UNIVERSIDADE DE SÃO PAULO
CAMPUS DE SÃO CARLOS
INSTITUTO DE CIÊNCIAS MATEMÁTICAS DE SÃO CARLOS

I.C.M.S.C.

# ACCURACY OF LAPLACE APPROXIMATIONS FOR POSTERIOR MOMENTS

JORGE ALBERTO ACHCAR

Nº 39

Notas do I.C.M.S.C. - USP

ISSN 0103-2577

# ACCURACY OF LAPLACE APPROXIMATIONS FOR POSTERIOR MOMENTS

JORGE ALBERTO ACHCAR

Nº 39

DEDALUS - Acervo - ICMSC
30300023513

SÃO CARLOS. - SP

1988

# ACCURACY OF LAPLACE APPROXIMATIONS FOR POSTERIOR MOMENTS

Jorge Alberto Achcar

SUMMARY

"The use of good parametrization is essential to get accurate Laplace approximations for posterior moments. We present a sufficient condition to be used as a diagnostic checking criterion of the accuracy of Laplace approximations for posterior moments using a given parametrization. The proposed procedure also can be used as a discriminatory method to choose the best parametrization".

Key words: Laplace method. Posterior moment. Parametrization. Accuracy.

## 1. INTRODUCTION

Denoting the likelihood function for a parameter $\theta$ in the uniparameter case by $lik(\theta)$ and the prior density for $\theta$ by $\pi(\theta)$, the posterior moment for selected real-valued functions $g(\theta)$ is given by:

$$E[g(\theta)\,|\,data] = \frac{\int g(\theta)\pi(\theta)lik(\theta)d\theta}{\int \pi(\theta)lik(\theta)d\theta} \tag{1}$$

Usually, the integrations required for the numerator and denominator of (1) cannot be carried out exactly using analytical methods and the statistician uses approximation techniques. One possibility is to use numerical methods (see for example, Naylor and Smith, 1982; and Smith et al, 1987). Another possibility is to use approximation methods for integrals or ratio of integrals. One of the most popular approximation method, is given by the Laplace method for integrals (see for example, Tierney and Kadane, 1986; Tierney, Kass and Kadane, 1986; kass, Tierney and Kadane, 1987). In The application of Laplace method, usually the statistician is concerned about the accuracy of the approximation for (1), specially for small values of n. Methods justifying derivations of asymptotic expansions for predictive densities, odds factors, marginal posterior densities and posterior moments are discussed by various authors (see for example, Johnson, 1967, 1970; and Kass, Tierney and Kadane, 1987), but the practical problem to choose good parametrization is not very well explored in the literature.

In this paper, we present a sufficient condition to the used as a criterion to check the accuracy of Laplace approximations for (1), considering a parametrization θ and a prior density π(θ). In this context, the effect of different priors and different parametrizations in the accuracy of Laplace approximations for (1), can be verified using the proposed method.

## 2. THE LAPLACE METHOD

The posterior moment (1), can be written as,

$$E\{g(\theta)|\text{data}\} = \frac{\int e^{-nh^*(\theta)}d\theta}{\int e^{-nh(\theta)}d\theta}, \tag{2}$$

where $-nh = \ell n(\pi)+L$, $-nh^* = \ell n(g)+\ell n(\pi)+L$ with L denoting the log-likelihood function and π the prior density.

In the uniparameter case, if $\hat{\theta}$ maximizes $-h$ and $\sigma=[h''(\hat{\theta})]^{-1/2}$, the Laplace method of approximation gives:

$$\int e^{-nh(\theta)}d\theta \doteq \sqrt{2\pi}\ \sigma n^{-1/2} e^{-nh(\hat{\theta})} \tag{3}$$

Similarly, if $g(\theta)$ is a smooth positive function, $\hat{\theta}^*$ maximizes $-h^*$ and $\sigma^* = [h^{*''}(\hat{\theta}^*)]^{-1/2}$, the Laplace approximation for the numerator of (2) is given by $\sqrt{2\pi}\ \sigma^* n^{-1/2}\exp\{-nh^*(\hat{\theta}^*)\}$. The implied approximation for (2) is given by:

$$\hat{E}\{g(\theta)|\text{data}\} \doteq (\sigma^*/\sigma)\exp\{-n[h^*(\hat{\theta}^*)-h(\hat{\theta})]\} \tag{4}$$

