Journal of Statistical Planning and Inference 138 (2008) 4068 – 4080

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Asymptotic posterior normality for multiparameter problems

Ruby C. Weng, Wen-Chi Tsai*

Department of Statistics, National Chengchi University, Wen-San, Taipei 11623, Taiwan

ARTICLE INFO

Article history:
Received 8 May 2006
Received in revised form
21 October 2007
Accepted 7 March 2008
Available online 28 March 2008

Keywords:
Maximum likelihood estimator
Multiparameter cases
Posterior normality
Stein's identity
Stochastic processes

ABSTRACT

For asymptotic posterior normality in the one-parameter cases, Weng [2003. On Stein's identity for posterior normality. Statist. Sinica 13, 495–506] proposed to use a version of Stein's Identity to write the posterior expectations for functions of a normalized quantity in a form that is more transparent and can be easily analyzed. In the present paper we extend this approach to the multi-parameter cases and compare our conditions with earlier work. Three examples are used to illustrate the application of this method.

1. Introduction

The study of asymptotic posterior normality can be traced back to the time of Laplace and it has attracted the attention of many authors. A conventional approach to such problems starts from a Taylor series expansion of the loglikelihood function around the maximum likelihood estimator (MLE) and proceeds from there to develop expansions that have standard normal as a leading term and hold in probability or almost surely, given the data. For contributions of previous work, see, for instance, Cam (1953), Walker (1969), and Johnson (1970) for i.i.d. observations; Heyde and Johnstone (1979), Chen (1985), and Sweeting and Adekola (1987) for stochastic processes. In these papers there are three basic conditions, which regard information growth, information continuity, and the tail behavior of the likelihood function. More specifically, the information growth assumes that the norm of the observed information matrix goes to infinity; the information continuity involves assuming that the information function is smooth over either a fixed or a shrinking neighborhood of the maximum likelihood estimate (or the true underlying parameter value); and the tail behaviors essentially concern with how fast the loglikelihood values decline outside the specified neighborhood. Earlier papers consider fixed neighborhoods; later, Chen (1985) introduces the idea of shrinking neighborhood, but does not specify the shrinking rate; Sweeting (1992) studies a two-parameter nonhomogeneous Poisson process, gives explicit shrinking rate, and provides some workable conditions relating to the tail behavior of the likelihood. In fact, these two papers show that in practice the first two conditions are relatively easy to check while the third one for nonlocal behavior of the likelihood is more complicated.

Recently, Weng (2003) proposed an alternative approach for posterior normality of stochastic processes in the one-parameter cases. This approach begins with a suitable parameter transformation Z_t, then for any bounded measurable function h, a version of Stein's Identity is employed to isolate the remainder terms of the posterior expectations of $h(Z_t)$ so that the posterior normality becomes more transparent and can be easily established. For a detailed account of Stein's Identity, we refer readers to Woodroofe (1989, 1992) and Woodroofe and Coad (1997). However, due to some technical difficulties (see Section 5), the conditions in Weng (2003) cannot be directly modified for k-parameter models. The purposes of this paper are to extend this method to multiparameter problems and provide comparisons of the conditions with earlier work. In the following section we introduce

* Corresponding author. Fax: +886 2 29398024.
E-mail address: wctsai@nccu.edu.tw (W.-C. Tsai).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.03.034

the model. Section 3 reviews Stein's Identity and gives some preliminary results. Model conditions and the main theorems are provided in Section 4, where the asymptotic results are established first by assuming that the prior has a compact support and is continuously differentiable on $\mathfrak{R}^k$, and then extending to more general priors. Section 5 compares our conditions with Weng (2003) and Sweeting (1992). Finally, we use three examples in Section 6 to illustrate the application of this approach: Section 6.1 gives a bivariate normal model in which our conditions hold, but Sweeting's (1992) fail; Sections 6.2 and 6.3 consider the exponential family and a nonhomogeneous Poisson model, respectively.

