行政院國家科學委員會專題研究計畫 成果報告

Stein's 方法在貝氏分析之應用
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2118-M-004-002-
執 行 期 間 ：96 年 08 月 01 日至 97 年 08 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 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 98 年 05 月 13 日

Applications of Stein's method in Bayesian analysis NSC96-2118-M-004-002 96.8.1-97.7.31

Ruby C. Weng
Department of Statistics, National Chengchi University

May 11, 2009

Abstract

This project describes applications of a version of Stein's Identity in Bayesian asymptotics. We show that the use of Stein's Identity provides an alternative to traditional Laplace method for obtaining approximations of the marginal posterior densities.

Key words: Laplace method; posterior distributions; Stein's identity.

1 Introduction

Let $g(\theta)$ be a smooth function on the parameter space Θ. We are interested in the estimation of the posterior mean of $g(\theta)$, given a sample of observations $x^{(t)}$; that is,

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

where ℓ_t is the log-likelihood function and ξ the prior. Nowadays, modern computing techniques like Markov chain Monte Carlo and importance sampling have made many computations possible. Still, such methods are computational intensive and the sampling schemes vary from distribution to distribution. It is therefore of importance to have good analytic approximations which are simpler to compute. A traditional analytic approach to this problem (1) starts from a Taylor series expansion at the maximum likelihood estimator (or at the modes of the integrands), proceeds from there to develop expansions on both the numerator and denominator, and then obtains approximations by formal division of the two series. For example, Johnson [1, 2] derived expansions associated with posterior distribution of some pivotal quantity; Lindley [3, 4] and Mosteller and Wallace [5] obtained second order approximations for the integral by applying standard Laplace method to both numerator and denominator and taking the ratio. Tierney and Kadane [6] renewed interest in Laplace method by applying it in a special form in which g is assumed to be positive.

In related work, Woodroofe [10, 11] developed a version of Stein's Identity, which can be used to write posterior expectations in a particular form. Though this identity has a close Bayesian connection, the main focus of Woodroofe [10, 11] and some follow up work is on developing frequentist confidence regions. The first study of this tool in Bayesian context is Weng [7], which showed asymptotic posterior normality of nonhomogeneous Poisson model. Recently, Weng [8] further applied this identity for estimating predictive densities, and approximating marginal posterior distributions and posterior quantiles for individual parameters. Some formulas obtained are new, and some are shown to equivalent to the existing ones.

2 Stein's Identity and the Model

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. For $s > 0$, denote H_s as the collection of all measurable functions $h: \Re^p \to \Re$ for which $|h(z)|/b \leq 1 + ||z||^s$ for some $b > 0$. Given $h \in H_s$, let $h_0 = \Phi_p h$, $h_p = h$,

$$h_k(y_1, ..., y_k) = \int_{\Re^{p-k}} h(y_1, ..., y_k, w)\Phi_{p-k}(dw), \quad (2)$$

$$g_k(y_1, ..., y_p) = e^{\frac{1}{2}y_k^2} \int_{y_k}^{\infty} [h_k(y_1, ..., y_{k-1}, w) - h_{k-1}(y_1, ..., y_{k-1})]e^{-\frac{1}{2}w^2} dw, \quad (3)$$

for $-\infty < y_1, ..., y_p < \infty$ and $k = 1, ..., p$. Then let $Uh = (g_1, ..., g_p)^T$ and $Vh = (U^2h + U^2h^T)/2$, where U^2h is the $p \times p$ matrix whose k-th column is Ug_k and g_k is as in (3). 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 $Uh(z) = z$. Simple calculations by taking $f(z)$ in Lemma 2.1

below as z_i and $z_i z_j$ yield

$$\Phi_p(Uh) = \int_{\Re^p} z h(z) \Phi_p(dz), \tag{4}$$

$$\Phi_p(U^2 h) = \int_{\Re^p} \frac{1}{2}(zz^T - 1) h(z) \Phi_p(dz). \tag{5}$$

Lemma 2.1 *(Stein's Identity) Let r be a nonnegative integer. Suppose that f is a differentiable function on $\Re^p$, and*

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

then

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

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

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

for all $h \in H_r$.

