行政院國家科學委員會專題研究計畫　成果報告

後驗分配的近似計算(2/2)

計畫類別：個別型計畫
計畫編號：NSC93-2118-M-004-001-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：國立政治大學統計學系

計畫主持人：翁久幸

報告類型：完整報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 94 年 10 月 18 日

Approximate Computations for Posterior Distributions
NSC93-2118-M-004-001
92.8.1-94.7.31

Ruby C. Weng
Department of Statistics, National Chengchi University

October 18, 2005

Abstract

This project describes a method for approximating posterior expectations of functions of the parameter. First the posterior density of a data dependent transformation Z_t of the parameter is expressed as a form close to a normal density. Next, a version of Stein's Identity is applied to the posterior distribution to obtain posterior moments of Z_t. Then the results are converted to derive second-order approximations to posterior expectations of functions (not necessarily positive) of the parameter.

Key words: maximum likelihood estimator; posterior distributions; Stein's identity.

1 Introduction

Let $g(\theta)$ be a smooth function on the parameter space Θ. The estimation of the posterior mean of $g(\theta)$, given a sample of observations $x^{(t)}$, requires integraion over Θ of the form

$$E_\xi^t[g(\theta)] = E_\xi[g(\theta)||x^{(t)}] = \frac{\int_\Theta g(\theta)e^{\ell_t(\theta)}\xi(\theta)d\theta}{\int_\Theta e^{\ell_t(\theta)}\xi(\theta)d\theta}, \tag{1}$$

where ℓ_t is the log-likelihood function. If the likelihood function has a dominant mode, Laplace method will be suitable for approximating the integrals. Many authors have applied Laplace method to find approximations to the ratios of integrals in (1). For example, Lindley [2] derived the second order approximation for the integral.

Tierney and Kadane [5] applied the Laplace method in a special form in which g is assumed to be positive, the integrand of the numerator in (1) is expressed as $\exp[\ell_t(\theta) + \log g(\theta) + \log \xi(\theta)]$ (called *fully exponential* Laplace approximations) and is expanded at the mode of the integrand itself, rather than at the posterior mode. For a general function g (possibly non-positive), Tierney, Kass, and Kadane [6] obtained a second order expansion of the posterior expectation by applying the fully exponential method to approximate the moment generating function $E^t_\xi[\exp(sg(\theta))]$ and then differentiating (called the *MGF* method).

In this project we present a method, based on a version of Stein's Identity, for the problem of estimating the posterior mean of a smooth function of the parameter. First the posterior density of a data dependent transformation Z_t (2) of the parameter is converted into a form close to a normal density. Next, a version of Stein's Identity is applied to the posterior distribution to obtain posterior moments of Z_t. Then the results are converted to derive second-order approximations to posterior expectations of functions (not necessarily positive) of the parameter.

2 The Model and Stein's Identity

Let X_t be a random vector distributed according to a family of probability densities $p_t(x_t|\theta)$, where t is a discrete or continuous parameter and $\theta \in \Theta$, an open subset in $\Re^p$. Assume that the log-likelihood function, denoted by $\ell_t(\theta)$, is twice continuously differentiable with respect to θ. Throughout let $\hat{\theta}_t$ be a root of the likelihood equation satisfying $\nabla \ell_t(\hat{\theta}_t) = 0$, where ∇ indicates differentiation with respect to θ. Whenever such a root exists and $-\nabla^2 \ell_t(\hat{\theta}_t)$ is positively definite, we define Σ_t and the data dependent transformation Z_t as

$$\begin{aligned} \Sigma_t' \Sigma_t &= -\nabla^2 \ell_t(\hat{\theta}_t) \\ Z_t &= \Sigma_t(\theta - \hat{\theta}_t); \end{aligned} \tag{2}$$

otherwise, define Σ_t and Z_t arbitrarily (in a measurable way).

