UNIVERSIDADE DE SÃO PAULO
CAMPUS DE SÃO CARLOS
INSTITUTO DE CIÊNCIAS MATEMÁTICAS DE SÃO CARLOS
I.C.M.S.C.

Laplace approximations for posterior moments: an invariant reparametrization

ACHCAR, J.A.

nº 50

Notas do ICMSC - USP

ISSN 0103-2577

Laplace approximations for posterior moments: an invariant reparametrization

ACHCAR, J.A.

nº 50

São Carlos (SP)
1989

LAPLACE APPROXIMATIONS FOR POSTERIOR MOMENTS: AN INVARIANT REPARAMETRIZATION

Jorge Alberto Achcar
Departamento de Estatística
Universidade Federal de São Carlos
C. Postal 676, 13560 - São Carlos, SP, Brazil

SUMMARY

For the application of Laplace's method for integrals in the approximation of posterior moments, the choice of good parametrization is important to get accurate results. In the multiparameter case and considering a parametrization with a locally uniform prior, we show that linear transformations of the parameters give invariant Laplace approximations for posterior moments of interest.

1. INTRODUCTION

The Laplace's method for approximation of integrals is a powerful general technique that has being extensively used by Bayesian statisticians in the approximation of marginal posterior densities and posterior moments of interest (see for example, Tierney and Kadane, 1986; see also, Tierney, Kass and Kadane, 1986, and Kass, Tierney and Kadane, 1987, for further extensions of the method). Usually, to get accurate approximations for posterior moments, it is necessary a good choice of parametrization, specially for small sample sizes. Since the Laplace's method for integrals is based on second-order Taylor approximation applied to the exponent term in the integrand, the approximate posterior moments will be very accurate when we have posterior distributions close to normality. In the situation with an uniform prior, we can consider a reparametrization to have the likelihood approximately normal, by ensuring that third derivatives of the log-likelihood are zero (see for example, Anscombe, 1964; Sprott, 1973, 1980). This parametrization is equivalent, in the Bayesian approach, to consider a reparametrization with a Jeffreys locally uniform prior.

Considering this parametrization with a locally uniform prior, we show that Laplace approximations for posterior moments are invariant if we consider linear transformations of the parameters. This result could be of great practical interest, since in general the approximation procedures are not invariant to changes in the parametrization chosen when specifying the likelihood and prior density functions (see for example, Achcar and Smith, 1988).

2. AN INVARIANT REPARAMETRIZATION

THEOREM: "Let $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)'$ be a parameter vector of dimension m with a locally uniform Jeffreys prior and let $g(\underset{\sim}{\theta})$ be a nonnegative function of $\underset{\sim}{\theta}$. Consider the reparametrization $\underset{\sim}{\phi} = A\underset{\sim}{\theta}$, where A is a nonsingular mxm matrix. In the parametrization $\underset{\sim}{\phi}$, we also have a locally uniform prior and the

Laplace approximation for the posterior moment $E\{g(\underset{\sim}{\theta})|\underset{\sim}{y}\}$, where $\underset{\sim}{y}$ is the data vector, coincides in both parametrizations $\underset{\sim}{\theta}$ and $\underset{\sim}{\phi}$".

PROOF.: For selected nonnegative functions $g_\theta(\underset{\sim}{\theta})$ and a parametrization $\underset{\sim}{\theta}$ with a Jeffreys locally uniform prior, the posterior moment for $g_\theta(\underset{\sim}{\theta})$ is given by,

$$E\{g_\theta(\underset{\sim}{\theta})|\underset{\sim}{y}\} = \frac{\int g_\theta(\underset{\sim}{\theta})\ell_\theta(\underset{\sim}{\theta}|\underset{\sim}{y})d\underset{\sim}{\theta}}{\int \ell_\theta(\underset{\sim}{\theta}|\underset{\sim}{y})d\underset{\sim}{\theta}} \tag{1}$$

where $\ell_\theta(\underset{\sim}{\theta}|\underset{\sim}{y})$ is the likelihood function for $\underset{\sim}{\theta}$.

