UNIVERSIDADE DE SÃO PAULO

Laplace's approximations for posterior expectations when the mode is not in the parameter space

JORGE ALBERTO ACHCAR

MARIA JOSÉ PEGORIN

Nº 15

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Laplace's approximations for posterior expectations when the mode is not in the parameter space

JORGE ALBERTO ACHCAR

MARIA JOSÉ PEGORIN

Nº 15

NOTAS DO ICMSC
Série Estatística

São Carlos
jan./95

Laplace's Approximations for Posterior Expectations when the Mode is not in the Parameter Space

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

Abstract

We propose a generalization of Laplace's approximations for posterior moments (see Tierney and Kadane, 1986) when the mode of the posterior density is not in the parameter space. We illustrate the proposed methodology in some selected applications, and we also show the importance of an appropriate parametrization to improve the accuracy of the obtained approximations, especially for small sample sizes.

Keywords: Laplace's method, mode of posterior density not in the parameter space, reparametrization.

1 Introduction

Laplace's method for approximation of integrals has been extensively used by Bayesian statisticians to approximate posterior expectations and marginal posterior densities of interest (see for example, Tierney and Kadane, 1986; or Tierney, Kass and Kadane, 1989). This approach assumes that the main contribution to the relevant integrals comes from a peak in the interior of the parameter space Θ.

In this paper, we present a simple way to generalize Laplace's approximations for posterior expectations when the posterior mode is not in the interior of the parameter space Θ. We also discuss some aspects of reparametrization to improve the accuracy of the approximate results.

The posterior expectation of a function $g(\theta)$ can be written as the ratio,

$$E(g(\theta)|\mathcal{D}) = \frac{\int g(\theta)L_n(\theta)\pi(\theta)d\theta}{\int L_n(\theta)\pi(\theta)d\theta} \tag{1}$$

where n is the sample size of a sample denoted by $\mathcal{D}$, g is a smooth function of θ, L_n is the likelihood function, and π is the prior density.

Considering the fully exponential case (see Tierney and Kadane, 1986), the basic integrals to be approximated in the numerator and denominator of (1), can be given by

$$I_1 = \int_{-\infty}^{\infty} \epsilon^{-nh(\theta)}d\theta \tag{2}$$

where h is a smooth function of θ.

When $-h$ has a maximum at a point $\widehat{\theta}$ in the interior of the parameter space, and considering first the case in which θ is one-dimensional, Laplace's method gives the approximation,

$$\widehat{I}_1 \cong \left(\frac{2\pi}{n}\right)^{1/2} \widehat{v}\ \epsilon^{-nh(\widehat{\theta})} \tag{3}$$

where $\widehat{v} = \{h''\left(\widehat{\theta}\right)\}^{-1/2}$.

In the multiparameter case, with $\theta \in R^m$, we have

$$\widehat{I}_1 \cong (2\pi)^{m/2}\left\{det\left(nD^2h\left(\widehat{\theta}\right)\right)\right\}^{-1/2} \epsilon^{-nh(\widehat{\theta})} \tag{4}$$

where $\hat{\theta}$ maximizes $-h$ and $D^2h(\theta)$ is the Hessian matrix of h evaluated at $\hat{\theta}$.

2 Laplace's Approximations when $\hat{\theta}$ is not in the Interior of the Parameter Space

Assuming θ one-dimensional and a parameter space given by $\Theta = \{\theta; -\infty < \theta < a\}$ where a is a constant, consider the approximation for the integral,

$$I_2 = \int_{-\infty}^{a} \epsilon^{-nh(\theta)} d\theta \tag{5}$$

where the maximum $\hat{\theta}$ is not in the interior of the parameter space Θ.

Expanding h in a Taylor series about the maximum $\hat{\theta}$ of $-h$, we have,

$$\hat{I}_2 \cong \epsilon^{-nh(\hat{\theta})} \int_{-\infty}^{a} \epsilon^{-\frac{nh''(\hat{\theta})}{2}(\theta-\hat{\theta})} d\theta \ . \tag{6}$$

Since,

$$\int_{-\infty}^{a} \epsilon^{-\frac{1}{2\delta^2}(x-\mu)^2} dx = \sqrt{2\pi}\delta \ \Phi\left(\frac{a-\mu}{\delta}\right) \ ,$$

where Φ denotes the distribution function of a standard normal distribution $N(0,1)$, we have the approximation.

$$\hat{I}_2 \cong \left(\frac{2\pi}{n}\right)^{1/2} \hat{v}\, e^{-nh(\hat{\theta})} \Phi\left(\frac{a-\hat{\theta}}{\hat{v}}\right) \tag{7}$$

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

Observe that if $(a-\hat{\theta})/\hat{v} > 3$, Laplace's approximation (7) reduces to (3), since $\Phi(\frac{a-\hat{\theta}}{\hat{v}}) \cong 1$. Also observe that if $a = \hat{\theta}$, that is, the maximum of $-h$ is at the boundary of the parameter space, $\Phi(\frac{a-\hat{\theta}}{\hat{v}}) = \Phi(0) = 1/2$, and Laplace's approximation for I_2 is given by,

$$\hat{I}_2 \cong \left(\frac{\pi}{2n}\right)^{1/2} \hat{v}\, \epsilon^{-nh(\hat{\theta})}. \tag{8}$$