2. The model

Let X_t be a random vector distributed according to a family of probability densities $p_\theta^t(x_t)$, where t is a discrete or continuous parameter and $\theta \in \Theta$, an open subset of $\mathscr{R}^k$. Let P_θ^t and E_θ^t be the associated probability measure and expectation of p_θ^t. Assume that the loglikelihood function $\ell_t(\theta)=\log p_\theta^t(x_t)$ is twice continuously differentiable with respect to θ. Denote $\nabla\ell_t(\theta)$ as the vector of first-order partial derivatives, and $\nabla^2\ell_t(\theta)$ as the matrix of second-order partial derivatives. Throughout let $\hat{\theta}_t$ be the MLE, satisfying $\nabla\ell_t(\hat{\theta}_t)=0$. Whenever such a root exists and $-\nabla^2\ell_t(\hat{\theta}_t)$ is positive definite, we define B_t and Z_t as

$$B_t' B_t = -\nabla^2 \ell_t(\hat{\theta}_t), \tag{1}$$

$$Z_t = B_t(\theta - \hat{\theta}_t); \tag{2}$$

otherwise, define B_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 $\xi_t(\theta) \propto \mathrm{e}^{\ell_t(\theta)}\xi(\theta)$, and the posterior density of Z_t is

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

where the relation of θ and z is given in (2). Let P_ξ^t and E_ξ^t denote the conditional probability and expectation given data x_t. Suppose that θ_0 is the true underlying parameter. The goal is to establish explicit conditions under which

$$P_\xi^t(Z_t \in B) \to \Phi_k(B) \quad \text{as } t\to\infty \text{ in } P_{\theta_0}^t\text{-probability}, \tag{4}$$

where B is any Borel set in $\mathfrak{R}^k$ and Φ_k is the standard k-variate normal distribution. In what follows, $P_{\theta_0}^t$ is abbreviated as P_{θ_0} for convenience.

3. Preliminary results

To begin, we review Stein's Identity. Let Γ denote a finite signed measure of the form $\mathrm{d}\Gamma=f\,\mathrm{d}\Phi_k$, where f is a real-valued function defined on $\mathscr{R}^k$ satisfying $\int |f|\,\mathrm{d}\Phi_k<\infty$. Next, write $\Phi_k h=\int h\,\mathrm{d}\Phi_k$ for functions h for which the integral is finite, and similarly write $\Gamma h=\int h\,\mathrm{d}\Gamma$. For $p\geqslant 0$, denote by H_p the collection of all measurable functions $h:\mathscr{R}^k\to\mathscr{R}$ for which $|h(z)|\leqslant c(1+\|z\|^p)$ for some $c>0$, and define $H=\bigcup_{p\geqslant 0}H_p$. Given $h\in H_p$, let $h_0=\Phi_k h$, $h_k=h$,

$$h_j(y_1,\ldots,y_j)=\int_{\mathfrak{R}^{k-j}} h(y_1,\ldots,y_j,w)\Phi_{k-j}(\mathrm{d}w),$$

$$g_j(y_1,\ldots,y_k)=\mathrm{e}^{(1/2)y_j^2}\int_{y_j}^{\infty}[h_j(y_1,\ldots,y_{j-1},w)-h_{j-1}(y_1,\ldots,y_{j-1})]\mathrm{e}^{-(1/2)w^2}\,\mathrm{d}w, \tag{5}$$

for $-\infty<y_1,\ldots,y_k<\infty$ and $j=1,\ldots,k$. Then let $Uh=(g_1,\ldots,g_k)'$. Note that $g_j(y_1,\ldots,y_k)=g_j(y_1,\ldots,y_j)$, which does not involve $y_{j+1},\ldots,y_k$.

Proposition 3.1 (*Stein's Identity*). *Let q be a nonnegative integer. Suppose that $\mathrm{d}\Gamma=f\,\mathrm{d}\Phi_k$, where f is differentiable on $\mathscr{R}^k$ for which*

$$\int_{\mathfrak{R}^k}|f|\,\mathrm{d}\Phi_k+\int_{\mathfrak{R}^k}(1+\|z\|^q)\|\nabla f(z)\|\Phi_k(\mathrm{d}z)<\infty.$$

Then,

$$\Gamma h=\Gamma 1\cdot\Phi_k h+\int (Uh(z))'\nabla f(z)\Phi_k(\mathrm{d}z), \tag{6}$$

for all $h\in H_q$.