## 3. STUDY OF THE ERROR IN LAPLACE APPROXIMATION FOR $E\{g(\theta)|\text{data}\}$

Considering a Taylor expansion of $h(\theta)$ defined in (2) about the mode $\hat{\theta}$ up to third derivative, we find a criterion to study the error of the Laplace approximation (4) in different parametrizations. Assuming $h(\theta)$ thrice differentiable and since $h'(\hat{\theta}) = 0$, we have:

$$e^{-nh(\theta)} \doteq e^{-nh(\hat{\theta})}.e^{-\frac{n(\theta-\hat{\theta})^2 h''(\hat{\theta})}{2}}.e^{-\frac{n(\theta-\hat{\theta})^3 h^{(3)}(\hat{\theta})}{6}} \tag{5}$$

Using $e^{-x} \cong 1-x+x^2/2$, we have $\exp[-\frac{n(\theta-\hat{\theta})^3 h^{(3)}(\hat{\theta})}{6}] \cong 1-\frac{n(\theta-\hat{\theta})^3 h^{(3)}(\hat{\theta})}{6} + \frac{n^2(\theta-\hat{\theta})^6 (h^{(3)}(\hat{\theta}))^2}{72}$. Since $E(\theta-\hat{\theta})^3 = 0$ for normal posterior densities for $\theta$, we find:

$$\int e^{-nh(\theta)} d\theta \doteq e^{-nh(\hat{\theta})} \int e^{-\frac{n(\theta-\hat{\theta})^2 h''(\hat{\theta})}{2}} d\theta +$$

$$+ \frac{e^{-nh(\hat{\theta})} n^2 [h^{(3)}(\hat{\theta})]^2}{72} \int (\theta-\hat{\theta})^6 e^{-\frac{n(\theta-\hat{\theta})^2 h''(\hat{\theta})}{2}} d\theta.$$

Since $E(\theta-\hat{\theta})^6 = 15n^{-3}\{h''(\hat{\theta})\}^{-3}$, we find:

$$\int e^{-nh(\theta)} d\theta \doteq \sqrt{2\pi}\, n^{-1/2} [h''(\hat{\theta})]^{-1/2} e^{-nh(\hat{\theta})} \{1 + \frac{15[h^{(3)}(\hat{\theta})]^2}{72n[h''(\hat{\theta})]^3}\} \tag{6}$$

Similarly, we find (6) for the numerator of (2) considering $h^*(\theta)$ and $\hat{\theta}^*$ the mode of $h^*(\theta)$. Thus, we find:

$$E\{g(\theta)|\text{data}\} \doteq (\varepsilon^\theta)\hat{E}\{g(\theta)|\text{data}\} \tag{7}$$

where $\hat{E}\{g(\theta)|\text{data}\}$ is the Laplace approximation (4), and,

$$\varepsilon^\theta = \{1 + \frac{15[h^{*(3)}(\hat{\theta}^*)]^2}{72n[h^{*''}(\hat{\theta}^*)]^3}\} / \{1 + \frac{15[h^{(3)}(\hat{\theta})]^2}{72n[h''(\hat{\theta})]^3}\}.$$

We observe that $\varepsilon^\theta$ can be used as a criterion to check the accuracy of the Laplace approximation (4). Thus, if $\varepsilon^\theta \cong 1$, the Laplace approximation (4) is very accurate. Also, we can find $\varepsilon$ for different parametrizations to be used as a discriminatory method to choose the best parametrization in the approximation for $E\{g(\theta)|\text{data}\}$. Therefore, we can write the following result:

