635-650) for the random effect. We illustrate the proposed methodology considering a data set introduced by Tiao and Ali (1971).

Keywords: Laplace's method, hierarchical classification models, nonnormality for the random effect.

1 Introduction

One class of practical problems very important is given by variance components (see for example, Box and Tiao, 1973). When the total variance σ_T^2 of the observed yield is large, we have interest to reduce it, by discovering the relative importance of the various sources of variation.

When there is only two variance components, with J groups (batches) of K observations (analyses), consider the model,

$$y_{jk} = \mu + \epsilon_j + \epsilon_{jk} \;, \tag{1}$$

where $j = 1, \ldots, J; k = 1, \ldots, K$, and the random variables ϵ_j and ϵ_{jk} are independent with $E(\epsilon_j) = E(\epsilon_{jk}) = 0$, $var(\epsilon_j) = \sigma_2^2$ and $var(\epsilon_{jk}) = \sigma_1^2$. Thus the total variance σ_T^2 of the (j, k) th observation y_{jk} becomes $\sigma_1^2 + \sigma_2^2$.

Let us assume that the random variable ϵ_{jk} has a normal distribution $N(0, \sigma_1^2)$ and the random effect ϵ_j has a mixture of normal distributions (see Tiao and Ali, 1971), given by the density,

$$\begin{aligned} p(\epsilon_j|\sigma_2^2, \theta, \delta, \lambda) = (1-\theta) f_N(\epsilon_j| - \delta\theta\sigma, \sigma^2) + \\ +\theta f_N(\epsilon_j| - \delta\theta\sigma + \delta\sigma, \lambda^2\sigma^2) \;, \end{aligned} \tag{2}$$

where $-\infty < \epsilon_j < \infty, \sigma > 0, -\infty < \delta < \infty, \lambda \geq 1, 0 \leq \theta \leq 1$ and $f_N(x|p, q)$ denotes the density of a normal distribution with mean p and variance q.

The distribution (2) can be interpreted as saying that ϵ_j could come from one of two populations, a central model $N(-\delta\theta\sigma, \sigma^2)$ and an alternative model $N(-\delta\theta\sigma + \delta\sigma, \lambda^2\sigma^2)$ with probabilities $1 - \theta$ and θ, respectively.

This distribution has zero mean and expressions for the variance, skewness and kurtosis measures γ_1 and γ_2 given by

$$var(e_j) = \sigma_2^2 = \sigma^2 \{1 + \theta(\lambda^2 - 1) + \delta^2\theta(1-\theta)\}$$

$$\gamma_1 = \frac{\theta(1-\theta)\delta\{\delta^2(1-2\theta) + 3(\lambda^2-1)\}}{\{1 + \theta(\lambda^2-1) + \theta(1-\theta)\delta^2\}^{3/2}} \tag{3}$$

$$\gamma_2 = \frac{3\{1+\theta(\lambda^4-1)\} + 6\theta(1-\theta)\delta^2\{\theta + (1-\theta)\lambda^2\} + \delta^4\theta(1-\theta)\{\theta^3 + (1-\theta)^3\}}{\{1+\theta(\lambda^2-1) + \theta(1-\theta)\delta^2\}^2}$$

As a special case of distribution (2), consider $\theta = 0.05$, that is, the process is out of control 5% of time. Also, assuming $|\delta| = \lambda - 1$, that is, $\delta = \phi(\lambda - 1), \phi = -1, 1$ we have the model (see Tiao and Ali, 1971),

$$p(e_j|\sigma_2^2, \lambda, \phi) = 0.95 f_N (e_j| - 0.05\phi(\lambda - 1)\sigma, \sigma^2) + \\ + 0.05 f_N (e_j|0.95\phi(\lambda - 1)\sigma, \lambda^2\sigma^2) \ . \tag{4}$$

Observe that, if $\lambda = 1$, we have a normal distribution for e_j (usual assumption). If $\lambda > 1$, the distribution is symmetric but leptokurtic for $\phi = 0$, skewed to the right for $\phi = 1$, and skewed to the left for $\phi = -1$.

For the general model (2). Tiao and Ali (1971) show that the distribution is unimodal for all θ in $(0, 1)$ if and only if

$$\delta^2 < \frac{27(1 - \lambda^{-2})^2}{(1 - 2\lambda^{-2})(2 + \lambda^{-2} - \lambda^{-4}) + 2(1 - \lambda^{-2} + \lambda^{-4})^{3/2}} \ . \tag{5}$$

Usually, inferences for models with mixture of normal distributions have great computational difficulties (sse for example, Titterington, Smith and Makov, 1985).

For a Bayesian analysis, the statistician could decide by one among many existing strategies: the use of numerical methods (see for example, Naylor and Smith, 1982); the use of approximation methods for integrals (see for example, Tierney and Kadane, 1986); or the use of Monte Carlo procedures or Gibbs sampling (see for example, Kloek and Van Dijk, 1978; or Gelfand and Smith, 1990).

In this paper, we explore the use of approximate Bayesian methods for hierarchical classification models assuming the mixture of normal distributions with density (4) based on Laplace's method for approximation of integrals.

One of the great advantages of Laplace's method to solve the Bayesian integrals of interest it, is in terms of very small computational cost and easy implementation which do not require sophisticated computational skill.

We show in one example, considering a data set introduced by Tiao and Ali (1971), the feasibility of the proposed method.

2 A Joint Posterior Density for $\sigma_1^2, \sigma_2^2, \phi$ and λ

Considering the hierarchical classification model (1) with $\mu = 0$ assuming a normal distribution $N(0, \sigma_1^2)$ for the error e_{jk} and a mixture of normal distributions with density

(4) for the randon effect ϵ_j, the likelihood function for σ_1^2, σ_2^2, ϕ and λ is (see Tiao and Ali, 1971) given by

$$l(\sigma_1^2,\sigma_2^2,\phi,\lambda|y) \propto (\sigma_1^2)^{-v_1/2} exp\left\{-\frac{v_1 m_1}{2\sigma_1^2}\right\} \times \times \prod_{j=1}^{J} p(y_{j.}|\sigma_1^2,\sigma_2^2,\phi,\lambda) \ , \tag{6}$$

where $y_{j.} = K^{-1}\sum_k y_{jk}, v_1 = J(K-1), v_1 m_1 = S_1$,

$$S_1 = \sum_j \sum_k (y_{jk} - y_{j.})^2, p(y_{j.} / \sigma_1^2,\sigma_2^2,\phi,\lambda) = A_{1j} + A_{2j} \ ,$$

$$A_{1j} = \left(0.95 \ exp\left\{-\frac{[y_{j.} + 0.05\phi(\lambda-1)\sigma]^2}{2(\sigma^2+\sigma_1^2/K)}\right\}\right) \Big/ (\sigma^2 + \sigma_1^2/K)^{1/2} \ .$$

$$A_{2j} = \left(0.05 \ exp\left\{-\frac{[y_{j.} - 0.95\phi(\lambda-1)\sigma]^2}{2(\lambda^2\sigma^2+\sigma_1^2/K)}\right\}\right) \Big/ (\lambda^2\sigma^2 + \sigma_1^2/K)^{1/2} \ ,$$

$$\sigma_2^2 = c(\lambda,\phi)\sigma^2, \ and \ c(\lambda,\varphi) = 0.95 + 0.05\lambda^2 + 0.0475\phi^2(\lambda-1)^2,$$

and y is vector of observed data.