We observe that (1) can be written by,

$$E\{g_\theta(\underset{\sim}{\theta})|\underset{\sim}{y}\} = \frac{\int e^{-nh^*_\theta(\underset{\sim}{\theta})}d\underset{\sim}{\theta}}{\int e^{-nh_\theta(\underset{\sim}{\theta})}d\underset{\sim}{\theta}} \tag{2}$$

where $-nh^*_\theta(\underset{\sim}{\theta}) = \ell n\{g_\theta(\underset{\sim}{\theta})\} + L_\theta(\underset{\sim}{\theta})$, $-nh_\theta(\underset{\sim}{\theta}) = L_\theta(\underset{\sim}{\theta})$ and $L_\theta(\underset{\sim}{\theta})$ is the log-likelihood function for $\underset{\sim}{\theta}$.

The Laplace approximation for $E\{g_\theta(\theta)|\underset{\sim}{y}\}$ (see for example, Tierney and Kadane, 1986) is given by,

$$\hat{E}\{g_\theta(\underset{\sim}{\theta})|\underset{\sim}{y}\} = \left\{\frac{\det D^2h^*_\theta(\hat{\underset{\sim}{\theta}}^*)}{\det D^2h_\theta(\hat{\underset{\sim}{\theta}})}\right\}^{-1/2} \exp\{-n[h^*_\theta(\hat{\underset{\sim}{\theta}}^*) - h_\theta(\hat{\underset{\sim}{\theta}})]\} \tag{3}$$

where D^2h_θ is the Hessian matrix of h_θ, $D^2h^*_\theta$ is the Hessian matrix of h^*_θ, $\hat{\underset{\sim}{\theta}}$ and $\hat{\underset{\sim}{\theta}}^*$ maximize $-h_\theta$ and $-h^*_\theta$, respectively.

Consider the linear transformation $\underset{\sim}{\phi} = A\underset{\sim}{\theta}$, where $A = (a_{ij})$, $i,j = 1,2,\ldots,m$ is a mxm nonsigular matrix. With

$A^{-1} = B = (b_{ij})$, $i,j = 1,2,\ldots,m$, the inverse matrix of A, we have $\underset{\sim}{\theta} = B\underset{\sim}{\phi}$.

We observe that the prior density in the parametrization $\underset{\sim}{\phi}$ is $\pi_\phi(\underset{\sim}{\phi}) \propto \text{constant}|\det B|$, that is, $\pi_\phi(\underset{\sim}{\phi}) \propto$ constant. We also observe that $h_\phi(\underset{\sim}{\phi}) = h_\theta(B\underset{\sim}{\phi})$, $h^*_\phi(\underset{\sim}{\phi}) = h^*_\theta(B\underset{\sim}{\phi})$ and $g_\phi(\underset{\sim}{\phi}) = g_\theta(B\underset{\sim}{\phi})$.

The Laplace approximation for the posterior moment $E\{g_\phi(\underset{\sim}{\phi})|\underset{\sim}{y}\}$ is given by,

$$\hat{E}\{g_\phi(\underset{\sim}{\phi})|\underset{\sim}{y}\} = \left\{\frac{\det D^2 h^*_\phi(\hat{\underset{\sim}{\phi}}^*)}{\det D^2 h_\phi(\hat{\underset{\sim}{\phi}})}\right\}^{-1/2} \exp\{-n[h^*_\phi(\hat{\underset{\sim}{\phi}}^*) - h_\phi(\hat{\underset{\sim}{\phi}})]\} \qquad (4)$$

where D^2h_ϕ and $D^2h^*_\phi$ are the Hessian matrix for h_ϕ and h^*_ϕ, and $\hat{\underset{\sim}{\phi}}$ and $\hat{\underset{\sim}{\phi}}^*$ maximize $-h_\phi$ and $-h^*_\phi$, respectively.