<u>RESULT 1:</u> The Laplace apprxomation $\hat{E}\{g(\theta)|\text{data}\}$ is very accurate if $\varepsilon^\theta \cong 1$. That is, if

$$\frac{[h^{*(3)}(\hat{\theta}^*)]^2}{[h^{*''}(\hat{\theta}^*)]^3} - \frac{[h^{(3)}(\hat{\theta})]^2}{[h''(\hat{\theta})]^3} \cong 0 \tag{8}$$

## 4. DETERMINATION OF $\varepsilon^{\theta}$ IN THE PARAMETRIZATION $\theta$

The first three derivatives of $h(\theta)$ defined in (2), are given by:

$$nh'(\theta) = -\frac{\pi'(\theta)}{\pi(\theta)} - L'(\theta)$$

$$nh''(\theta) = \left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2 - \frac{\pi''(\theta)}{\pi(\theta)} - L''(\theta) \qquad (9)$$

$$nh^{(3)}(\theta) = \frac{3\pi'(\theta)\pi''(\theta)}{\pi^2(\theta)} - 2\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^3 - \frac{\pi^{(3)}(\theta)}{\pi(\theta)} - L^{(3)}(\theta)$$

Since the mode $\hat{\theta}$ satisfies,

$$\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + L'(\hat{\theta}) = 0, \qquad (10)$$

we have,

$$\frac{[h^{(3)}(\hat{\theta})]^2}{[h''(\hat{\theta})]^3} = \frac{n[a_1(\hat{\theta})]^2}{[a_2(\hat{\theta})]^3}, \qquad (11)$$

where $a_1(\theta) = -3L'(\theta)\frac{\pi''(\theta)}{\pi(\theta)} + 2(L'(\theta))^3 - L^{(3)}(\theta) - \frac{\pi^{(3)}(\theta)}{\pi(\theta)}$ and $a_2(\theta) = (L'(\theta))^2 - L''(\theta) - \frac{\pi''(\theta)}{\pi(\theta)}$.

The first three derivatives of $h^*(\theta)$ are given by:

$$nh^{*\prime}(\theta) = -\frac{g'(\theta)}{g(\theta)} - \frac{\pi'(\theta)}{\pi(\theta)} - L'(\theta) \qquad (12)$$

$$nh^{*\prime\prime}(\theta) = -\frac{g''(\theta)}{g(\theta)} + \left(\frac{g'(\theta)}{g(\theta)}\right)^2 - \frac{\pi''(\theta)}{\pi(\theta)} + \left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2 - L''(\theta)$$

$$nh^{*(3)}(\theta) = -\frac{g^{(3)}(\theta)}{g(\theta)} + \frac{3g'(\theta)g''(\theta)}{g^2(\theta)} - 2\left(\frac{g'(\theta)}{g(\theta)}\right)^3 - \frac{\pi^{(3)}(\theta)}{\pi(\theta)} + \frac{3\pi'(\theta)\pi''(\theta)}{\pi^2(\theta)} - 2\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^3 - L^{(3)}(\theta).$$

Since the mode $\hat{\theta}^*$ of $h^*(\theta)$ satisfies,

$$\frac{\pi'(\hat{\theta}^*)}{\pi(\hat{\theta}^*)} = -\frac{g'(\hat{\theta}^*)}{g(\hat{\theta}^*)} - L'(\hat{\theta}^*), \qquad (13)$$

we have,

$$\frac{[h^{*(3)}(\hat{\theta}^*)]^2}{[h^{*''}(\hat{\theta}^*)]^3} = \frac{n[a_1(\hat{\theta}^*)+a_3(\hat{\theta}^*)]^2}{[a_2(\hat{\theta}^*)+a_4(\hat{\theta}^*)]^3}, \tag{14}$$

where $a_3(\theta) = -\frac{g^{(3)}(\theta)}{g(\theta)} + \frac{3g'(\theta)g''(\theta)}{g^2(\theta)} - \frac{3\pi''(\theta)g'(\theta)}{\pi(\theta)\; g(\theta)} + 6L'(\theta)(\frac{g'(\theta)}{g(\theta)})^2 +$

$+ 6(L'(\theta))^2 \frac{g'(\theta)}{g(\theta)}$, and $a_4(\theta) = -\frac{g''(\theta)}{g(\theta)} + 2(\frac{g'(\theta)}{g(\theta)})^2 + 2L'(\theta)\frac{g'(\theta)}{g(\theta)}$.