The joint prior density for $\sigma_1^2, \sigma_2^2, \phi$, and λ can be written in the form,

$$\pi\left(\sigma_1^2,\sigma_2^2,\phi,\lambda\right) = \pi\left(\sigma_1^2,\sigma_2^2|\phi,\lambda\right)\pi_0(\phi,\lambda) \ . \tag{7}$$

Considering a noninformative prior for σ_1^2 and σ_2^2 given (ϕ,λ), $\pi(\sigma_1^2,\sigma_2^2|\phi,\lambda) \propto (\sigma_1^2)^{-1}$ $(\sigma_1^2 + K\sigma_2^2)^{-1}$ (also considered by Tiao and Ali, 1971), the joint posterior density for $\sigma_1^2, \sigma_2^2, \phi$ and λ is given by

$$\pi\left(\sigma_1^2,\sigma_2^2,\phi,\lambda|y\right) \propto \pi_0(\phi,\lambda)\left(\sigma_1^2\right)^{-1}\left(\sigma_1 + K\sigma_2^2\right)^{-1} l\left(\sigma_1^2,\sigma_2^2,\phi,\lambda|y\right) \ , \tag{8}$$

where $\sigma_1^2 > 0, \sigma_2^2 > 0, \lambda \geq 1, \phi = -1,0,1; \pi_0(\phi,\lambda)$ is a joint prior density for ϕ and λ and $l(\sigma_1^2,\sigma_2^2,\phi,\lambda|y)$ is the likelihood function (6).

For the choice of prior $\pi_0(\phi,\lambda)$, we could assume prior independence for the parameters ϕ and λ. That is, $\pi_0(\phi,\lambda) = \pi_{01}(\phi)\pi_{02}(\lambda)$. In this case, one could consider many different choices for the prior densities for ϕ and λ. As a special case, consider $\pi_{01}(\phi) = 1/3$, for $\phi = -1,0,1$ and $\pi_{02}(\lambda) \propto 1/\lambda, \lambda \geq 1$. We also could consider an informative prior density for λ, since in many applications, we have prior opinion about λ.

3 Laplace's Method for Approximation of Integrals

Consider approximations for posterior moments of the form,

$$E(g(\psi)|y) = \frac{\int g(\psi)\pi(\psi)l(\psi|y)d\psi}{\int \pi(\psi)l(\psi|y)d\psi} \tag{9}$$

where $g(\psi)$ is a selected function of $\psi \in R^m$, $\pi(\psi)$ is a prior density, $l(\psi|y)$ is the likelihood function for ψ; and for posterior densities of the form,

$$\pi(\psi_1|y) = \int \pi\,(\psi_1, \psi_2|y)\, d\psi_2 \tag{10}$$

where $\pi(\psi_1, \psi_2|y)$ is the joint posterior density for $\psi = (\psi_1, \psi_2)$, $\psi_1 \in R^k$ and $\psi_2 \in R^{m-k}$.

The method of approximation for posterior moments introduced by Tierney and Kadane (1986) is based on Laplace's approximations for the integrals in both the numerator and denominator of (9). Laplace's method for approximation of integrals is used to solve integrals of the form

$$I = \int f(\psi) exp\left\{-nh(\psi)\right\} d\psi\ , \tag{11}$$

where $-nh(\psi)$ is a function having a maximum at $\hat{\psi}$ and which satisfies the usual regularity conditions.

To approximate integrals of the form (11), Laplace's method assumes an expansion of h and f in Taylor series about $\hat{\psi}$ (see Tierney and Kadane, 1986; or Tierney, Kass and Kadane, 1989a, 1989b).

With ψ one-dimensional, Laplace's approximation for I is given by

$$\hat{I} \cong \left(\frac{2\pi}{n}\right)^{1/2} \sigma_L f\left(\hat{\psi}\right)\ exp\left\{-nh\left(\hat{\psi}\right)\right\} \tag{12}$$

where $\sigma_L = \left\{h''\left(\hat{\psi}\right)\right\}^{-1/2}$.

In the m-dimensional case,

$$\hat{I} \cong (2\pi)^{m/2}\left\{det\left(n\ \Sigma_h^{-1}\left(\hat{\psi}\right)\right)\right\}^{-1/2} f\left(\hat{\psi}\right) exp\left\{-nh\left(\hat{\psi}\right)\right\}\ , \tag{13}$$

where $\Sigma_h^{-1}(\hat{\psi})$ is the Hessian matrix of h at $\hat{\psi}$, given by

$$\sum\nolimits^{-1}(\hat{\psi}) = \left(\frac{\partial^2 h}{\partial\psi_i\partial\psi_j}\right)\Big|_{\hat{\psi}} ; i,j = 1,\ldots,m .$$

To aproximate the posterior moment (9), we could consider $\pi(\psi)l(\psi|y) = exp\{-nh(\psi)\}$ for the numerator and denominator of (9), with f equals to g and 1, respectively. Other choice for f that gives more accurate approximations for (9) is given by $f = 1$ in both integrals in (9).

In the same way, we find Laplace's approximations for the marginal posterior density (10).

4 An Approximate Marginal Posterior Density for ϕ

The marginal posterior density for ϕ is (from (8)) given by

$$\pi(\phi|y) \propto \int\int\int \pi_0(\phi,\lambda)\left(\sigma_1^2\right)^{-1}\left(\sigma_1^2 + K\sigma_2^2\right)^{-1} l\left(\sigma_1^2,\sigma_2^2,\phi,\lambda|y\right) d\sigma_1^2 d\sigma_2^2 d\lambda . \qquad (14)$$

Considering

$$f_c(\sigma_1^2,\sigma_2^2,\lambda) = \pi_0(\phi,\lambda)(\sigma_1^2)^{-1}\ (\sigma_1^2 + K\sigma_2^2)^{-1}$$

and

$$-nh_\phi(\sigma_1^2,\sigma_2^2,\lambda) = ln l(\sigma_1^2,\sigma_2^2,\phi,\lambda|y)$$

in (11), a Laplace's approximation for the marginal posterior density for ϕ is given by

$$\hat{\pi}(\phi|y) \propto \frac{\pi_0\left(\phi,\hat{\lambda}\right)(\hat{\sigma}_1^2)^{-1}(\hat{\sigma}_1^2 + K\hat{\sigma}_2^2)^{-1} l\left(\hat{\sigma}_1^2,\hat{\sigma}_2^2,\phi,\hat{\lambda}|y\right)}{\left\{det\left(n\ \sum\nolimits_{h_\phi}^{-1}\left(\hat{\sigma}_1^2,\hat{\sigma}_2^2,\hat{\lambda}\right)\right)\right\}^{1/2}} \qquad (15)$$

where $\phi = -1,0,1$ and $(\hat{\sigma}_1^2,\hat{\sigma}_2^2,\hat{\lambda})$ maximizes $-nh_\phi(\sigma_1^2,\sigma_2^2,\lambda)$ for each value of ϕ.

Since we have difficulties to find the second derivatives of $-nh_\phi(\sigma_1^2,\sigma_2^2,\lambda)$, we could consider the use of numerical derivatives locally at $(\hat{\sigma}_1^2,\hat{\sigma}_2^2,\hat{\lambda})$ to find an approximate value for the determinant of the Hessian matrix in (15).

Also observe that the approximation (15) is valid for any choice of prior density for the parameters $\sigma_1^2,\sigma_2^2,\phi$ and λ.

5 An Approximate Marginal Posterior Density for λ with ϕ Known

The marginal posterior density for λ is given (from (8)) by

$$\pi(\lambda|y) = \sum_{\phi=-1,0,1} \int\int \pi\left(\sigma_1^2, \sigma_2^2, \phi, \lambda|y\right) d\sigma_1^2 d\sigma_2^2 . \tag{16}$$

Assuming ϕ known, we consider the prior density,

$$\begin{aligned} \pi\left(\sigma_1^2, \sigma_2^2, \lambda|\phi\right) &= \pi\left(\sigma_1^2, \sigma_2^2|\phi, \lambda\right) \pi_0(\lambda|\phi) \\ &\propto \left(\sigma_1^2\right)^{-1} \left(\sigma_1^2 + K\sigma_2^2\right)^{-1} \pi_0(\lambda|\phi) . \end{aligned} \tag{17}$$

where $\pi_0(\lambda|\phi)$ is a prior density for λ given ϕ.