Consider a Bayesian model in which θ has a prior density ξ. Then the posterior density of θ given data x_t is $\xi_t(\theta) \propto e^{\ell_t(\theta)}\xi(\theta)$, and the posterior density of Z_t is

$$\zeta_t(z) \propto \xi_t(\theta(z)) \propto e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)}\xi(\theta), \tag{3}$$

where the relation of θ and z is given in (2). Now a Taylor's expansion gives

$$\ell_t(\theta) \quad = \quad \ell_t(\hat{\theta}_t) + \frac{1}{2}(\theta - \hat{\theta}_t)'\nabla^2\ell_t(\theta_t^*)(\theta - \hat{\theta}_t),$$

where θ_t^* lies between θ and $\hat{\theta}_t$. Let

$$u_t(\theta) = -\frac{1}{2}(\theta - \hat{\theta}_t)'[\nabla^2\ell_t(\hat{\theta}_t) - \nabla^2\ell_t(\theta_t^*)](\theta - \hat{\theta}_t),$$

it follows that

$$\ell_t(\theta) = \ell_t(\hat{\theta}_t) - \frac{1}{2}||z_t||^2 + u_t(\theta) \tag{4}$$

and (3) can be rewritten as

$$\zeta_t(z) \propto \phi_p(z) f_t(z), \tag{5}$$

where $f_t(z) = \xi(\theta(z))\exp[u_t(\theta)]$ and $\phi_p(z)$ denotes the standard p-variate normal density.

Throughout $\nabla\xi$ and $\nabla^2\xi$ denote the gradient and Hessian of ξ with respect to θ, ∇f and $\nabla^2 f$ the gradient and Hessian of f with respect to Z, and E_ξ^t and V_ξ^t the conditional expectation and variance given data x_t.

Stein's Identity Let Φ_p denote the standard p-variate normal distribution and write

$$\Phi_p h = \int h d\Phi_p$$

for functions h for which the integral is finite. Next let Γ denote a finite signed measure of the form $d\Gamma = f d\Phi_p$, where f is a real-valued function defined on $\Re^p$ satisfying $\Phi_p|f| = \int |f| d\Phi_p < \infty$. For $k > 0$, denote H_k as the collection of all measurable functions $h: \Re^p \to \Re$ for which $|h(z)|/b \leq 1 + ||z||^k$ for some $b > 0$ and $H = \cup_{k \geq 0} H_k$. Given $h \in H_k$, let $h_0 = \Phi_p h$, $h_p = h$,

$$h_j(y_1, ..., y_j) = \int_{\Re^{p-j}} h(y_1, ..., y_j, w)\Phi_{p-j}(dw), \tag{6}$$

and

$$g_j(y_1, ..., y_p) = e^{\frac{1}{2}y_j^2} \int_{y_j}^{\infty} [h_j(y_1, ..., y_{j-1}, w) - h_{j-1}(y_1, ..., y_{j-1})]e^{-\frac{1}{2}w^2} dw, \tag{7}$$

for $-\infty < y_1, ..., y_p < \infty$ and $j = 1, ..., p$. Then let $Uh = (g_1, ..., g_p)^T$. Note that U may be iterated. Let $Vh = (U^2h + U^2h')/2$, where U^2h is the $p \times p$ matrix whose j-th column is Ug_j and g_j is as in (7). Then Vh is a symmetric matrix. For example, for $z \in \Re^p$, if $h(z) = z_1$, then $Uh(z) = (1, 0, ..., 0)^T$ and if $h(z) = ||z||^2$, then $Uh(z) = z$.

Lemma 2.1 *(Stein's Identity) Let r be a nonnegative integer. Suppose that $d\Gamma = fd\Phi_p$ as above, where f is a differentiable function on $\Re^p$, for which*

$$\int_{\Re^p} |f|d\Phi_p + \int_{\Re^p} (1 + ||z||^r)||\nabla f(z)||\Phi_p(dz) < \infty,$$

then

$$\Gamma h = \Gamma 1 \cdot \Phi_p h + \int_{\Re^p} (Uh(z))^T \nabla f(z) \Phi_p(dz),$$

for all $h \in H_r$. If $\partial f/\partial z_j$, $j = 1, ..., p$, are differentiable, and

$$\int_{\Re^p} (1 + ||z||^r)||\nabla^2 f(z)||\Phi_p(dz) < \infty,$$

then

$$\Gamma h = \Gamma 1 \cdot \Phi_p h + \Phi_p(Uh)^T \int_{\Re^p} \nabla f(z)\Phi_p(dz) + \int_{\Re^p} tr[(Vh(z))\nabla^2 f(z)]\Phi_p(dz),$$

for all $h \in H_r$.