For the two-parameter case, we have,

$$
\begin{aligned}
I_3 &= \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \epsilon^{-nh(\theta_1,\theta_2)} d\theta_1 d\theta_2 \\
&\cong (2\pi)\, \epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \left\{ det\left(nD^2h\left(\hat{\theta}_1,\hat{\theta}_2\right)\right)\right\}^{-1/2} F(a_1,a_2)
\end{aligned}
\tag{9}
$$

where $\left(\hat{\theta}_1,\hat{\theta}_2\right)$ maximizes $-h$ and $F(a_1,a_2)$ is the distribution function of a bivariate normal distribution with mean $\left(\hat{\theta}_1,\hat{\theta}_2\right)$ and variance-covariance matrix $\left\{nD^2h\left(\hat{\theta}_1,\hat{\theta}_2\right)\right\}^{-1}$.

If θ_1 and θ_2 are orthogonal parameters with $n\partial^2 h\left(\hat{\theta}_1,\hat{\theta}_2\right)/\partial\theta_1\partial\theta_2 = 0$ (see for example, Cox and Reid, 1987), Laplace's approximation (9) is reduced to,

$$
\hat{I}_3 \cong (2\pi)\, \epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \hat{v}_1\, \hat{v}_2 \Phi\left(\frac{a_1-\hat{\theta}_1}{\hat{v}_1}\right)\, \Phi\left(\frac{a_2-\hat{\theta}_2}{\hat{v}_2}\right) \tag{10}
$$

where $\hat{v}_1 = \left\{\frac{n\partial^2 h((\hat{\theta}_1,\hat{\theta}_2)}{\partial\theta_1^2}\right\}^{-1/2}$ and $\hat{v}_2 = \left\{n\, \frac{\partial^2 h(\hat{\theta}_1,\hat{\theta}_2)}{\partial\theta_2^2}\right\}^{-1/2}$.

Similar results are given for other cases; for example, considering $-\infty < \theta_1 < a_1$ and $-\infty < \theta_2 < \infty$, we have,

$$
\begin{aligned}
I_4 &= \int_{-\infty}^{a_1} \int_{-\infty}^{\infty} \epsilon^{-nh(\theta_1,\theta_2)} d\theta_1 d\theta_2 \\
&\cong (2\pi)\epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \left\{ det\left(nD^2h\left(\hat{\theta}_1,\hat{\theta}_2\right)\right)\right\}^{-1/2} \Phi\left(\frac{a_1-\hat{\theta}_1}{\hat{v}_1}\right),
\end{aligned}
\tag{11}
$$

and considering $a_1 < \theta_1 < \infty, -\infty < \theta_2 < \infty$, we have,

$$
\begin{aligned}
I_5 &= \int_{a_1}^{\infty} \int_{-\infty}^{\infty} \epsilon^{-nh(\theta_1,\theta_2)} d\theta_1 d\theta_2 \\
&\cong (2\pi)\, \epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \left\{ det\left(nD^2h\left(\hat{\theta}_1,\hat{\theta}_2\right)\right)\right\}^{-1/2} \left\{1-\Phi\left(\frac{a_1-\hat{\theta}_1}{\hat{v}_1}\right)\right\},
\end{aligned}
\tag{12}
$$

where $\hat{v}_1 = \left\{\frac{n\partial^2 h\left(\hat{\theta}_1,\hat{\theta}_2\right)/\partial\theta_2^2}{\Delta}\right\}^{1/2}$.

$\Delta = \left(n\, \frac{\partial^2 h\left(\hat{\theta}_1,\hat{\theta}_2\right)}{\partial\theta_1^2}\right)\left(n\, \frac{\partial^2 h\left(\hat{\theta}_1,\hat{\theta}_2\right)}{\partial\theta_2^2}\right) - \left(n\, \frac{\partial^2 h\left(\hat{\theta}_1,\hat{\theta}_2\right)}{\partial\theta_1\partial_2}\right)^2$ in (11) and (12).

If θ_1 and θ_2 are orthogonal parameters, we have,

$$\hat{I}_4 \cong (2\pi)\, \epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \hat{v}_1 \hat{v}_2\, \Phi\left(\frac{a_1 - \hat{\theta}_1}{\hat{v}_1}\right) \tag{13}$$

and when $-\infty < \theta_1 < \infty, -\infty < \theta_2 < a_2$.

$$\begin{aligned} I_6 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a_2} \epsilon^{-nh(\theta_1,\theta_2)} d\theta_1\, d\theta_2 \\ &\cong (2\pi)\, \epsilon^{-nh(\hat{\theta}_1,\hat{\theta}_2)} \hat{v}_1 \hat{v}_2 \Phi\left(\frac{a_2 - \hat{\theta}_2}{\hat{v}_2}\right) \end{aligned} \tag{14}$$

where

$$\hat{v}_1 = \left\{n\partial^2 h\left(\hat{\theta}_1, \hat{\theta}_2\right) / \partial\theta_1^2\right\}^{-1/2}$$

and

$$\hat{v}_2 = \left\{n\partial^2 h\left(\hat{\theta}_1, \hat{\theta}_2\right) / \partial\theta_2^2\right\}^{-1/2}.$$

We also get similar results for the multiparameter case $\theta \in R^m$.