With this prior, the joint posterior density for σ_1^2, σ_2^2 and λ is given by

$$\begin{aligned} \pi\left(\sigma_1^2, \sigma_2^2, \lambda|\phi, y\right) &\propto \pi_0(\lambda|\phi) \left(\sigma_1^2\right)^{-1} \left(\sigma_1^2 + K\sigma_2^2\right)^{-1} \times \\ &\times\ l\left(\sigma_1^2, \sigma_2^2, \lambda|\phi, y\right) . \end{aligned} \tag{18}$$

With the choice $f_\lambda(\sigma_1^2, \sigma_2^2) = \pi_0(\lambda|\phi)(\sigma_1^2)^{-1}(\sigma_1^2 + K\sigma_2^2)^{-1}$ and $-nh_\lambda(\sigma_1^2, \sigma_2^2) = ln l(\sigma_1^2, \sigma_2^2, \lambda|\phi, y)$ in (11), a Laplace's approximate marginal posterior density for λ is given by

$$\hat{\pi}(\lambda|\phi, y) \propto \frac{\pi_0(\lambda|\phi) \left(\hat{\sigma}_1^2\right)^{-1} \left(\hat{\sigma}_1^2 + K\hat{\sigma}_2^2\right)^{-1} l\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \lambda|\phi, y\right)}{\left\{det\left(n\ \Sigma_{h_\lambda}^{-1}\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2\right)\right)\right\}^{1/2}} \tag{19}$$

where $(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$ maximizes $-nh_\lambda(\sigma_1^2, \sigma_2^2)$ for each value of λ.

6 A Bayesian Analysis with ϕ and λ known

Assuming ϕ and λ known, the likelihood function for σ_1^2 and σ_2^2 is (see (6)) given by

$$l\left(\sigma_1^2, \sigma_2^2|\phi, \lambda, y\right) \propto \left(\sigma_1^2\right)^{-v1/2} exp\left\{-\frac{v_1 m_1}{2\sigma_1^2}\right\} \prod_{j=1}^{J} \left\{A_{1j} + A_{2j}\right\} , \tag{20}$$

5 An Approximate Marginal Posterior Density for λ with ϕ Known

The marginal posterior density for λ is given (from (8)) by

$$\pi(\lambda|y) = \sum_{\phi=-1,0,1} \int\int \pi\left(\sigma_1^2, \sigma_2^2, \phi, \lambda|y\right) d\sigma_1^2 d\sigma_2^2 \,. \tag{16}$$

Assuming ϕ known, we consider the prior density,

$$\begin{aligned} \pi\left(\sigma_1^2, \sigma_2^2, \lambda|\phi\right) &= \pi\left(\sigma_1^2, \sigma_2^2|\phi, \lambda\right) \pi_0(\lambda|\phi) \\ &\propto \left(\sigma_1^2\right)^{-1} \left(\sigma_1^2 + K\sigma_2^2\right)^{-1} \pi_0(\lambda|\phi) \,. \end{aligned} \tag{17}$$

where $\pi_0(\lambda|\phi)$ is a prior density for λ given ϕ.

With this prior, the joint posterior density for σ_1^2, σ_2^2 and λ is given by

$$\begin{aligned} \pi\left(\sigma_1^2, \sigma_2^2, \lambda|\phi, y\right) &\propto \pi_0(\lambda|\phi) \left(\sigma_1^2\right)^{-1} \left(\sigma_1^2 + K\sigma_2^2\right)^{-1} \times \\ &\times \; l\left(\sigma_1^2, \sigma_2^2, \lambda|\phi, y\right) \,. \end{aligned} \tag{18}$$

With the choice $f_\lambda(\sigma_1^2, \sigma_2^2) = \pi_0(\lambda|\phi)(\sigma_1^2)^{-1}(\sigma_1^2 + K\sigma_2^2)^{-1}$ and $-nh_\lambda(\sigma_1^2, \sigma_2^2) = \ln l(\sigma_1^2, \sigma_2^2, \lambda|\phi, y)$ in (11), a Laplace's approximate marginal posterior density for λ is given by

$$\hat{\pi}(\lambda|\phi, y) \propto \frac{\pi_0(\lambda|\phi) \left(\hat{\sigma}_1^2\right)^{-1} \left(\hat{\sigma}_1^2 + K\hat{\sigma}_2^2\right)^{-1} l\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \lambda|\phi, y\right)}{\left\{det\left(n \; \Sigma_{h_\lambda}^{-1} \left(\hat{\sigma}_1^2, \hat{\sigma}_2^2\right)\right)\right\}^{1/2}} \tag{19}$$

where $(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$ maximizes $-nh_\lambda(\sigma_1^2, \sigma_2^2)$ for each value of λ.

6 A Bayesian Analysis with ϕ and λ known

Assuming ϕ and λ known, the likelihood function for σ_1^2 and σ_2^2 is (see (6)) given by

$$l\left(\sigma_1^2, \sigma_2^2|\phi, \lambda, y\right) \propto \left(\sigma_1^2\right)^{-v1/2} exp\left\{-\frac{v_1 m_1}{2\sigma_1^2}\right\} \prod_{j=1}^{J} \left\{A_{1j} + A_{2j}\right\} \,, \tag{20}$$

where

$$A_{1j} = \left(0.95exp\left\{-\frac{(y_j + a_1\sigma)^2}{2(\sigma^2 + \sigma_1^2/K)}\right\}\right) \Big/ (\sigma^2 + \sigma_1^2/K)^{1/2} \; ,$$

$$A_{2j} = \left(0.05exp\left\{-\frac{(y_j - a_2\sigma)^2}{2(\lambda^2\sigma^2 + \sigma_1^2/K)}\right\}\right) \Big/ (\lambda^2\sigma^2 + \sigma_1^2/K)^{1/2} \; ,$$

$a_1 = 0.05\phi(\lambda - 1), a_2 = 0.95\phi(\lambda - 1), \sigma^2 = \sigma_2^2/b$ and

$b = 0.95 + 0.05\lambda^2 + 0.0475\phi^2(\lambda - 1)^2.$

The joint posterior density for σ_1^2 and σ_2^2 is given by

$$\pi\left(\sigma_1^2, \sigma_2^2|\phi, \lambda, y\right) \propto \pi\left(\sigma_1^2, \sigma_2^2|\phi, \lambda\right) l\left(\sigma_1^2, \sigma_2^2|\phi, \lambda, y\right) \; . \tag{21}$$

where $\sigma_1^2 > 0, \sigma_2^2 > 0; \pi(\sigma_1^2, \sigma_2^2|\phi, \lambda)$ is a prior density for (σ_1^2, σ_2^2) with ϕ and λ known and $l(\sigma_1^2, \sigma_2^2|\phi, \lambda, y)$ is the likelihood function (20).

6.1 An Approximate Marginal Posterior Density for σ_1^2

The marginal posterior density for σ_1^2 can be written (from (21)) in the form,

$$\pi\left(\sigma_1^2|\phi, \lambda, y\right) \propto \int f_{\sigma_1^2}\left(\sigma_2^2\right) e^{-nh_{\sigma_1^2}\left(\sigma_2^2\right)} d\sigma_2^2 \; . \tag{22}$$

where $f_{\sigma_1^2}(\sigma_2^2) = \pi(\sigma_1^2, \sigma_2^2|\phi, \lambda)$ (a prior density for σ_1^2 and σ_2^2 given ϕ and λ) and $-nh_{\sigma_1^2}(\sigma_2^2) = lnl(\sigma_1^2, \sigma_2^2|\phi, \lambda, y)$.

Therefore, an approximate marginal posterior density for σ_1^2 using Laplace's method is given by

$$\hat{\pi}\left(\sigma_1^2|\phi, \lambda, y\right) \propto \frac{\pi(\sigma_1^2, \hat{\sigma}_2^2|\phi, \lambda)\, l(\sigma_1^2, \hat{\sigma}_2^2|\phi, \lambda, y)}{\left\{-\dfrac{\partial^2 g(\sigma_1^2, \hat{\sigma}_2^2)}{\partial(\sigma_2^2)^2}\right\}^{1/2}} \tag{23}$$

where $\sigma_1^2 > 0, \hat{\sigma}_2^2$ maximizes $-nh_{\sigma_1^2}(\sigma_2^2)$ for each value of σ_1^2 and $g(\sigma_1^2, \sigma_2^2) = \sum_{j=1}^{J} ln(A_{1j} + A_{2j})$.