That is, $\varepsilon^\theta$ given in (7) is:

$$\varepsilon^\theta = \{1 + \frac{15[a_1(\hat{\theta}^*)+a_3(\hat{\theta}^*)]^2}{72[a_2(\hat{\theta}^*)+a_4(\hat{\theta}^*)]^3}\}/\{1 + \frac{15[a_1(\hat{\theta})]^2}{72[a_2(\hat{\theta})]^3}\} \tag{15}$$

Therefore, we can write the following result:

<u>RESULT 2:</u> A sufficient condition to have a very accurate Laplace approximation $\hat{E}\{g(\theta)|\text{data}\}$ in the parametrization $\theta$, is given if,

$$A^\theta = \frac{[a_1(\hat{\theta}^*)+a_3(\hat{\theta}^*)]^2}{[a_2(\hat{\theta}^*)+a_4(\hat{\theta}^*)]^3} - \frac{[a_1(\hat{\theta})]^2}{[a_2(\hat{\theta})]^3} \cong 0, \tag{16}$$

where $a_1(\theta)$, $a_2(\theta)$ are defined in (11) and $a_3(\theta)$, $a_4(\theta)$ are defined in (14).

A special case is given in the parametrization $\theta$ with a locally uniform prior. Since $\pi(\theta) \propto \text{constant}$, we find $\hat{\theta}^*$ (from (13)), by $\frac{g'(\hat{\theta}^*)}{g(\hat{\theta}^*)} +$ $+ L'(\hat{\theta}^*) = 0$. Thus, we have:

$$a_1(\hat{\theta}^*) = -2(\frac{g'(\hat{\theta}^*)}{g(\hat{\theta}^*)})^3 - L^{(3)}(\hat{\theta}^*),$$

$$a_2(\hat{\theta}^*) = (\frac{g'(\hat{\theta}^*)}{g(\hat{\theta}^*)})^2 - L''(\hat{\theta}^*) \tag{17}$$

$$a_3(\hat{\theta}^*) = -\frac{g^{(3)}(\hat{\theta}^*)}{g(\hat{\theta}^*)} + \frac{3g''(\hat{\theta}^*)g'(\hat{\theta}^*)}{g^2(\hat{\theta}^*)}$$

and $$a_4(\hat{\theta}^*) = -\frac{g''(\hat{\theta}^*)}{g(\hat{\theta}^*)}.$$

Also, if $L'(\hat{\theta})=0$, $a_1(\hat{\theta}) = -L^{(3)}(\hat{\theta})$ and $a_2(\hat{\theta}) = -L''(\hat{\theta})$. If we have interest in the first posterior moment $E\{\theta|\text{data}\}$ where $\theta$ has a locally uniform prior, (17) is reduced to $a_1(\hat{\theta}^*) = -2\hat{\theta}^{*-3} - L^{(3)}(\hat{\theta}^*)$, $a_2(\hat{\theta}^*) = \hat{\theta}^{*-2} - L''(\hat{\theta}^*)$, $a_3(\hat{\theta}^*) = 0$ and $a_4(\hat{\theta}^*) = 0$. Also, $a_1(\hat{\theta}) = -L^{(3)}(\hat{\theta})$ and $a_2(\hat{\theta}) = -L''(\hat{\theta})$.