The model 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 $\Re^p$. Consider a Bayesian model in which θ has a prior density ξ which is twice differentiable in $\Re^p$ and vanishes off of Θ. Assume that the log-likelihood function $\ell_t(\theta)$ is twice differentiable with respect to θ. Let B_t denote the set of sample points for which the maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla \ell_t(\hat{\theta}_t) = 0$, where ∇ indicates differentiation with respect to θ; therefore, $-\nabla^2 \ell_t(\hat{\theta}_t)$ is positive definite in B_t. The expressions for posterior expansions in (11) and (12) below are valid on B_t.

The model 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 $\Re^p$. Consider a Bayesian model in which θ has a prior density ξ which is twice differentiable in $\Re^p$ and vanishes off of Θ. Assume that the log-likelihood function $\ell_t(\theta)$ is twice differentiable with respect to θ. Let B_t denote the set of sample

points for which the maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla \ell_t(\hat{\theta}_t) = 0$, where ∇ indicates differentiation with respect to θ; therefore, $-\nabla^2 \ell_t(\hat{\theta}_t)$ is positive definite in B_t. The expressions for posterior expansions in (11) and (12) below are valid on B_t.

Define Σ_t and Z_t as

$$\Sigma_t^T \Sigma_t = -\nabla^2 \ell_t(\hat{\theta}_t), \tag{6}$$

$$Z_t = \Sigma_t(\theta - \hat{\theta}_t). \tag{7}$$

Then the posterior density of θ given data x_t is $\xi_t(\theta) \propto \exp(\ell_t(\theta))\xi(\theta)$, and the posterior density of Z_t is

$$\zeta_t(z) \propto \xi_t(\theta(z)) \propto \exp[\ell_t(\theta) - \ell_t(\hat{\theta}_t)]\xi(\theta), \tag{8}$$

where the relation of θ and z is given in (7). Now define

$$u_t(\theta) = \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \frac{1}{2}||z_t||^2. \tag{9}$$

So, (8) can be rewritten as

$$\zeta_t(z) \propto \phi_p(z) f_t(z), \tag{10}$$

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

Observe that the posterior distribution of Z_t in (10) is of a form suitable for Stein's Identity. Since ξ is twice differentiable in $\Re^p$ and vanishes off of Θ, $f_t(z)(= \xi(\theta(z))\exp[u_t(\theta)])$ also has the properties. So, by Lemma 2.1,

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + E_\xi^t\{[Uh(Z_t)]^T \frac{\nabla f_t(Z_t)}{f_t(Z_t)}\}, \tag{11}$$

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + (\Phi_p U h)^T E_\xi^t[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}] + E_\xi^t\{\mathrm{tr}[Vh(Z_t)\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}]\}. \tag{12}$$

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 posterior expectation and variance given data x_t. Some calculations are useful for later reference.

$$\frac{\nabla f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}[\frac{\nabla\xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta)], \tag{13}$$

$$\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}[\frac{\nabla^2\xi}{\xi} + \frac{\nabla\xi}{\xi}\nabla u_t^T + \nabla u_t \frac{\nabla\xi^T}{\xi} + \nabla^2 u_t + \nabla u_t \nabla u_t^T]\Sigma_t^{-1}, \tag{14}$$

where by (9) we can derive

$$\nabla u_t(\theta) = \nabla \ell_t(\theta) - \nabla^2 \ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t), \quad (15)$$

$$\nabla^2 u_t(\theta) = \nabla^2 \ell_t(\theta) - \nabla^2 \ell_t(\hat{\theta}_t). \quad (16)$$

3 Marginal Posterior Distributions

All asymptotic posterior expansions in this section are valid for sample points which lie on B_t (the set in which maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla \ell_t(\hat{\theta}_t) = 0$; see Section 2) and satisfy the following lemma.