6.2 An Approximate Marginal Posterior Density for σ_2^2

The marginal posterior density for σ_2^2 can be written (from (21)) in the form,

$$\pi\left(\sigma_2^2|\phi,\lambda,y\right) \propto \int f_{\sigma_2^2}\left(\sigma_1^2\right) e^{-nh_{\sigma_2^2}(\sigma_1^2)} d\sigma_1^2 \tag{24}$$

where $f_{\sigma_2^2}(\sigma_1^2) = \pi(\sigma_1^2,\sigma_2^2|\phi,\lambda)$ and $-nh_{\sigma_2^2}(\sigma_1^2) = lnl(\sigma_1^2,\sigma_2^2|\phi,\lambda,y)$.

An approximate marginal posterior density for σ_2^2 is given by,

$$\hat{\pi}\left(\sigma_2^2|\phi.\lambda,y\right) \propto \frac{\pi\left(\hat{\sigma}_1^2.\sigma_2^2|\phi.\lambda,y\right)}{\left\{\frac{v_1 m_1}{\left(\hat{\sigma}_1^2\right)^3} - \frac{v_1}{2\left(\hat{\sigma}_1^2\right)^2} - \frac{\partial^2 g\left(\hat{\sigma}_1^2.\sigma_2^2\right)}{\partial\left(\sigma_1^2\right)^2}\right\}^{1/2}} \tag{25}$$

where $\sigma_2^2 > 0, \hat{\sigma}_1^2$ maximizes $-nh_{\sigma_2^2}(\sigma_1^2)$ for each value of σ_2^2.

6.3 Approximate Posterior Moments

We also could find Laplace's approximations for posterior moments of the form,

$$E\left\{m\left(\sigma_1^2,\sigma_2^2\right)|y\right\} = \frac{\int\int m\left(\sigma_1^2,\sigma_2^2\right)\pi\left(\sigma_1^2,\sigma_2^2|\phi.\lambda\right) l\left(\sigma_1^2,\sigma_2^2|\varphi,\lambda,y\right) d\sigma_1^2 d\sigma_2^2}{\int\int \pi\left(\sigma_1^2,\sigma_2^2|\phi,\lambda\right) l\left(\sigma_1^2.\sigma_2^2|\phi,\lambda,y\right) d\sigma_1^2 d\sigma_2^2} \tag{26}$$

As a special case, with the choice $f(\sigma_1^2,\sigma_2^2) = m(\sigma_1^2,\sigma_2^2)$ for the integral in the numerator of (26) and $f(\sigma_1^2,\sigma_2^2) = 1$ for the integral in the denominator of (26) (f given in (11)), we find a Laplace's approximation for (26), given by

$$\hat{E}\left\{m\left(\sigma_1^2,\sigma_2^2\right)|y\right\} \cong m\left(\tilde{\sigma}_1^2,\tilde{\sigma}_2^2\right)\left(1+0\left(n^{-1}\right)\right), \tag{27}$$

where $(\tilde{\sigma}_1^2,\tilde{\sigma}_2^2)$ is the mode of (21) (see Tierney, Kass and Kadane, 1989a).

As a special case, we have (from (27)), $E(\sigma_2^2/\sigma_1^2|y) \cong \tilde{\sigma}_2^2/\tilde{\sigma}_1^2, E(\sigma_1^2|y) \cong \tilde{\sigma}_1^2$ and $E(\sigma_2^2/y) \cong \tilde{\sigma}_2^2$.

More accurate approximations for (26) are given for other choices of $f(\sigma_1^2,\sigma_2^2)$ for the integrals in (26).

6.4 Approximate Predictive Distribution for a Future Group Average $Y_{(J+1)\cdot}$

Assuming ϕ and λ known, the predictive density for a future group average $y_{(J+1)\cdot}$ is given by

$$p\left(y_{(J+1)\cdot}|y\right) \propto \int\int p\left(y_{(J+1)\cdot}|\sigma_1^2,\sigma_2^2,\phi,\lambda\right)\pi\left(\sigma_1^2,\sigma_2^2|\phi,\lambda,y\right)d\sigma_1^2 d\sigma_2^2 \tag{28}$$

where $p(y_{(J+1)\cdot}|\sigma_1^2,\sigma_2^2,\phi,\lambda) = A_{1(J+1)} + A_{2(J+1)}$ and A_{ij} is defined in (20).

Considering $f(\sigma_1^2,\sigma_2^2) = \pi(\sigma_1^2,\sigma_2^2|\phi,\lambda)$ and $-nh(\sigma_1^2,\sigma_2^2) = lnp(y_{(J+1)\cdot}|\sigma_1^2,\sigma_2^2,\phi,\lambda) + lnl(\sigma_1^2,\sigma_2^2,\phi,\lambda|y)$ in (11), we also obtain an approximate predictive density for a future group average $y_{(J+1)\cdot}$ by

$$\hat{p}\left(y_{(J+1)\cdot}|y\right) \propto \frac{p\left(y_{(J+1)\cdot}|\hat{\sigma}_1^2,\hat{\sigma}_2^2,\phi,\lambda\right)\pi\left(\hat{\sigma}_1^2,\hat{\sigma}_2^2|\phi,\lambda\right)l\left(\hat{\sigma}_1^2,\hat{\sigma}_2^2|\phi,\lambda,y\right)}{\left\{det\left(n\ \Sigma_{*}^{-1}\left(\hat{\sigma}_1^2,\hat{\sigma}_2^2\right)\right)\right\}^{1/2}} \tag{29}$$

where $-\infty < y_{(J+1)\cdot} < \infty$; $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ maximizes $-nh(\sigma_1^2,\sigma_2^2)$ for each value of $y_{(J+1)\cdot}$.

7 A Bayesian Analysis Assuming $\lambda = 1$

Assuming $\lambda = 1$ in the mixture of normal densities (4), that is, a normal distribution for the random effect ϵ_j, the likelihood function for σ_1^2 and σ_2^2 (with $\mu = 0$ in the hierarchical classification model (1)), is given by,

$$\begin{aligned} l\left(\sigma_1^2,\sigma_2^2|y\right) &\propto \left(\sigma_1^2\right)^{-v_1/2}\ exp\left\{-\frac{v_1 m_1}{2\sigma_1^2}\right\}\left(\sigma_1^2+K\sigma_2^2\right)^{-J/2}\times \\ &\times exp\left\{-\frac{K\sum y_{j\cdot}^2}{2\left(\sigma_1^2+K\sigma_2^2\right)}\right\}. \end{aligned} \tag{30}$$

Considering f equals to the prior density $\pi(\sigma_1^2,\sigma_2^2)$ in (11), we find approximate marginal posterior for σ_1^2 and σ_2^2 given by

$$\hat{\pi}\left(\sigma_1^2|y\right) \propto \frac{\pi\left(\sigma_1^2,\hat{\sigma}_2^2\right)l\left(\sigma_1^2,\hat{\sigma}_2^2|y\right)}{\left\{\frac{K^3\sum y_{j\cdot}^2}{\left(\sigma_1^2+K\hat{\sigma}_2^2\right)^3} - \frac{JK^2}{2\left(\sigma_1^2+K\hat{\sigma}_2^2\right)^2}\right\}^{1/2}} \tag{31}$$

where $\sigma_1^2 > 0$ and $\hat{\sigma}_2^2$ maximizes $l(\sigma_1^2, \sigma_2^2|y)$ for each fixed value of σ_1^2, and ,

$$\tilde{\pi}\left(\sigma_2^2|y\right) \propto \frac{\pi\left(\hat{\sigma}_1^2, \sigma_2^2\right) l\left(\hat{\sigma}_1^2, \sigma_2^2|y\right)}{\left\{\frac{v_1 m_1}{\left(\hat{\sigma}_1^2\right)^3} - \frac{v_1}{2\left(\hat{\sigma}_1^2\right)^2} - \frac{J}{2\left(\hat{\sigma}_1^2+K\sigma_2^2\right)^2} + \frac{K\sum y_{j\cdot}^2}{\left(\hat{\sigma}_1^2+K\sigma_2^2\right)^3}\right\}^{1/2}} \tag{32}$$

where $\sigma_2^2 > 0$ and $\hat{\sigma}_1^2$ maximizes $l(\sigma_1^2, \sigma_2^2|y)$ for each fixed value of σ_2^2.