## 5. SOME EXAMPLES

### 5.1 - EXPONENTIAL DISTRIBUTION

Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ of an exponential distribution with parameter $\theta$. The Jeffreys prior for $\theta$ is proportional to $\theta^{-1}$, $\theta > 0$ and $L(\theta) \propto -n\ell n\theta - n\bar{x}\theta^{-1}$. To approximate $E\{\theta|\text{data}\}$ using Laplace method, we have $nh(\theta) = (n+1)\ell n\theta + n\bar{x}\theta^{-1}$ and $nh^*(\theta) = n\ell n\theta + n\bar{x}\theta^{-1}$. From $h'(\theta) = 0$, we find $\hat{\theta} = n\bar{x}/(n+1)$, $h''(\hat{\theta}) = (n+1)^3/n^3\bar{x}^2$ and $h^{(3)}(\hat{\theta}) = -4(n+1)^4/n^4\bar{x}^3$. From $h^{*\prime}(\theta) = 0$, we find $\hat{\theta}^* = \bar{x}$, $h^{*\prime\prime}(\hat{\theta}^*) = 1/\bar{x}^2$ and $h^{*(3)}(\hat{\theta}^*) = -4/\bar{x}^3$.

That is,

$$\varepsilon^{\theta} = 1 + \frac{240}{n(72n+312)} \tag{18}$$

Considering the reparametrization $\psi = \ell n\theta$, we have a locally uniform Jeffreys prior for $\psi$, where $-\infty < \psi < \infty$, and $L(\psi) \propto -n\psi - n\bar{x}e^{-\psi}$. To approximate $E\{\theta|\text{data}\}$ using Laplace method, we have $\theta = g(\psi) = e^{\psi}$, $h(\psi) = \psi + \bar{x}e^{-\psi}$ and $h^*(\psi) = (n-1)n^{-1}\psi + \bar{x}e^{-\psi}$. From $h'(\psi) = 0$, we find $\hat{\psi} = \ell n\bar{x}$, $h''(\hat{\psi}) = 1$ and $h^{(3)}(\hat{\psi}) = -1$. From $h^{*\prime}(\psi) = 0$, we find $e^{\hat{\psi}^*} = n\bar{x}/(n-1)$, $h^{*\prime\prime}(\hat{\psi}^*) = (n-1)/n$ and $h^{*(3)}(\hat{\psi}^*) = -(n-1)/n$.

That is,

$$\varepsilon^{\psi} = 1 + \frac{15}{(n-1)(72n+15)} \tag{19}$$

Since $\varepsilon^{\theta} = \varepsilon^{\psi} + B_n$, where $B_n = \{16200(n+0.1704)(n-1.3038)\}/\{n(n-1)(72n+312)(72n+15)\} > 0$, we have $\varepsilon^{\theta} > \varepsilon^{\psi}$ for all $n > 1$. Thus, we have more accurate Laplace approximation for $E\{\theta|\text{data}\}$ in the parametrization $\psi$, specially for small values of $n$.

### 5.2 - POISSON DISTRIBUTION

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size $n$ of a Poisson distribution with parameter $\theta$. The Jeffreys prior for $\theta$ is proportional to $\theta^{-1/2}$, $\theta > 0$ and $L(\theta) \propto n\bar{y}\ell n\theta - n\theta$. To approximate $E\{\theta|\text{data}\}$ using Laplace method, we have $nh(\theta) = (1/2 - n\bar{y})\ell n\theta + n\theta$ and $nh^*(\theta) = -(1/2 + n\bar{y})\ell n\theta + n\theta$. We find $\hat{\theta} = (n\bar{y} - 1/2)/n$, $\hat{\theta}^* = (n\bar{y} + 1/2)/n$, $a_1(\hat{\theta}) = -2n^3/(n\bar{y} - 1/2)^2$ and $a_1(\hat{\theta}^*) + a_3(\hat{\theta}^*) = -2n^3/(n\bar{y} + 1/2)^2$. Also, $a_2(\hat{\theta}) = n^2/(n\bar{y} - 1/2)$ and $a_2(\hat{\theta}^*) + a_4(\hat{\theta}^*) = n^2/(n\bar{y} + 1/2)$. That is, $A^{\theta} = -16/(4n^2\bar{y}^2-1)$. Considering the parametrization $\phi = \theta^{1/2}$, we have a locally uniform Jeffreys prior for $\phi$. In this parametrization we have $g(\phi)=\phi^2$, $L(\phi) \propto 2n\bar{y}\ln\phi - n\phi^2$, $nh(\phi) = -\ln$ (constant)$-2n\bar{y}\ln\phi+n\phi^2$ and $nh^*(\phi) = -\ln$ (constant)$-2(n\bar{y}+1)\ln\phi+n\phi^2$. Since $\hat{\phi}=\bar{y}$ and $\hat{\phi}^{*2}=(n\bar{y}+1)/n$, we have $a_1(\hat{\phi}) = -4n/\bar{y}^{1/2}$, $a_2(\hat{\phi})=4n$, $a_1(\hat{\phi}^*)+a_3(\hat{\phi}^*) = -4n^{3/2}/(n\bar{y}+1)^{1/2}$, and $a_2(\hat{\phi}^*)+a_4(\hat{\phi}^*) = 4n$. That is, $A^{\phi} = -1/\{4n\bar{y}(n\bar{y}+1)\}$. Since $A^{\theta}=A^{\phi}-C_n$ where $C_n = \{60(n\bar{y}+1.0508)(n\bar{y}+0.0158)\}/\{16n\bar{y}(n\bar{y}+1)(n^2\bar{y}^2-1/4)\}$, we have $|A^{\theta}|>|A^{\phi}|$ for all $n$ such that $n\bar{y}>1/2$. We observe that the sufficient condition to have very accurate Laplace approximation $\hat{E}\{\theta|\text{data}\}$, given in result 2, is better verified in the parametrization $\phi$ with a locally uniform prior. In Table 1, we have a numerical illustration considering n=1.