3 Some Selected Applications

3.1 Binomial Likelihood: One Sample

Let Y be a random variable with a binomial distribution $b(n, \theta)$, where the likelihood function for θ is given by $L_n(\theta) \propto \theta^y(1-\theta)^{n-y}, 0 < \theta < 1$, and assume a Jeffreys prior density (see for example, Box and Tiao, 1973) given by $\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}, 0 < \theta < 1$.

The posterior moment (1) with $g(\theta) = \theta$, when $\theta < a, a < 1$, is given by,

$$E(\theta|\mathcal{D}) = \frac{\int_0^a \epsilon^{-nh(\theta)} d\theta}{\int_0^a \epsilon^{-nh^*(\theta)} d\theta}, \tag{15}$$

where $-nh(\theta) = (y + 1/2)ln\theta + (n - y - 1/2)ln(1-\theta)$

and $-nh^*(\theta) = (y - 1/2)ln\theta + (n - y - 1/2)ln(1-\theta)$.

Using Laplace's approximation (7) for both integrals in the numerator and denominator of (15), we get,

$$\hat{E}(\theta|\mathcal{D}) \cong \frac{(y+1/2)^{y+1}(n-1)^{n+1/2}\Phi\left(\frac{a-\hat{\theta}}{\hat{v}}\right)}{(y-1/2)^{y}n^{n+3/2}\Phi\left(\frac{a-\hat{\theta}^{*}}{\hat{v}^{*}}\right)} \tag{16}$$

where $\hat{\theta} = (y+1/2)/n$ maximizes $-nh(\theta), \hat{v} = (y+1/2)^{1/2}(n-y-1/2)^{1/2}/n^{3/2}, \hat{\theta}^{*} = (y-1/2)/(n-1)$ maximizes $-nh^{*}(\theta)$, and $\hat{v}^{*} = (y-1/2)^{1/2}(n-y-1/2)^{1/2}/(n-1)^{3/2}$.

As a numerical illustration, consider $n=5, y=3$ and $a=0.8$. Since $\hat{\theta}=0.7, \hat{\theta}^{*} = 0.625, \hat{v} = 0.20494$, and $\hat{v}^{*} = 0.24206$, we get $\hat{E}(\theta|\mathcal{D}) \cong 0.50561$.

It is interesting to observe that the exact posterior moment (15) is given by

$$E(\theta|\mathcal{D}) = \frac{(y+1/2)I_a(y+3/2, n-y+1/2)}{(n+1)I_a(y+1/2, n-y+1/2)} \tag{17}$$

where $I_x(\alpha,\beta)$ is the incomplete Beta function.

$$I_x(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt\ .$$

From (17), we get $E(\theta|\mathcal{D}) = 0.54056$.

The accuracy of Laplace's approximations could be improved by considering an appropriate reparametrization (see for example, Achcar and Smith, 1990). In parametrization $\xi = sin^{-1}\sqrt{(\theta)}$ (that is, $\theta = sin^2\xi$), with a locally uniform Jeffreys prior density. Laplace's approximation for (15) is given by

$$\hat{E}(\theta|\mathcal{D}) \cong \frac{n^{n+1/2}(y+1)^{y+1}\Phi\left(\frac{sin^{-1}\sqrt{a}-\hat{\xi}}{\hat{v}}\right)}{(n+1)^{n+3/2}y^{y}\ \Phi\left(\frac{sin^{-1}\sqrt{a}-\hat{\xi}^{*}}{\hat{v}^{*}}\right)}, \tag{18}$$

where $\hat{\xi} = tg^{-1}\sqrt{\frac{y+1}{n-y}}, \hat{v} = \frac{1}{2\sqrt{n+1}}, \hat{\xi}^{*} = tg^{-1}\sqrt{\frac{y}{n-y}}$ and $\hat{v}^{*} = \frac{1}{2\sqrt{n}}$.

With $n=5, y=3$ and $a=0.8$, we find $\hat{\xi} = 0.96, \hat{v} = 0.20412, \hat{\xi}^{*} = 0.89$ and $\hat{v}^{*} = 0.22361$. From (18), we get $\hat{E}(\theta|\mathcal{D}) \cong 0.53336$.

In table 1, we have Laplace's approximations for the posterior moment (1) considering $g(\theta)$ equals to θ and θ^2 and different values of a. We observe very accurate approximations considering parametrization $\xi = sin^{-1}\sqrt{\theta}$, especially for small values of n.