8 An Example

Consider the simulated data set of table 1, introduced by Tiao and Ali(1971) assuming the hierarchical classification model (1) with density (4) for the random effect ϵ_j where $\mu = 0, \sigma_1^2 = 1, \sigma_2^2 = 4, \phi = 1$ and $\lambda = 3$.

Group j	1	2	3	4	5	6
$y_{j\cdot}$	-3.682	-2.057	-1.780	-1.238	-0.797	-0.671
Group j	7	8	9	10	11	12
$y_{j\cdot}$	-0.646	-0.471	-0.436	-0.401	-0.378	0.000
Group j	13	14	15	16	17	18
$y_{j\cdot}$	0.112	0.791	0.923	1.571	1.712	4.223
Group j	19	20				
$y_{j\cdot}$	6.415	7.072				

Table 1 - Ordered Group Averages $y_{y\cdot}$ (simulated data with $\mu = 0, \sigma_1^2 = 1, \sigma_2^2 = 4, \phi = 1, \lambda = 3, J = 20$ and $K = 3$).

From the data of table 1, we have $J = 20, K = 3, v_1 = J(K-1) = 40, JKy_{\cdot\cdot} = \sum_j \sum_k y_{jk} = 0.5131, m_1 = 1.1525$ and $v_1 m_1 = S_1 = 46.1$ (see (6)). The logarithm of the likelihood function (6) is given by,

$$\begin{aligned} ln\, l\left(\sigma_1^2, \sigma_2^2, \phi, \lambda|y\right) &\propto -20 ln\left(\sigma_1^2\right) - \frac{23.05}{\sigma_1^2} + \\ &+ \sum_{j=1}^{20} ln\left(A_{1j} + A_{2j}\right) \ , \end{aligned} \tag{33}$$

where A_{1j} and $A_{2j}, j = 1, \ldots, J$ are given in (6). The maximum likelihood estimators for $\sigma_1^2, \sigma_2^2, \phi$ and λ (see table 2) are given by $\hat{\sigma}_1^2 = 1.15, \hat{\sigma}_2^2 = 3.13, \hat{\phi} = 1$ and $\hat{\lambda} = 3.76$.

ϕ	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\lambda}$	$ln l\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \phi, \hat{\lambda} \mid y\right)$
-1	1.15	9.80	3.12	-52.4795
0	1.15	3.90	4.10	-51.9312
1	1.15	3.13	3.76	-50.4169

Table 2 - Maximum Likelihood Estimators for σ_1^2, σ_2^2, and λ with ϕ known

Using numerical second derivatives of $\sum_{j=1}^{20} ln(A_{1j} + A_{2j})$ locally at the maximum likelihood estimators $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\phi}$ and $\hat{\lambda}$, we find the observed Fisher information matrix given $\phi = 1$,

$$I = \begin{pmatrix} 15.3031 & 0.1574 & -0.1574 \\ 0.1574 & 0.3862 & -0.0620 \\ -0.1574 & -0.0620 & 0.3672 \end{pmatrix}$$

Considering the usual normal limiting distribution for the maximun likelihood estimators $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\lambda})$ given $\phi = 1$, we find 95% confidence intervals for σ_1^2, σ_2^2 and λ given by $0.6471 < \sigma_1^2 < 1.6528, -0.0722 < \sigma_2^2 < 6.3322$ and $0.4753 < \lambda < 7.0447$.

In table 3, we have the approximate marginal posterior density (15) for ϕ considering different choices for the prior density $\pi_0(\phi, \lambda)$. We also considered the use of numerical derivatives to find the determinant of the Hessian matrix given in Laplace's approximation (15), for each value of ϕ. The mode of the marginal posterior densities for ϕ considering the different priors are all given by $\tilde{\phi} = 1$.

$\pi_0(\phi,\lambda)=\pi_{01}(\phi)\pi_{02}(\lambda)$	ϕ	$\hat{\pi}(\phi\|y)$
$\pi_{01}(\phi)=1/3, \phi=-1,0,1$	-1	0.2956
$\pi_{02}(\lambda)\propto$ constant	0	0.1711
	1	0.5333
$\pi_{01}(\phi)=1/3, \phi=-1,0,1$	-1	0.2541
$\pi_{02}(\lambda)\propto 1/\lambda$	0	0.1933
	1	0.5525
$\pi_{01}(\phi)=1/3, \phi=-1,0,1$	-1	0.3731
$\pi_{02}(\lambda)\propto exp\left\{-\frac{1}{2}(\lambda-3)^2\right\}$	0	0.1188
	1	0.5080
$\pi_{01}(\phi)=1/2, \phi=1$	-1	0.1637
$\pi_{01}(\phi)=1/4, \phi=0$		
	0	0.1245
$\pi_{01}(\phi)=1/4, \phi=-1$		
$\pi_{02}(\lambda)\propto 1/\lambda$	1	0.7118

Table 3 - Approximate Marginal Posterior Density (15) for ϕ Considering Different Priors $\pi_0(\phi,\lambda)$.

Assuming $\phi=1$ known, we have in figure 1, the plot of the approximate marginal posterior density (19) for λ considering some different choices of prior densities for σ_1^2, σ_2^2 and λ given $\phi=1$:

$$\pi_1(\sigma_1^2,\sigma_2^2,\lambda|\phi)\propto\lambda^{-1}(\sigma_1^2)^{-1}(\sigma_1^2+3\sigma_2^2)^{-1}.$$

$$\pi_2(\sigma_1^2,\sigma_2^2,\lambda|\phi)\propto exp\left\{-\tfrac{1}{8}(\lambda-3)^2\right\}(\sigma_1^2)^{-1}(\sigma_1^2+3\sigma_2^2)^{-1}, \tag{34}$$

$$\pi_3(\sigma_1^2,\sigma_2^2,\lambda|\phi)\propto exp\left\{-\tfrac{1}{2}(\lambda-3)^2\right\}(\sigma_1^2)^{-1}(\sigma_1^2+3\sigma_2^2)^{-1}.$$

We also used numerical second derivatives to find the determinant of the Hessian matrix in (19). The mode of the marginal posterior density for λ is given by $\tilde{\lambda}\cong 3.5$ considering all prior densities π_1,π_2 and π_3.