Observe from (5) that the posterior distribution of Z_t is of a form suitable for Stein's Identity. Let B_t denote the event $\{\nabla \ell_t(\hat{\theta}_t) = 0, -\nabla^2 \ell_t(\hat{\theta}_t)$ is positively definite$\}$. Suppose that ξ has a compact support $\Theta_1 \in \Theta$ and $\nabla \xi$ is continuous. Then, $||\nabla \xi||$ is bounded on Θ_1 and we can verify that

$$\int_{\Re^p} |f|d\Phi_p + \int_{\Re^p} (1 + ||z||^r)||\nabla f_t(z)||\Phi_p(dz) < \infty.$$

Hence, by Lemma 2.1

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + E_\xi^t\{[Uh(Z_t)]^T \frac{\nabla f_t(Z_t)}{f_t(Z_t)}\}, \quad (8)$$

$a.e.$ on B_t, for all $h \in H$. If also $\nabla^2 \xi$ is continuous, then similar arguments lead to

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + (\Phi_p U h)^T E_\xi^t[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}] + E_\xi^t\{\mathrm{tr}[Vh(Z_t)\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}]\} \quad (9)$$

$a.e.$ on B_t, for all $h \in H$.

3 Main Results

In this section we present approximations of posterior moments of Z_t and use it to derive posterior means and variances of $g(\theta)$.

Lemma 3.2 *If $h(z) = z_i z_j$, $1 \leq i \leq j \leq p$, then*
(i) $g_j(z) = z_i$ and $g_k(z) = 0$ for $k \neq j$
(ii) $\Phi U h = (0, \cdots, 0)^T$
(iii) $tr[Vh(z)\frac{\nabla^2 f_t(z)}{f_t(z)}] = [\frac{\nabla^2 f_t(z)}{f_t(z)}]_{ij}$.

$$E_\xi^t(\frac{\nabla \xi}{\xi}) = \frac{\nabla \hat{\xi}}{\hat{\xi}} + O(t^{-1}) \quad \text{and} \quad E_\xi^t(\frac{\nabla^2 \xi}{\xi}) = \frac{\nabla^2 \hat{\xi}}{\hat{\xi}} + O(t^{-1}). \tag{10}$$

Theorem 3.1

$$\begin{aligned} (i) E_\xi^t Z_t &= (\Sigma_t^T)^{-1}[(\frac{\nabla \hat{\xi}}{\hat{\xi}}) + \frac{1}{2}U] + O(t^{-3/2}), \\ (ii) V_\xi^t Z_t &= I_p + (\Sigma_t^T)^{-1}[(\frac{\nabla^2 \hat{\xi}}{\hat{\xi}}) + (\frac{\nabla \hat{\xi}}{\hat{\xi}})(\frac{\nabla \hat{\xi}^T}{\hat{\xi}}) + W]\Sigma_t^{-1} + O(t^{-2}), \end{aligned}$$

where U is a vector and W a matrix involving higher order derivatives of ℓ.

4 Applications

4.1 Linkage example

Here we consider an example presented in Rao [3] and reexamined by Tanner and Wong [4] and references therein. From a genetic linkage model, it is believed that 197 animals are distributed multinomially into four categories, $y = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$, with cell probabilities specified by $(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4})$.

Tanner and Wong [4] also consider a second version of the sata in which the sample size is reduced by a factor of 10, $y = (125, 18, 20, 34)$. As suggested in their paper, here we choose the uniform prior for $\theta \in (0, 1)$ and assess the performance of our method using both the large sample and small sample data. Table 1 reports the exact posterior means and variances of θ (carried out by matlab), and the approximations using our approach.

Table: Linkage example

	Large sample		Small sample	
Method	posterior mean	posterior variance	posterior mean	posterior variance
Exact	0.6228	0.0026	0.5704	0.0225
Stein	0.6233	0.0028	0.5615	0.0220

5 Conclusions

In conclusion, we use Stein's Identity to approximate posterior moments of a suitably normalized quantity. These moments are useful in the evaluation of posterior means and variances of $g(\theta)$. Unlike Laplace method (for positive function), ours requires third derivatives of ℓ_t, but we need only posterior mode for all g. Some formulas presented here are new, while some agree with results in earlier approaches such as Johnson [1] and Tierney, Kass, and Kadane [6].

References

[1] R. Johnson. Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.*, 41:851–864, 1970.

[2] D. V. Lindley. The use of prior probability distributions in statistical inference and decisions. *Proc. 4th. Berkeley Symp.*, 1:453–468, 1961.

[3] C. R. Rao. *Linear Statistical Inference and its Applications*. John Wiley, New York, Second edition, 2001.

[4] M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of American Statistical Association*, 82:528–540, 1987.

[5] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of American Statistical Association*, 81:82–86, 1986.

[6] L. Tierney, R. E. Kass, and J. B. Kadane. Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of American Statistical Association*, 84:710–716, 1989.