		$E(\theta\|\mathcal{D})$		$E(\theta^2\|\mathcal{D})$	
	a	$n=5, y=3$	$n=10, y=3$	$n=5, y=3$	$n=10, y=3$
Exact	0.2	0.15366(0%)	0.14489(0%)	0.02588(0%)	0.02273(0%)
	0.4	0.29994(0%)	0.25496(0%)	0.09566(0%)	0.07228(0%)
	0.6	0.43393(0%)	0.30894(0%)	0.20212(0%)	0.11038(0%)
	0.8	0.54056(0%)	0.31815(0%)	0.31817(0%)	0.11909(0%)
Laplace in parametrization θ	0.2	0.10457(46.95%)	0.16875(16.47%)	0.00831(211.52%)	0.03121(37.32%)
	0.4	0.22867(31.16%)	0.25217(1.11%)	0.04773(100.42%)	0.07113(1.62%)
	0.6	0.38375(13.08%)	0.30696(0.65%)	0.15225(32.75%)	0.10862(1.62%)
	0.8	0.50561(6.91%)	0.31731(0.27%)	0.28183(12.90%)	0.11821(0.74%)
Laplace in parametrization $\xi = sin^{-1}\sqrt{\theta}$	0.2	0.15766(2.61%)	0.15267(5.37%)	0.02425(6.72%)	0.02540(11.78%)
	0.4	0.29148(2.90%)	0.25373(0.49%)	0.08907(7.40%)	0.07191(0.5%)
	0.6	0.42579(1.91%)	0.30714(0.59%)	0.19404(4.16%)	0.10950(0.8%)
	0.8	0.53336(1.35%)	0.31662(0.48%)	0.31081(2.37%)	0.11826(0.7%)

Table 1: Laplace's approximations for posterior moments (percentage errors in parentheses)

3.2 Binomial Likelihood: Two Samples

Considering Y_i a random variable with a binomial distribution $b(n_i, \theta_i), 0 < \theta_i < 1, i = 1, 2$, the likelihood function for θ_1 and θ_2 is given by,

$$L_n(\theta_1, \theta_2) \propto \theta_1^{y_1}\theta_2^{y_2}(1-\theta_1)^{n_1-y_1}(1-\theta_2)^{n_2-y_2} \tag{19}$$

Assuming a Jeffreys prior density $\pi(\theta_1, \theta_2) \propto \theta_1^{-1/2}\theta_2^{-1/2}(1-\theta_1)^{-1/2}(1-\theta_2)^{-1/2}$, the posterior moment for $g(\theta_1, \theta_2) = \theta_1/\theta_2$ when $0 < \theta_i < a_i, i = 1, 2$ is given by

$$E\left(\frac{\theta_1}{\theta_2}|\mathcal{D}\right) = \frac{\int_0^{a_1}\int_0^{a_2} e^{-nh(\theta_1,\theta_2)}d\theta_1 d\theta_2}{\int_0^{a_1}\int_0^{a_2} e^{-nh^*(\theta_1,\theta_2)}d\theta_1 d\theta_2} \tag{20}$$

where $-nh(\theta_1, \theta_2) = (y_1+1/2)ln\theta_1 + (y_2-3/2)ln\theta_2 + (n_1-y_1-1/2)ln(1-\theta_1) + (n_2-y_2-1/2)ln(1-\theta_2)$ and $-nh^*(\theta_1, \theta_2) = (y_1-1/2)ln\theta_1 + (y_2-1/2)ln\theta_2 + (n_1-y_1-1/2)ln(1-\theta_1) + (n_2-y_2-1/2)ln(1-\theta_2)$.

Observe that is this case we have orthogonal parameters θ_1 and θ_2 and we use Laplace's approximations (10) for both integrals in the numerator and denominator of (20) to get,

$$\hat{E}\left(\frac{\theta_1}{\theta_2} \mid \mathcal{D}\right) \cong \frac{(n_1-1)^{n_1+1/2}(n_2-1)^{n_2+1/2}(y_1+1/2)^{y_1+1}(y_2-3/2)^{y_2-1}}{n_1^{n_1+3/2}(n_2-2)^{n_2-1/2}(y_1-1/2)^{y_1}(y_2-1/2)^{y_2}} \times$$

$$\times \frac{\Phi\left(\frac{a_1-\hat{\theta}_1}{\hat{v}_1}\right)\Phi\left(\frac{a_2-\hat{\theta}_2}{\hat{v}_2}\right)}{\Phi\left(\frac{a_1-\hat{\theta}_1^*}{\hat{v}_1^*}\right)\Phi\left(\frac{a_2-\hat{\theta}_2^*}{\hat{v}_2^*}\right)} \tag{21}$$

where

$$\hat{\theta}_1 = (y_1+1/2)\,/\,n_1, \hat{\theta}_2 = (y_2-3/2)\,/\,(n_2-2)\ ,$$

$$\hat{\theta}_1^* = (y_1-1/2)\,/\,(n_1-1), \hat{\theta}_2^* = (y_2-1/2)\,/\,(n_2-1)\ ,$$