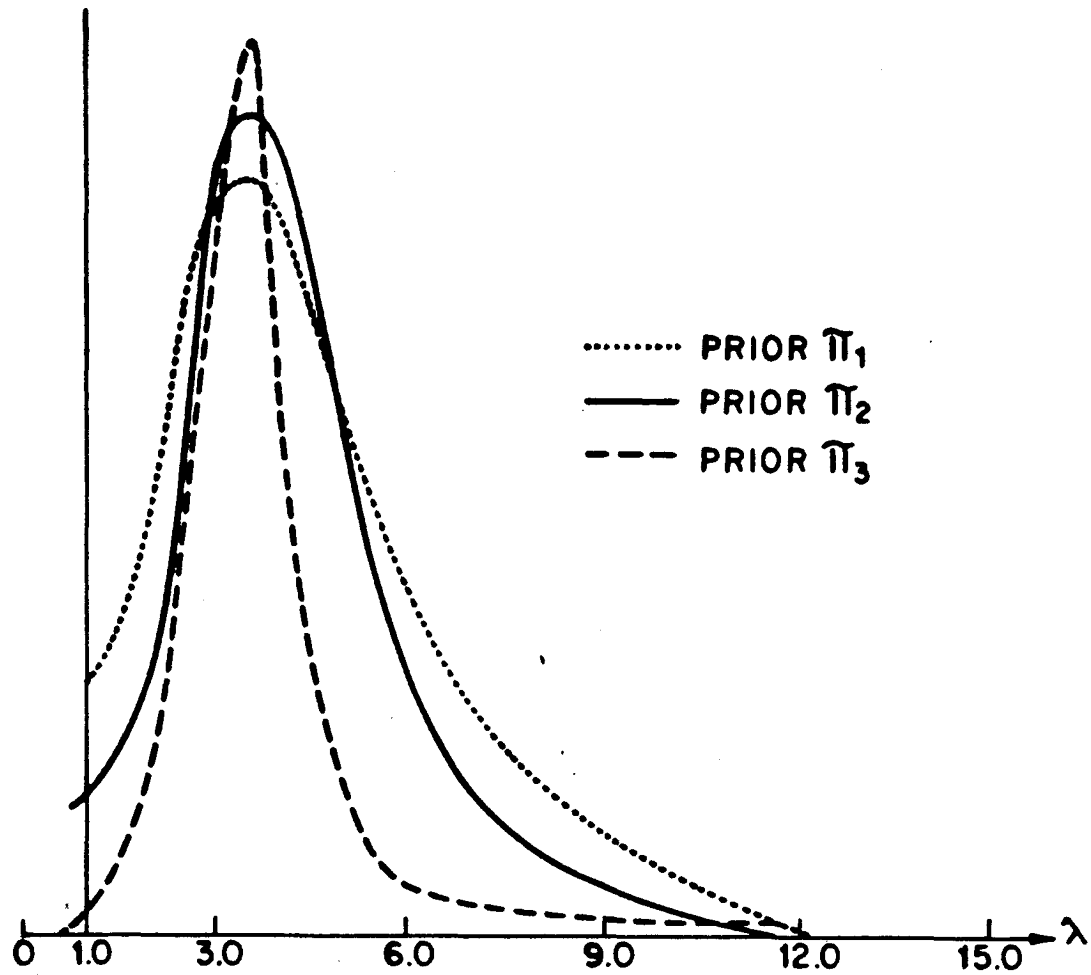

Figure 1 - Approximate Marginal Posterior Density for λ Assuming $\phi = 1$ Known

Assuming $\phi = 1$ and $\lambda = 3.5$ known, we have in figure 2, the plot of the approximate marginal posterior density (23) for σ_1^2 considering the prior density $\pi(\sigma_1^2, \sigma_2^2|\phi, \lambda) \propto (\sigma_1^2)^{-1}(\sigma_1^2 + 3\sigma_2^2)^{-1}, \sigma_1^2, \sigma_2^2 > 0$.

We also have in figure 2, the plot of the approximate marginal posterior density for σ_1^2 (31) assuming $\lambda = 1$, that is, a normal distribution for the random effect ϵ_j. We observe very close inference results for σ_1^2 considering normality or nonnormality of $\epsilon_j, j = 1, \ldots, J$.

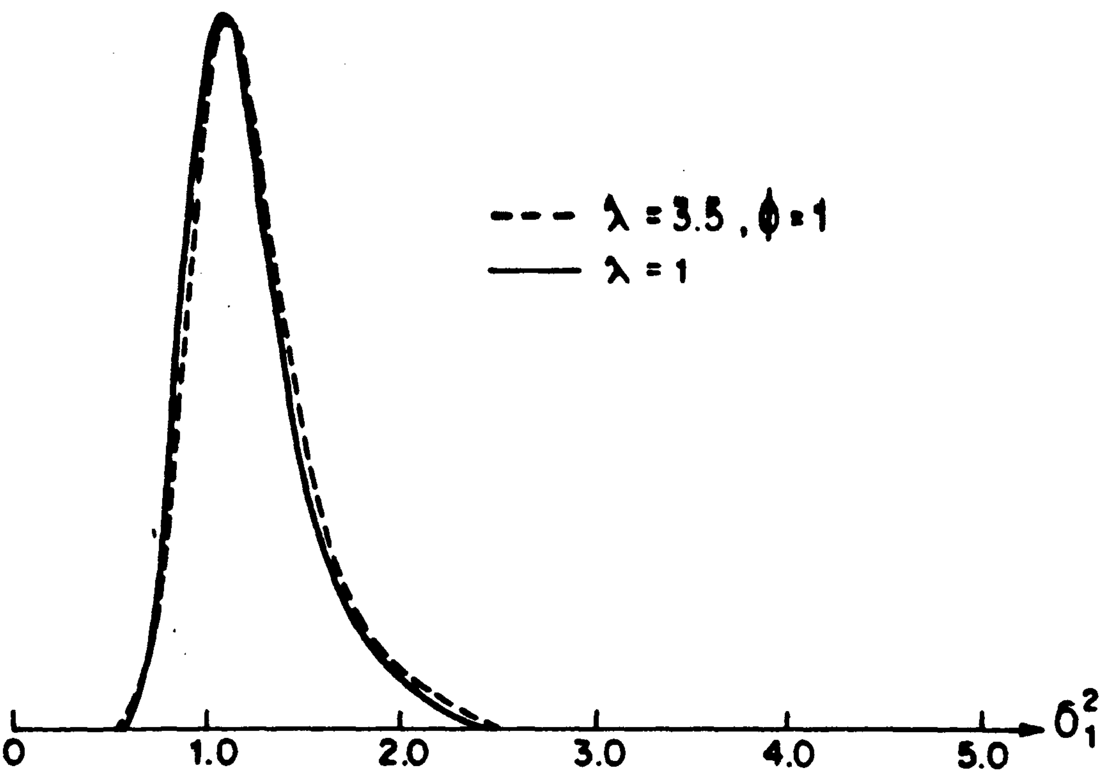

Figure 2 - Approximate Marginal Posterior Density for σ_1^2

Also, with $\phi = 1$ and $\lambda = 3.5$ known, and considering the same noninformative prior for σ_1^2 and σ_2^2 above, we have in figure 3 the plot of the approximate marginal posterior density (25) for σ_2^2. We also have in figure 3, the plot of the approximate marginal posterior density (32) for σ_2^2 considering $\lambda = 1$. In this case, we observe very different inference results for σ_2^2 considering normality or non-normality for the random effect $\epsilon_j, j = 1, \ldots, J$. It is important, to point out that we can use these Laplace's approximations considering any choice of prior opinion about σ_1^2 and σ_2^2.