The vector of first derivatives of $h_\phi(\underset{\sim}{\phi})$ with respect to ϕ_j, $j = 1,2,\ldots,m$ is given by,

$$\frac{\partial h_\phi(\underset{\sim}{\phi})}{\partial \underset{\sim}{\phi}} = \left(\frac{\partial h_\phi}{\partial \phi_1}, \ldots, \frac{\partial h_\phi}{\partial \phi_m}\right) = \left(\frac{\partial h_\theta(B\underset{\sim}{\phi})}{\partial \underset{\sim}{\theta}}\right)B,$$

where $\partial h_\theta(B\underset{\sim}{\phi})/\partial\underset{\sim}{\theta} = (\partial h_\theta(B\underset{\sim}{\phi})/\partial\theta_1, \ldots, \partial h_\theta(B\underset{\sim}{\phi})/\partial\theta_m)$. The maximum $\hat{\underset{\sim}{\phi}}$ of $h_\phi(\underset{\sim}{\phi})$ is given from $\partial h_\phi(\underset{\sim}{\phi})/\partial\underset{\sim}{\phi} = \underset{\sim}{0}$. Since A is a mxm nonsingular matrix, we have $\hat{\underset{\sim}{\phi}} = A\hat{\underset{\sim}{\theta}}$; also, $\hat{\underset{\sim}{\phi}}^* = A\hat{\underset{\sim}{\theta}}^*$. Therefore, $h^*_\phi(\hat{\underset{\sim}{\phi}}^*) = h^*_\theta(\hat{\underset{\sim}{\theta}}^*)$ and $h_\phi(\hat{\underset{\sim}{\phi}}) = h_\theta(\hat{\underset{\sim}{\theta}})$.

The Hessian matrix for h_ϕ and h^*_ϕ are given by, $D^2h_\phi = B'(D^2h_\theta(B\underset{\sim}{\phi}))B$, and $D^2h^*_\phi = B'(D^2h^*_\theta(B\underset{\sim}{\phi}))B$, respectively, where B' is the transpose of the matrix B.

Since $\det B'(D^2h_\theta(B\underset{\sim}{\phi}))B = (\det B'B)(\det D^2h_\theta(B\underset{\sim}{\phi}))$ and $\det B'(D^2h^*_\theta(B\underset{\sim}{\phi}))B = (\det B'B)(\det D^2h^*_\theta(B\underset{\sim}{\phi}))$, we have,

$$\frac{\det D^2 h^*_\phi(\hat{\underset{\sim}{\phi}}^*)}{\det D^2 h_\phi(\hat{\underset{\sim}{\phi}})} \equiv \frac{\det D^2 h^*_\theta(\hat{\underset{\sim}{\theta}}^*)}{\det D^2 h_\theta(\hat{\underset{\sim}{\theta}})} \tag{5}$$

Thus, we conclude that (3) coincides with (4).

3. SOME EXAMPLES

3.1. *Ratio of Exponential Means*

Let $y_{11}, y_{12}, \ldots, y_{1n}$ be a random sample of size n of an exponential distribution with parameter ξ and let $y_{21}, y_{22}, \ldots, y_{2n}$ be a random sample of size n of an exponential distribution with parameter λ. The likelihood function for ξ and λ is given by,

$$\ell(\xi,\lambda|\underset{\sim}{y}) \propto (\xi\lambda)^{-n} \exp\{-n\bar{y}_1\xi^{-1} - n\bar{y}_2\lambda^{-1}\} \tag{6}$$

where $\xi,\lambda > 0$. The Jeffreys prior for ξ and λ (see for example, Box and Tiao, 1973) is given by $\pi(\xi,\lambda) \propto (\xi\lambda)^{-1}$, where $\xi > 0$ and $\lambda > 0$.

If the parameter of interest is given by the ratio ξ/λ, we consider the reparametrization $\theta_1 = \ell n\xi$ and $\theta_2 = \ell n\lambda$, where $\pi_\theta(\theta_1,\theta_2) \propto$ constant (from the Jeffreys prior for ξ and λ) where $-\infty < \theta_1,\theta_2 < \infty$ and $g_\theta(\theta_1,\theta_2) = e^{\theta_1-\theta_2}$ (the ratio ξ/λ).