$$\hat{v}_1 = (y_1+1/2)^{1/2}(n_1-y_1-1/2)^{1/2}\,/\,n_1^{3/2}\ ,$$

$$\hat{v}_2 = (y_2-3/2)^{1/2}(n_2-y_1-1/2)^{1/2}\,/\,(n_2-2)^{3/2}\ ,$$

$$\hat{v}_1^* = (y_1-1/2)^{1/2}(n_1-y_1-1/2)^{1/2}\,/\,(n_1-1)^{3/2}$$

and

$$\hat{v}_2^* = (y_2-1/2)^{1/2}(n_2-y_2-1/2)^{1/2}\,/\,(n_2-1)^{3/2}\ .$$

In parametrization $\xi_1 = sin^{-1}\sqrt{\theta_1}$ and $\xi_2 = sin^{-1}\sqrt{\theta_2}$ (that is, $\theta_1 = sin^2\xi_1$ and $\theta_2 = sin^2\xi_2$), where we have a locally uniform Jeffreys prior density, Laplace's approximation for the posterior moment of $g(\theta_1,\theta_2) = \theta_1/\theta_2$ is (from (10)) given by

$$\hat{E}\left(\frac{\theta_1}{\theta_2} \mid \mathcal{D}\right) \cong \frac{n_1^{n_1+1/2}n_2^{n_2+1/2}(y_1+1)^{y_1+1}(y_2-1)^{y_2-1}}{(n_1+1)^{n_1+3/2}(n_2-1)^{n_2-1/2}y_1^{y_1}y_2^{y_2}} \times$$

$$\times \frac{\Phi\left(\frac{sin^{-1}\sqrt{a_1}-\hat{\xi}_1}{\hat{v}_1}\right)\Phi\left(\frac{sin^{-1}\sqrt{a_2}-\hat{\xi}_2}{\hat{v}_2}\right)}{\Phi\left(\frac{sin^{-1}\sqrt{a_1}-\hat{\xi}_1^*}{\hat{v}_1^*}\right)\Phi\left(\frac{sin^{-1}\sqrt{a_2}-\hat{\xi}_2^*}{\hat{v}_2^*}\right)} \tag{22}$$

where

$$\hat{\xi}_1 = cot^{-1}\sqrt{\frac{n_1-y_1}{y_1+1}}, \hat{\xi}_2 = cot^{-1}\sqrt{\frac{n_2-y_2}{y_2-1}},$$

$$\hat{\xi}_1^* = cot^{-1}\sqrt{\frac{n_1-y_1}{y_1}}, \hat{\xi}_2^* = cot^{-1}\sqrt{\frac{n_2-y_2}{y_2}},$$

$$\hat{v}_1 = 1/2\sqrt{n_1+1}, \hat{v}_2 = 1/2\sqrt{n_2-1},$$

$$\hat{v}_1^* = 1/2\sqrt{n_1} \text{ and } \hat{v}_2^* = 1/2\sqrt{n_2}.$$

It is interesting to observe that the exact posterior moment (20) is given by

$$E\left(\frac{\theta_1}{\theta_2} \mid \mathcal{D}\right) = \frac{n_2(y_1+1/2)\,I_{a_1}(y_1+3/2, n_1-y_1+1/2)\,I_{a_2}(y_2-1/2, n_2-y_2+1/2)}{(n_1+1)(y_2-1/2)\,I_{a_1}(y_1+1/2, n_1-y_1+1/2)\,I_{a_2}(y_2+1/2, n_2-y_2+1/2)} \tag{23}$$

where $I_x(\alpha,\beta)$ is the incomplete Beta function (see (17)).

As a numerical illustration, consider $n_1 = 20, n_2 = 30, y_1 = 3, y_2 = 8$ and $a_1 = a_2 = 0.2$. From (23), we find the exact posterior moment $E(\theta_1/\theta_2|\mathcal{D}) = 0.77759$. In the original parametrization θ_1 and θ_2, we find Laplace's approximation given in (21) by $\hat{E}(\theta_1/\theta_2|\mathcal{D}) \cong 0.74278$ (percentage error of 4.48%). In parametrization $\xi_1 = sin^{-1}\sqrt{\theta_1}$ and $\xi_2 = sin^{-1}\sqrt{\theta_2}$ we find more accurate Laplace's approximation (see (22)) given by $\hat{E}(\theta_1/\theta_2|\mathcal{D}) \cong 0.75645$ (percentage error of 2.72%).