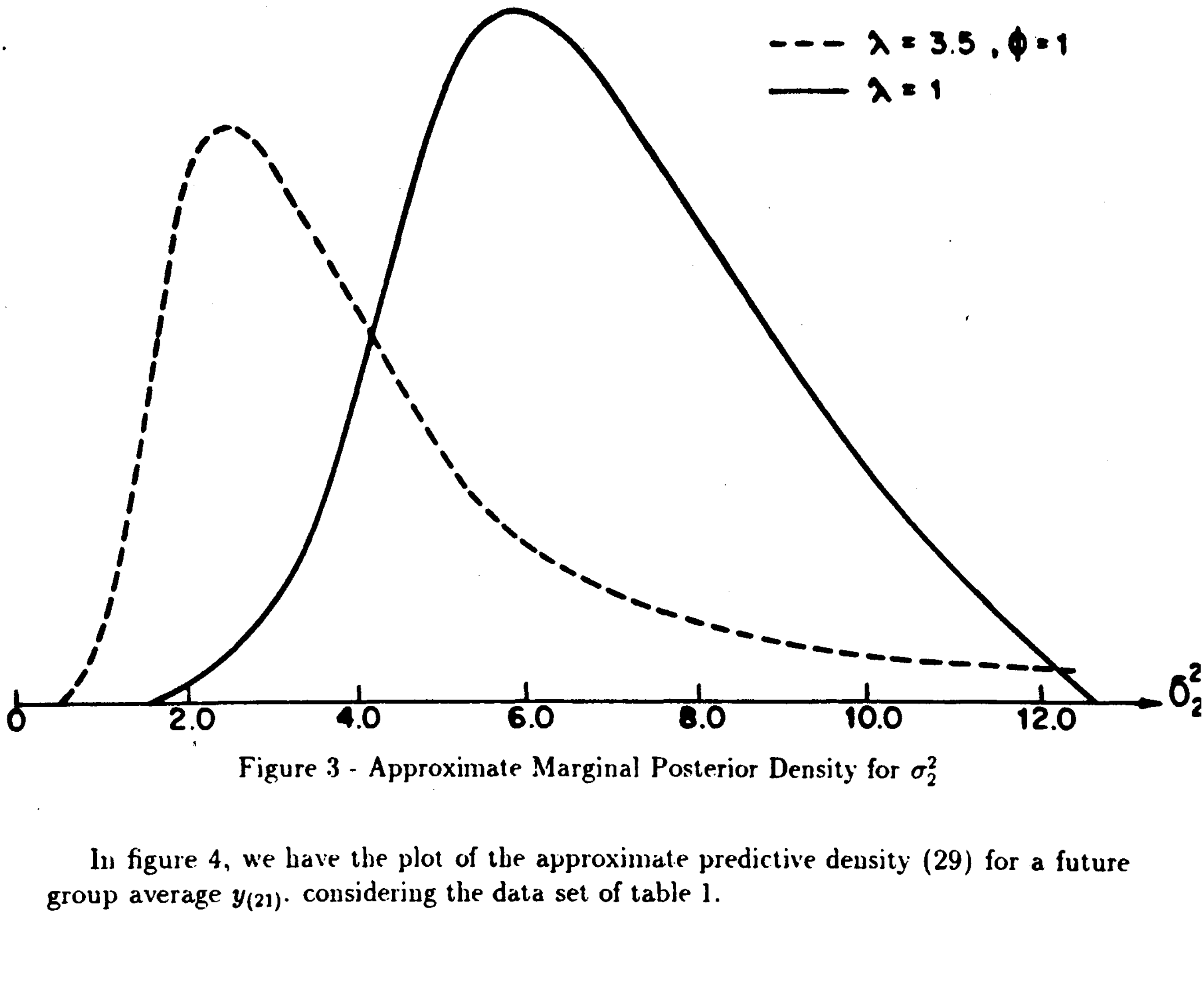

Figure 3 - Approximate Marginal Posterior Density for σ_2^2

In figure 4, we have the plot of the approximate predictive density (29) for a future group average $y_{(21)}$. considering the data set of table 1.

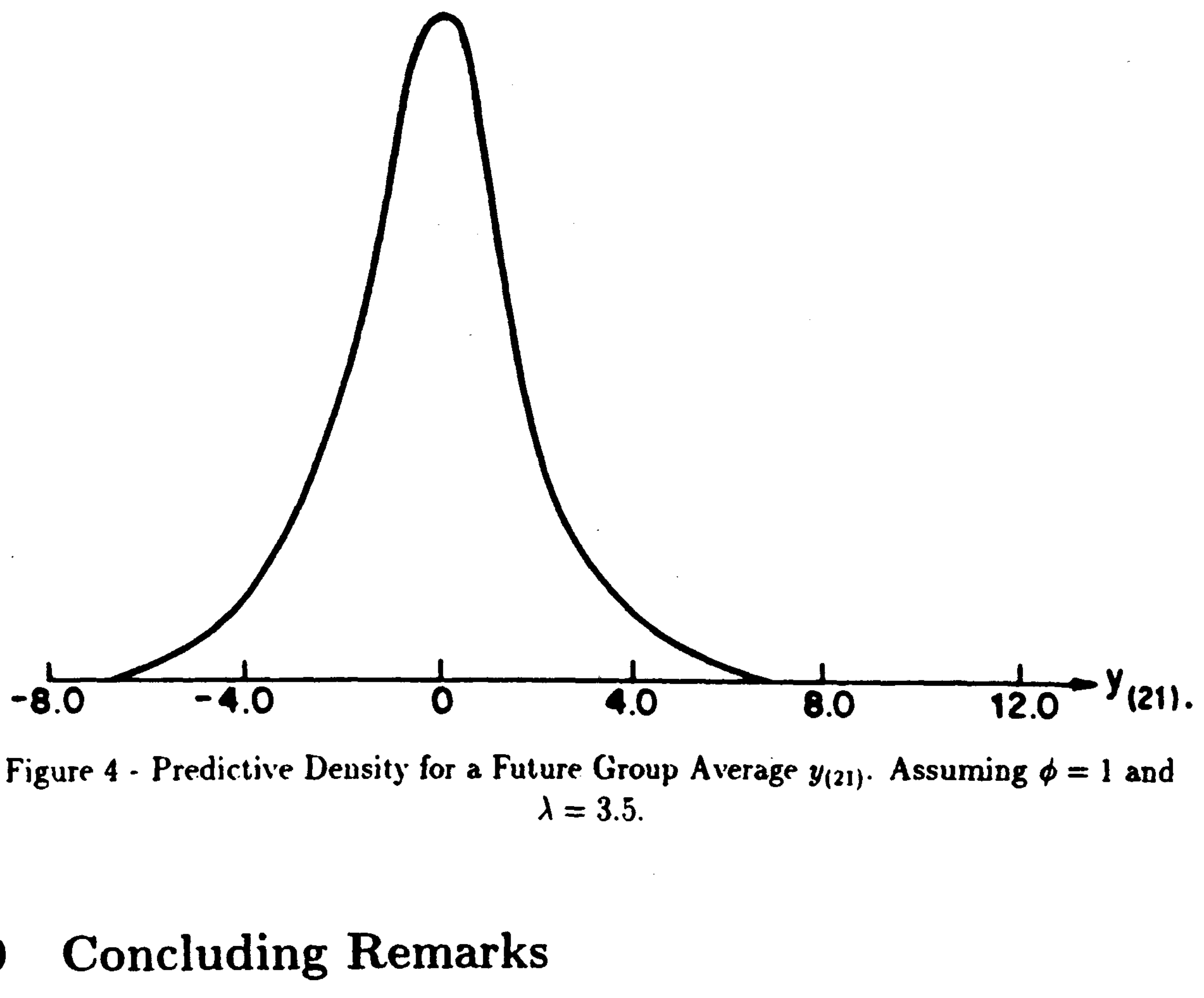

Figure 4 - Predictive Density for a Future Group Average $y_{(21)}$. Assuming $\phi = 1$ and $\lambda = 3.5$.

9 Concluding Remarks

The use of approximate Bayesian methods based on Laplace's approximations for integrals is a suitable alternative to present a Bayesian analysis for hierarchical classification models with two variance components assuming non-normality for the random effect. Similar results also could be obtained for three or more variance components models.

We also could consider θ unknown to present a similar approximate Bayesian analysis for two variance components models considering the mixture of normal distributions (2) for the random effect e_j in model (1). With this assumption, an approximate marginal posterior density for θ is given by,

$$\hat{\pi}(\theta|y) \propto \frac{\pi\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \theta, \hat{\delta}, \hat{\lambda}\right) l\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \theta, \hat{\delta}, \hat{\lambda}|y\right)}{\left\{det\left(n \ \Sigma_{h_\theta}^{-1}\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\delta}, \hat{\lambda}\right)\right)\right\}^{1/2}} \tag{35}$$

where $0 \leq \theta \leq 1, \pi(\sigma_1^2, \sigma_2^2, \theta, \delta, \lambda)$ is a prior density for the parameters and $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\delta}, \hat{\lambda})$ maximizes $-nh_\theta(\sigma_1^2, \sigma_2^2, \delta, \lambda) = lnl(\sigma_1^2, \sigma_2^2, \theta, \delta, \lambda|y)$, the logarithm of the likelihood function (6) with

$$p(y_j.|\sigma_1^2,\sigma_2^2,\theta,\delta,\lambda) = (1-\theta)f_N(y_j.|-\delta\theta\sigma,\sigma^2+\sigma_1^2/K) + \\ +\theta f_N(y_j.|-\delta\theta\sigma+\delta\sigma,\lambda^2\sigma^2+\sigma_1^2/K)\ , \tag{36}$$

where $\sigma_2^2 = \sigma^2\{1+\theta(\lambda^2-1)+\delta^2\theta(1-\theta)\}$ for each value of θ.