TABLE 1. POISSON DISTRIBUTION (n=1).

| $\bar{y}$ | $\lvert A^{\theta}\rvert$ | $\lvert A^{\phi}\rvert$ |
|---|---|---|
| 1 | 5.3333 | 0.1250 |
| 4 | 0.2539 | 0.0125 |
| 9 | 0.0495 | 0.0028 |

In Table 2, we have the exact values of $E\{\theta|\text{data}\}$ considering n=1, and the Laplace approximations using both parametrizations $\theta$ and $\phi$ (see Achcar and Smith, 1988). We observe that the Laplace approximation $\hat{E}\{\theta|\text{data}\}$ is very accurate in the parametrization $\phi$.

TABLE 2. LAPLACE APPROXIMATIONS $\hat{E}\{\theta|\text{DATA}\}$ (n=1) ERROR IN PARENTESES.

| | | LAPLACE IN PARAMETRIZATION | |
|---|---|---|---|
| $\bar{y}$ | EXACT | $\theta$ | $\phi$ |
| 1 | 1.5000 | 1.6555 (9.40%) | 1.4715 (1.93%) |
| 4 | 4.5000 | 4.5237 (0.52%) | 4.4907 (0.21%) |
| 9 | 9.500 | 9.5098 (0.10%) | 9.4956 (0.05%) |

REFERENCES

ACHCAR, J.A.; SMITH, A.F.M. (1988). Parameter transformations in approximation of posterior moments using Laplace's method, technical report, Department of Mathematics, University of Nottingham.

JOHNSON, R.A. (1967). An asymptotic expansion for posterior distributions, Ann. Math. Statist., 38, 1266-1272.

Instituto de Ciências Matemáticas S. Carlos - Biblioteca

JOHNSON, R.A. (1970). Asymptotic expansions associated with posterior distributions, Ann. Math. Statist., 41, 851-864.

KASS, R.E.; TIERNEY, L.; KADANE, J.B. (1987). The validity of posterior expansions based on Laplace's method, technical report, Department of Statistics, Carnegie-Mellon University.

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

SMITH, A.F.M.; SKENE, A.M.; SHAW, J.E.H.; NAYLOR, J.C. (1987). Progress with numerical and graphical methods for practical Bayesian statistics, The Statistician, 36, 75-82.

TIERNEY, L.; KADANE, J.B. (1986). Accurate approximations for posterior moments and marginal densities, J. Amer. Statist. Ass., 81, 82-86.

TIERNEY, L.; KASS, R.E.; KADANE, J.B. (1986). Approximations of posterior expectations and variances using Laplace's method, technical report, Department of statistics, Carnegie-Mellon University.