3.3 Two Variance Components

Consider the random effects model $y_{jk} = \theta + \epsilon_j + \epsilon_{jk}, j = 1,\ldots,J, k = 1,\ldots,\mathbb{K}$, where $E(\epsilon_j) = E(\epsilon_{jk}) = 0, var(\epsilon_j) = \delta_2^2$ and $var(\epsilon_{jk}) = \delta_1^2$. Assuming a noninformative prior density for θ, δ_1^2 and δ_2^2 given by

$$\pi\left(\theta, \delta_1^2, \delta_2^2\right) \propto \delta_1^{-2}\left(\delta_1^2 + \mathbb{K}\delta_2^2\right)^{-1}, \tag{24}$$

(see for example, Box and Tiao, 1973), it is easily verified that the corresponding joint marginal posterior density for δ_1^2 and δ_2^2 is given by

$$\pi(\delta_1^2, \delta_2^2|\mathcal{D}) \propto (\delta_1^2)^{-(\frac{v_1}{2}+1)}(\delta_1^2+\mathbb{K}\delta_2^2)^{-(\frac{v_2}{2}+1)} \times \times exp\left\{-\frac{1}{2}\left[\frac{v_1 m_1}{\delta_1^2} + \frac{v_2 m_2}{\delta_1^2+\mathbb{K}\delta_2^2}\right]\right\} \tag{25}$$

where

$$v_1 = J(\mathbb{K}-1), v_2 = J-1, v_1 m_1 = S_1, v_2 m_2 = S_2$$

with

$$S_1 = \sum_j \sum_k \left(y_{jk} - \bar{y}_{j\cdot}\right)^2, S_2 = \mathbb{K} \sum_j \left(\bar{y}_{j\cdot} - \bar{y}_{\cdot\cdot}\right)^2 .$$

Observe that the parameter space is restricted to positive values of δ_1^2 and δ_2^2. It is important to point out that for some applications, if we ignore the constraints $\delta_1^2 > 0$ and $\delta_2^2 > 0$, it could be possible to have negative values for the mode of the joint posterior density (25). In these situations, we should use Laplace's method for approximation of integrals using results of section 2, for different choices of prior density.

As a numerical illustration, consider the data set of table 2 generated from a table of random normal deviates for the two-component model (data set introduced by Box and Tiao, 1973, with $\theta = 5, \delta_1 = 4$ and $\delta_2 = 2$), where $m_1 = 14.9459, m_2 = 8.3363, S_1 = 358.7014, S_2 = 41.6816, v_1 = 24$ and $v_2 = 5$.

Batch	1	2	3	4	5	6
	7.298	5.220	0.110	2.212	0.282	1.722
	3.846	6.556	10.386	4.852	9.014	4.782
	2.434	0.608	13.434	7.092	4.458	8.106
	9.566	11.788	5.510	9.288	9.446	0.758
	7.990	-0.892	8.166	4.980	7.198	3.758
$y_{j\cdot}$	6.2268	4.6560	7.5212	5.6848	6.0796	3.8252
			$y_{\cdot\cdot} = 5.6656$			

Table 2: Data generated from a table of random normal deviates for the two-component model ($\theta = 5, \delta_1 = 4, \delta_2 = 2$).

The posterior mean for δ_1^2 is given by

$$E\left(\delta_1^2|\mathcal{D}\right) = \frac{\int_0^\infty \int_0^\infty exp\left\{-nh\left(\delta_1^2, \delta_2^2\right)\right\} d\delta_1^2 d\delta_2^2}{\int_0^\infty \int_0^\infty exp\left\{-nh^*\left(\delta_1^2, \delta_2^2\right)\right\} d\delta_1^2 d\delta_2^2} \quad (26)$$

where

$$-nh\left(\delta_1^2,\delta_2^2\right) = -\frac{v_1}{2} ln\left(\delta_1^2\right) - \left(\frac{v_2+2}{2}\right) ln\left(\delta_1^2 + \mathbb{K}\delta_2^2\right) -$$

$$-\frac{m_1 v_1}{2\delta_1^2} - \frac{m_2 v_2}{2\left(\delta_1^2 + \mathbb{K}\delta_2^2\right)}$$

and

$$-nh^*\left(\delta_1^2,\delta_2^2\right) = -\frac{(v_1+2)}{2} ln\delta_1^2 - \frac{(v_2+2)}{2} ln\left(\delta_1^2 + \mathbb{K}\delta_2^2\right)$$

$$-\frac{m_1 v_1}{2\delta_1^2} - \frac{m_2 v_2}{2\left(\delta_1^2 + \mathbb{K}\delta_2^2\right)}.$$

The unrestricted maximum of $-h$ is given by $\hat{\delta}_1^2 = m_1$ and $\hat{\delta}_2^2 = \left(\frac{m_2 v_2}{v_2+2} - m_1\right)/\mathbb{K}$. That is, $\hat{\delta}_1^2 = 14.9459$ and $\hat{\delta}_2^2 = -1.7983 < 0$.

Also, the maximum of $-h^*$ is given by $\hat{\delta}_1^{*2} = m_1 v_1/(v_1+2)$ and $\hat{\delta}_2^{*2} = \left(\frac{m_2 v_2}{v_2+2} - \frac{m_1 v_1}{v_1+2}\right)/\mathbb{K}$ That is, $\hat{\delta}_1^{*2} = 13.7962$ and $\hat{\delta}_2^{*2} = -1.5683$.