In the same way, we could find approximate marginal posterior densities for the other parameters.

The use of Laplace's method for approximation of integrals could be justified by comparing numerically integrated marginal posterior with Laplace's approximations. Assuming the data of table 1 with $\phi = 1$ and $\lambda = 3.5$ known, we have in figure 5 the graphs for the marginal posterior densities of σ_1^2 considering Laplace's method and a numerical procedure based on Gaussian quadrature (Gauss-Hermite with $n = 9$ roots of the polynomial equation of Hermite). We observe close results for both integration methods (see also table 4).

It is important to point out that the accuracy of the obtained approximations usually depend on good parametrization and data, especially for small sample sizes (see for example, Achcar and Smith, 1990).

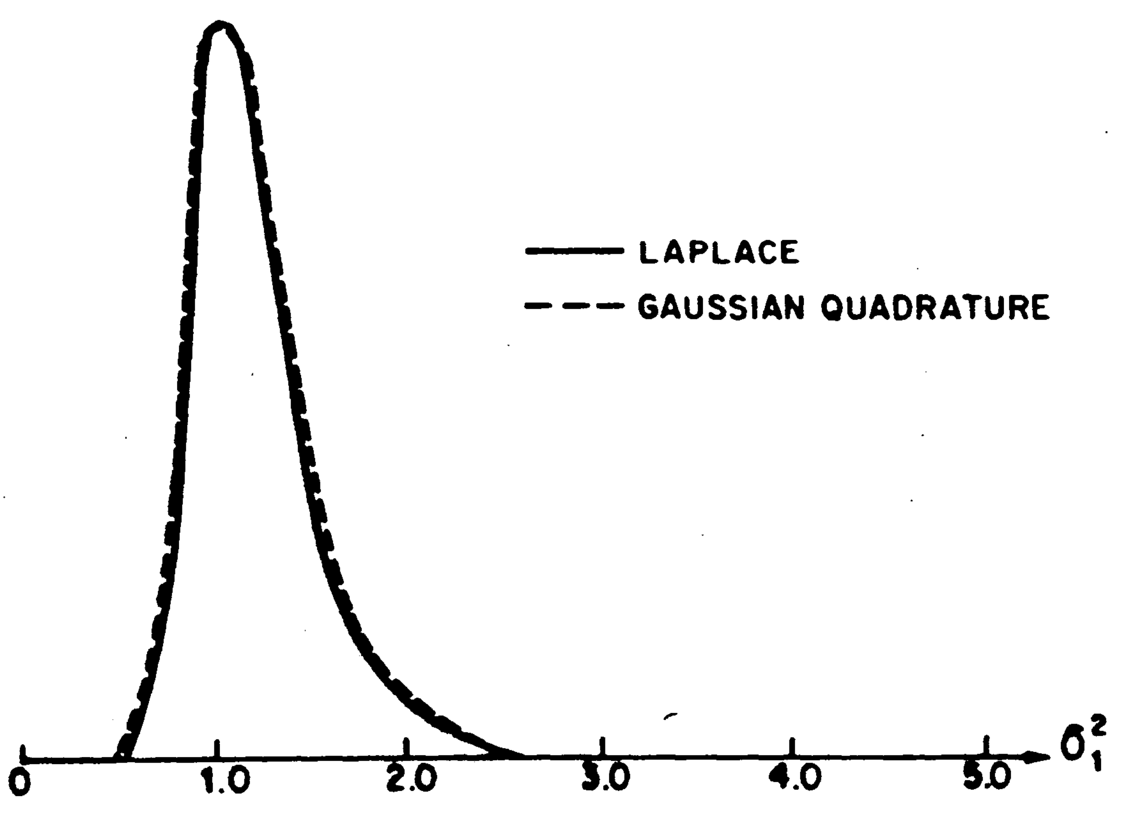

Figure 5 - Marginal Posterior Density for σ_1^2 with $\phi = 1$ and $\lambda = 3.5$ known.

σ_1^2	Gaussian Quadrature	Laplace
0.2	0.000000	0.000000
0.3	0.000000	0.000000
0.4	0.000000	0.000000
0.5	0.000048	0.000048
0.6	0.002273	0.002278
0.7	0.021630	0.021588
0.8	0.080391	0.080664
0.9	0.166467	0.167022
1.0	0.235640	0.235896
1.2	0.237517	0.239892
1.4	0.143797	0.142019
1.6	0.067360	0.066111
1.8	0.027712	0.027502
2.0	0.010713	0.010548
2.2	0.004042	0.004032
2.4	0.001521	0.001522
2.6	0.000578	0.000569
2.8	0.000223	0.000221
3.0	0.000088	0.000087
4.0	0.000001	0.000001

Table 4 - Marginal Posterior Density for σ_1^2 with $\phi = 1$ and $\lambda = 3.5$ known

References

Achcar,J.A.; Smith,A.F.M. (1990), "Aspects of reparametrization in approximate Bayesian inference", in Essays in honor of George A. Barnard, ed. J. Hodges, Amsterdam: North-Holland, 431-452.

Box,G.E.P.; Tiao,G.C. (1973), Bayesian inference in statistical analysis, New York: Addison-Wesley.

Gelfand,A.E.; Smith,A.F.M. (1990), "Sampling based approaches to calculating margi densities", Journal of the American Statistical Association, 85, 398-409.

Kloek,T.; Van Dijk,H.K. (1978), "Bayesian estimates of equation system parameters: an application of integration by Monte Carlo", Econometrica, 46, 1-19.

Naylor,J.C.; Smith,A.F.M. (1982), "Applications of a method for the efficient computation of posterior distributions", Applied Statistics, 31, 214-225.

Tiao,G.C.; Ali,M.M. (1971), "Effect of non-normality on inferences about variance components", Technometrics, 13(3), 635-650.

Tierney,L.; Kadane, J.B. (1986), "Accurate approximations for posterior moments and marginal densities", Journal of the American Statistical Association, 81, 82-86.

Tierney,L.; Kass, R.E.; Kadane,J.B. (1989a), "Approximate marginal densities for nonlinear functions", Biometrika, 76, 425-433.

Tierney,L.; Kass,R.E.; Kadane, J.B. (1989b), "Fully exponential Laplace approximations to expectations and variances of nonpositive functions", Journal of the American Statistical Association, 84, 710-716.

Titterington,D.M.; Smith,A.F.M.; Makov, U.E. (1985), Statistical analysis of finite mixtures distributions, New York: John Wiley & Sons.

NOTAS DO ICMSC

Serie Estatistica

Nº 010/94 RODRIGUES, J.; RODRIGUES, E.F. Bayesian estimation in the study of tampered Random variables

Nº 009/94 ACHCAR, J.A.; FOGO, J.C. An useful reparametrization for the reliability in the Weibull case

Nº 008/94 RODRIGUES, J.; LEITE, J.G. A note on Bayesian analysis in M/M/1 queues from confidence intervals

Nº 007/94 ACHCAR, J.A.; DAMACENO, V.L. An useful reparametrization for the survival function considering an exponential regression model

Nº 006/94 RODRIGUES, J. Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations

Nº 005/93 ACHCAR, J.A.; FOGO, J.C. Some useful reparametrizations for the reliability function considering an exponential regession model

Nº 004/93 LEME, E.C. A m.v.u. estimator v.s. a m.l. estimator: a way of comparing their precisions by using the Pareto distribution

Nº 003/93 ACHCAR, J.A. Some practical aspects of approximate bayesian inference

Nº 002/93 RODRIGUES, J.; BABA, M.Y. Bayesian estimation of a simple regression model with measurement errors

Nº 001/93 ACHCAR, J.A. Some aspects of reparametrization for statistical models