Lemma 3.2 *Let $M_t(r; r_1, ..., r_p)$ denote rth joint posterior moments of Z_t with $r > 0$; that is, $M_t(r; r_1, ..., r_p) = E^t_\xi h(Z_t)$, where $h(z) = \prod_{i=1}^p z_i^{r_i}$ with $\sum r_i = r$. Then*
(i) $E^t_\xi h(Z_t) = O(t^{-1/2})$ for odd r;
(ii) $E^t_\xi h(Z_t) = \Phi h + O(t^{-1})$ for even r.

The above lemma is well known and we state it here for later use. The proof is in, for instance, Johnson [2]. We can also establish it using Stein's Identity.

Recall that X_t is a random vector from $p_t(x_t|\theta)$, where θ is chosen according to the prior density ξ. Let θ_0 denote the true underlying parameter. All asymptotic posterior expansions below are valid for sample points which lie on B_t (see Section 2) and satisfy the following conditions:

(C0) $\lim\limits_{t\to\infty} t^{-1}\nabla^2 \hat{\ell}_t$ is positive definite,

(C1) $t^{-1}\hat{\ell}_t^{(k)} = O(1)$ for $k > 0$,

(C2) $t^2 E^t_\xi[a(\theta) - a(\hat{\theta}_t) - \sum_{s=1}^3 (s!)^{-1} a^{(s)}(\hat{\theta}_t; \theta - \hat{\theta}_t)]^2 = O(1)$,

(C3) $E^t_\xi ||Z_t||^n = O(1)$ for $n > 0$,

where $a(\theta)$ is $\ell_i^{(1)}$ or $\ell_{ij}^{(2)}$, $a^{(s)}(\hat{\theta}_t; \theta - \hat{\theta}_t) = \sum_{i_1\cdots i_s} a^{(s)}_{i_1\cdots i_s}(\hat{\theta}_t)\delta_{i_1}\cdots\delta_{i_s}$, and $O(1)$ means convergence of a sequence of real numbers. So, the integrand in (C2) is square of remainder terms in a Taylor expansion. Condition (C1) is easy to check. Conditions (C1) and (C2) can be guaranteed by assuming some tail properties of ℓ_t and the local behavior that $\ell^{(k)}(\theta)$ is bounded in a small neighborhood of θ_0.

In the following we prove Lemma 3.2 using Stein's Identity. It should always be remembered that the derivatives of f_t are in (13) and (14), and ∇u_t and $\nabla^2 u_t$ are in

(15) and (16). First note that if h is a polynomial of order r, Uh and Vh are of orders $r-1$ and $r-2$ (see Weng and Woodroofe [9, Lemma 8]); and that by (4), $\Phi_p Uh = 0$ for even r. Then, by Taylor expansions,

$$[\nabla u_t(\theta)]_i = \frac{1}{2}\delta_t^T D_i \delta_t + (\text{Rem}_1) = \frac{1}{2} Z_t^T V_i Z_t + (\text{Rem}_1), \tag{17}$$

$$[\nabla^2 u_t(\theta)]_{ij} = [D_i]_{j.}\Sigma_t^{-1} Z_t + \frac{1}{2}\sum_{k,s} \hat{\ell}^{(4)}_{ijks}[Z_t^T(\Sigma_t^T)^{-1} e_k e_s^T \Sigma_t^{-1} Z_t] + (\text{Rem}_2), \tag{18}$$

where $(\text{Rem}_1) = (1/6)\sum_{jks}\hat{\ell}^{(4)}_{ijks}\delta_{tj}\delta_{tk}\delta_{ts} + (1/24)\sum_{jksq}\ell^{(5)}_{ijksq}(\tilde{\theta}_t)\delta_{tj}\delta_{tk}\delta_{ts}\delta_{tq}$, $\tilde{\theta}_t$ lies between θ and $\hat{\theta}_t$, and (Rem_2) has a similar form. So, $E^t_\xi\{[Uh(Z_t)]_i \text{Rem}_1\}$ is bounded by (C1)-(C3) and Cauchy-Schwartz inequality.