Therefore, we have $\delta_2^2 < 0$ for both integrals in the numerator and denominator of (26), and we use Laplace's approximation (12) with $a_1 = 0$ for both integrals in (26). With the data of table 2, we have,

$$\hat{E}\left(\delta_1^2|\mathcal{D}\right) \cong \frac{\epsilon\left(m_1 v_1\right)\left(v_1/2\right)^{(v_1-3)/2}\{1-\Phi(1.68)\}}{2\left(\frac{v_1}{2}+1\right)^{\frac{v_1-1}{2}}\{1-\Phi(1.57)\}} \quad (27)$$

That is, $\hat{E}(\delta_1^2|\mathcal{D}) \cong 13.1693$.

In the same way, we find Laplace's approximations for other posterior moments of interest. Writting $E(\delta_{12}^2|\mathcal{D})$, where $\delta_{12}^2 = \delta_1^2 + \mathbb{K}\delta_2^2$, in the same form of ratio of integrals as in (26), we also find negative values for δ_2^2 in the maximuns of $-h$ and $-h^*$. From (12), we find the approximation,

$$\hat{E}\left(\delta_{12}^2|\mathcal{D}\right) \cong \frac{\epsilon\left(m_2 v_2\right)\left(v_2/2\right)^{(v_2-3)/2}\{1-\Phi(0.84)\}}{2\left(\frac{v_2}{2}+1\right)^{(v_2-1)/2}\{1-\Phi(1.57)\}} \quad (28)$$

Also,

$$\hat{E}\left(\frac{\delta_{12}^2}{\delta_1^2} \mid \mathcal{D}\right) \cong \frac{(m_2 v_2)\left(\frac{v_1}{2}+2\right)^{\frac{v_1+1}{2}}\left(\frac{v_2}{2}\right)\{1-\Phi(0.71)\}}{(m_1 v_1)\left(\frac{v_1}{2}+1\right)^{\frac{v_1-1}{2}}\left(\frac{v_2}{2}+1\right)^{\frac{v_2-1}{2}}\{1-\Phi(1.57)\}} \tag{29}$$

That is, $\hat{E}(\delta_{12}^2|\mathcal{D}) \cong 40.3665$ and $\hat{E}(\frac{\delta_{12}^2}{\delta_1^2}|\mathcal{D}) \cong 3.2229$.

In this case, the exact posterior moments for $\delta_1^2, \delta_{12}^2$ and δ_{12}^2/δ_1^2 are given by

$$E\left(\delta_1^2|\mathcal{D}\right) = \frac{v_1 m_1 I_x\left(\frac{v_2}{2}, \frac{v_1}{2}-1\right)}{2\left(\frac{v_1}{2}-1\right) I_x\left(\frac{v_2}{2}, \frac{v_1}{2}\right)}$$

$$E\left(\delta_{12}^2|\mathcal{D}\right) = \frac{v_2 m_2 I_x\left(\frac{v_2}{2}-1, \frac{v_1}{2}\right)}{2\left(\frac{v_2}{2}-1\right) I_x\left(\frac{v_2}{2}, \frac{v_1}{2}\right)} \tag{30}$$

and,

$$E\left(\frac{\delta_{12}^2}{\delta_1^2}|\mathcal{D}\right) = \frac{v_2 m_2 I_x\left(\frac{v_2}{2}-1, \frac{v_1}{2}+1\right)}{(v_2-2)\, m_1 I_x\left(\frac{v_2}{2}, \frac{v_1}{2}\right)}$$

where $x = v_2 m_2/(v_1 m_1 + v_2 m_2)$.

With the generated data set of table 2, the exact values for the posterior moments (30) are given by $E(\delta_1^2|\mathcal{D}) = 14.2649$. $E(\delta_{12}^2|\mathcal{D}) = 28.8515$ and $E(\delta_{12}^2/\delta_1^2|\mathcal{D}) = 2.0555$.

The accuracy of Laplace's approximations for the posterior moments of $\delta_1^2, \delta_{12}^2$ and δ_{12}^2/δ_1^2 could be improved by searching for a better parametrization, especially with the generated data set of table 2, where the joint posterior density for δ_1^2 and δ_2^2 has a bad shape and is very far from normality (see Box and Tiao, 1973, pp 257).

In this way, we could explore different reparametrizations like $\xi_1' = ln(\delta_1^2)$ and $\xi_2' = ln(\delta_2^2+3)$ or $\xi_1 = ln(\delta_1^2)$ and ξ_2 given by the Box and Cox (1964) transformation.

$$\xi_2 = \begin{cases} \frac{(\delta_2^2+3)^{\lambda}-1}{\lambda} & if \quad \lambda \neq 0 \\ ln\,(\delta_2^2+3) & if \quad \lambda = 0 \end{cases} \tag{31}$$

To find an appropriate value for λ that gives good normality for the joint posterior density for ξ_1 and ξ_2, we choose λ in (31) that gives a small value for the third derivatives summaries of the joint posterior density for ξ_1 and ξ_2 (see, Kass and Slate, 1992).