From (3), we find the Laplace approximation,

$$\hat{E}\{g_\theta(\theta_1,\theta_2)|\underset{\sim}{y}\} = \frac{(n-1)^{n-3/2}(n+1)^{n+1/2}}{n^{2n-1}} \left(\frac{\bar{y}_1}{\bar{y}_2}\right) \tag{7}$$

Considering the reparametrization $\phi_1 = \theta_1-\theta_2$ and

$\phi_2 = \theta_1+\theta_2$, we also have a locally uniform prior for ϕ_1 and ϕ_2 and (from (4)) the same Laplace approximation (7) where $g_\phi(\phi_1,\phi_2) = e^{\phi_1}$ (the ratio ξ/λ).

3.2. *Normal Distribution*

Let $y_1,y_2,\ldots,y_n$ be a random sample of size n of a normal distribution with mean μ and variance σ^2. The likelihood function for μ and σ is given by,

$$\ell(\mu,\sigma|\underset{\sim}{y}) \propto \sigma^{-n} \exp\{-\frac{n(\mu-\bar{y})^2}{2\sigma^2} - \frac{(n-1)s^2}{2\sigma^2}\} \qquad (8)$$

where $n\bar{y} = \sum_{i=1}^{n} y_i$ and $(n-1)s^2 = \sum_{i=1}^{n} (y_i-\bar{y})^2$.

The Jeffreys prior for μ and σ (see for example, Box and Tiao, 1973) is given by $\pi(\mu,\sigma) \propto \sigma^{-1}$ where $-\infty<\mu<\infty$ and $\sigma > 0$.

In this original parametrization, the Laplace approximation for the posterior moment of $g(\mu,\sigma^2) = \sigma^2$ is given by,

$$\hat{E}\{\sigma^2|\underset{\sim}{y}\} = \frac{es^2(n-1)^{(n-2)/2}}{(n+1)^{(n-2)/2}} \qquad (9)$$

Considering the reparametrization $\theta_1 = \mu$ and $\theta_2 = \ell n\sigma$, we have (from the Jeffreys prior for μ and σ) a locally uniform prior for θ_1 and θ_2. In this parametrization, the Laplace approximation for the posterior moment of $g_\theta(\theta_1,\theta_2) = e^{2\theta_2}$ (the variance σ^2), is given by,

$$\hat{E}\{g_\theta(\underset{\sim}{\theta})|\underset{\sim}{y}\} = \frac{es^2(n-1)(n-2)^{n/2-2}}{n^{n/2-1}} \qquad (10)$$

In the same way, if we consider a linear transformation given by $\phi_1 = \mu$ and $\phi_2 = \ell n \sigma^2$, that is, $\phi_1 = \theta_1$ and $\phi_2 = 2\theta_2$, we also have a locally uniform prior for ϕ_1 and ϕ_2 and the same Laplace approximation (10) for the posterior moment of $g_\phi(\phi_1, \phi_2) = e^{\phi_2}$.

In table 1, we have a numerical illustration with $n = 20$, $\bar{y} = 3$ and $s^2 = 2$, where the exact posterior moment is given by $E\{\sigma^2 | \underset{\sim}{y}\} = (n-1)s^2/(n-3)$. We observe very accurate Laplace approximation in the parametrization $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ or in the invariant reparametrization $\underset{\sim}{\phi} = (\phi_1, \phi_2)$.

TABLE 1

LAPLACE APPROXIMATION FOR $E\{\sigma^2 | \underset{\sim}{y}\}$

(Percentage error in parenteses)

EXACT	2.2353	(0%)	
(μ, σ)	2.2087	(1.20%)	(9)
(θ_1, θ_2) or (ϕ_1, ϕ_2)	2.2232	(0.54%)	(10)

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, U.K.

. ANSCOMBE, F.J. (1964) - "Normal likelihood functions". Ann. Inst. Stat. Math. 16, 1-19.

. BOX, G.E.P.; TIAO, G.C. (1973) - *Bayesian Inference in Statistical Analysis*, Addison-Wesley.

. 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, U.S.A.

. SPROTT, D.A. (1973) - "Normal likelihoods and their relation to large sample theory of estimation", Biometrika, 60, 457-465.

. SPROTT, D.A. (1980) - "Maximum likelihood in small samples: estimation in the presence of nuisance parameters", Biometrika, 67, 515-523.

. 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, U.S.A.