Next, let q_k denote Hermite polynomials, given by $q_k(z)\phi(z) = (-d/dz)^k\phi(z)$. For instance, for $k = 1, ..., 4$ the Hermite polynomials are $q_1(z) = z$, $q_2(z) = z^2 - 1$, $q_3(z) = z^3 - 3z$, and $q_4(z) = z^4 - 6z^2 + 3$.

Theorem 3.1 *Take $h^*(z_{tp})$ in (??) as the indicator function $1(z_{tp} \le w)$, where $w \in \Re$. Then, the marginal posterior distribution for the individual parameter θ_p is*

$$\begin{aligned} & P^t_\xi(\theta_p \le a) = P^t_\xi(Z_{tp} \le w) \\ = \; & \Phi(w) - \sum_{i=1, i\neq 5}^{6} \frac{1}{i!} q_{i-1}(w)\phi(w) E^t_\xi(q_i(Z_{tp})) + O(t^{-3/2}), \end{aligned} \tag{19}$$

where $w = [\Sigma_t]_{pp}(a - \hat{\theta}_{tp})$.

By taking derivative of (19) with respect to a, we obtain the marginal posterior density

$$\xi^t_p(a) = [\Sigma_t]_{pp}\{\phi(w) + \sum_{i=1, i\neq 5}^{6}\frac{1}{i!} q_i(w)\phi(w) E^t_\xi(q_i(Z_{tp})) + O(t^{-3/2})\}. \tag{20}$$

Observe that no renormalization is needed for this approximation as $\int_\Re q_i(w)\phi(w)dw = 0$.

References

[1] R. Johnson. An asymptotic expansion for posterior distributions. *Ann. Math. Statist.*, 38:1899–1906, 1967.

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

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

[4] D. V. Lindley. Approximate bayesian methods. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. S. (Eds.), editors, *Bayesian Statistics*. University Press, 1980.

[5] F. Mosteller and D. L. Wallace. *Inference and Disputed Authorship: The Federalist Papers*. Addison-Wesley, Reading, Mass., 1964.

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

[7] R. C. Weng. On Stein's identity for posterior normality. *Statistica Sinica*, 13:495–506, 2003.

[8] R. C. Weng. Stein's identity for bayesian inference. *manuscript*, 2006.

[9] R. C. Weng and M. Woodroofe. Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels. *Statistica Sinica*, 10:693–713, 2000.

[10] M. Woodroofe. Very weak expansions for sequentially designed experiments: linear models. *The Annals of Statistics*, 17:1087–1102, 1989.

[11] M. Woodroofe. Integrable expansions for posterior distributions for one-parameter exponential families. *Statistica Sinica*, 2:91–111, 1992.

Report on attending the 7th World Congress in Probability and Statistics Singapore

The 7th World Congress in Probability and Statistics was jointly sponsored by the Bernoulli Society and the Institute of Mathematical Statistics, two of the major international statistical societies. This year the conference was held in Singapore from July 14 to 19, 2008. This meeting is a major international event in probability and statistics held every four years. It covers a wide range of topics and features the latest scientific developments in the fields of probability and statistics and their applications.

I arrived on July 13 and stayed for 6 days. I presented my recent work on Stein's Identity and its applications in Bayesian analysis. My talk was scheduled with some other Bayesian studies so that I got a good chance to see other Bayesian related work. I attended several other presentations and was impressed by some interesting talks such as "A picture is worth a thousand numbers: communicating uncertainties following statistical analysis" by David Spiegelhalter, "Probability and statistics in internet information retrieval" by Zhi-Ming Ma, and Luke Tierney's talk on statistical computing.

I also browsed books displayed in the book stand and purchased one book relevant to my research.