One of these summaries is given by,

$$m^2\overline{B}^2 = \sum_{i,j,k,l,m,n} b_{ij}b_{lm}b_{kn}d_{ijk}d_{lmn} \tag{32}$$

where d_{ijk} are the third derivatives of the logarithm of the joint posterior density and b_{ij} are the corresponding elements of the inverse of the matrix $\sum^{-1} = (-c_{ij})$ where c_{ij} are the second derivatives of the logarithm of the joint posterior density calculated at the mode of the posterior density. In the original parametrization δ_1^2 and δ_2^2 we have $m^2\overline{B}^2 = 5.8022$.

In the parametrization $\xi_1' = ln(\delta_1^2)$ and $\xi_2' = ln(\delta_2^2+3)$ we have $m\overline{B}^2 = 1.5168$, which indicates an improvement in the joint normality for the joint posterior density.

Considering $\lambda = -0.7$ in transformation (31), we find $m^2\overline{B}^2 = 0.8571$ which indicates still a better normality for the joint posterior density of $\xi_1 = ln(\delta_1^2)$ and $\xi_2 = [(\delta_2^2+3)^{-0.7} - 1]/(-0.7)$.

In table 3, we observe better accuracy of Laplace's approximation (12) for the posterior moments of $\delta_1^2, \delta_{12}^2$ and δ_{12}^2/δ_1^2 considering the parametrizations (ξ_1, ξ_2) and (ξ_1', ξ_2').

	Exact Value	Original Parametrization	$\xi_1 = ln(\delta_1^2)$ $\xi_2 = \frac{(\delta_2^2+3)^{-0.7}-1}{(-0.7)}$	$\xi_2' = ln(\delta_1^2)$ $\xi_2' = ln(\delta_2^2+3)$
$E(\delta_1^2\|\mathcal{D})$	14.2649 (0%)	13.1693 (8.32%)	15.1649 (6.31%)	14.7632 (3.49%)
$E(\delta_{12}^2\|\mathcal{D})$	28.8515 (0%)	40.3665 (39.91%)	32.4014 (12.30%)	34.2379 (18.67%)
$E\left(\frac{\delta_{12}^2}{\delta_1^2}\|\mathcal{D}\right)$	2.0555 (0%)	3.2229 (56.8%)	2.2572 (9.81%)	2.3980 (16.66%)

Table 3: Laplace's Approximations Considering Different Parametrizations (percentage errors in parentheses)

4 Concluding Remarks

The proposed generalization of Laplace's method to approximate posterior moments of interest could be very useful in applications where the mode of the posterior density is not in the interior of the parameter space.

The great advantage of Laplace's method is given in terms of simplicity which does not require sophisticated computational expertise as required by other existing approaches to solve Bayesian integrals.

It is very important to have appropriate parametrization to get good normality for the posterior densities of interest, which implies in good accuracy of the approximate posterior moments. We also observe that parameter orthogonality usually implies in great simplifications for the proposed approximations.

References

Achcar, J.A.; Smith, A.F.M. (1990). "Aspects of reparametrization in approximate Bayesian inference". In essays in honour of G.A. Barnard. Ed. J. Hodges, pp. 439-452, Amsterdam, North-Holland.

Box, G.E.P.; Cox, D.R. (1964). "An analysis of transformations", Journal of the Royal Statistical Society, series B, 26, 211-252.

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

Cox, D.R.; Reid, N. (1987). "Parameter orthogonality and approximate conditional inference", Journal of the Royal Statistical Society, series B, 49, 1, 1-39.

Kass, R.E.; Slate, E.H. (1992). " Reparametrization and diagnostics of posterior non-normality", in Bayesian statistics 4. Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, 289-306. Oxford University Press.

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. (1989). "Fully exponential Laplace approximations to expectations and variances of nonpositive functions". Journal of the American Statistical Association, 84, 710-716.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

014/94 ACHCAR, J. A.; FOGO, J.C. - *Accurate inferences for the reliability function considering accelerated life tests.*

013/94 RODRIGUES, J. - *Bayesian Solutions to a lass of selections problems using weighted loss functions.*

012/94 ACHCAR, J.A.; DAMASCENO, V.L. - *Extreme value models: an useful reparametrization for the survival function*

011/94 ACHCAR, J.A.- *Approximate bayesian analysis for non-normal hierarchical classification model.*

010/94 RODRIGUES, J.; RODRIGUES, E.F. - *Bayesian estimationin the study of tampered Random variables.*

009/94 ACHCAR, J.A.; FOGO, J.C. - *An useful reparametrization for the reliability in the Weibull case.*

008/94 RODRIGUES, J.; LEITE, J.G. - *A note on Bayesian analysis in M/M/1 queues from confidence intervals.*

007/94 ACHCAR, J.A.; DAMACENO, V.L. - *An useful reparametrization for the survival function considering an exponential regression model.*

006/94 RODRIGUES, J. - *Bayesian estimation of a normal mean parameter using liner loss function and robustness considerations.*

005/93 ACHCAR, J.A.; FOGO, J.C.- *Some useful reparametrizations for the reliability function considering an exponential regession model.*