Proof. See Woodroofe (1989, Proposition 1). □

Lemma 3.1. *If h is a bounded measurable function, then* $\|Uh(z)\| \leqslant c_0$ *for some* $c_0 > 0$ *and for all* $z \in \Re^k$. *Moreover, if* $h(z) = \|z\|^p$, $p \geqslant 1$, *then for some* $c_p > 0$

$$\|Uh(z)\| \leqslant c_p(1 + \|z\|^{p-1}).$$

Proof. See Woodroofe (1992) or Weng and Woodroofe (2000, Lemma 8). □

Some notations and calculations are needed. For converting the likelihood into a form close to normal, we first take a Taylor's expansion of $\ell_t(\theta)$ at $\hat{\theta}_t$,

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

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

$$u_t(\theta) = -\tfrac{1}{2}(\theta - \hat{\theta}_t)'[\nabla^2\ell_t(\hat{\theta}_t) - \nabla^2\ell_t(\theta_t^*)](\theta - \hat{\theta}_t), \tag{7}$$

it follows that

$$\ell_t(\theta) = \ell_t(\hat{\theta}_t) - \tfrac{1}{2}\|z_t\|^2 + u_t(\theta) \tag{8}$$

and (3) can be rewritten as

$$\zeta_t(z) \propto \phi_k(z) f_t(z), \tag{9}$$

where $f_t(z) = \xi(\theta(z))\exp[u_t(\theta(z))]$ and $\phi_k(z)$ denotes the standard k-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 $(\partial^2\ell_t/\partial\theta_i\,\partial\theta_j)(\eta^{ij})$ denotes the Hessian matrix of ℓ_t with its (i,j)-component evaluated at η^{ij}, respectively. More calculations are needed.

$$\frac{\nabla f_t(Z_t)}{f_t(Z_t)} = (B_t')^{-1}\left[\frac{\nabla\xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta)\right], \tag{10}$$

where from (8) we have

$$\nabla u_t(\theta) = \nabla\ell_t(\theta) - \nabla^2\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t). \tag{11}$$

Moreover, by Taylor's expansions,

$$\nabla u_t(\theta) = \left[\left(\frac{\partial^2\ell_t}{\partial\theta_i\,\partial\theta_j}(\theta^{*ij})\right) - \nabla^2\ell_t(\hat{\theta}_t)\right](\theta - \hat{\theta}_t),$$

where $\theta^{*il} = \theta^{*ir}$, $i, l, r = 1, \ldots, k$, lie between θ and $\hat{\theta}_t$. Therefore,

$$(B_t')^{-1}\nabla u_t(\theta) = \left\{I_k - (B_t')^{-1}\left[-\left(\frac{\partial^2\ell_t}{\partial\theta_i\,\partial\theta_j}(\theta^{*ij})\right)\right]B_t^{-1}\right\}Z_t. \tag{12}$$

From (9), the posterior distribution of Z_t is of a form suitable for Stein's Identity. The following two conditions on prior are needed for Proposition 3.2 and Theorem 4.1. However, in Theorems 4.2 and 4.3 the asymptotic posterior normality can be established for more general priors.

(P1) ξ is continuously differentiable on $\Re^k$.
(P2) ξ has a compact support $\Theta_\xi \subset \Re^k$.

In the proposition below, assume that there is a measurable $\hat{\theta}_t = \hat{\theta}_t(X_t)$ and let

$$D_t = \{\nabla\ell_t(\hat{\theta}_t) = 0,\ -\nabla^2\ell_t(\hat{\theta}_t) \text{ is positive definite}\}. \tag{13}$$

Proposition 3.2. *Let q be a nonnegative integer. Suppose that* ξ *satisfies* (P1) *and* (P2). *Then, for all* $h \in H_q$,

$$E_\xi^t[h(Z_t)] - \Phi h = E_\xi^t\left\{[Uh(Z_t)]'\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right\} \quad a.e.\ on\ D_t.$$

Proof. Write $\xi_t(\theta) = c_t \exp[\ell_t(\theta)]\xi(\theta)$ and $\zeta_t(z) = c_t^* \phi_k(z) f_t(z)$, where c_t and c_t^* are normalized constants which exist by (P1) and (P2). Taking f in Proposition 3.1 as $c_t^* f_t$, we have $\int c_t^* f_t \,\mathrm{d}\Phi_k = 1$. It then suffices to show that for fixed x_t

$$E_\xi^t \left\{ (1 + \|Z_t\|^q) \frac{\nabla f_t(Z_t)}{f_t(Z_t)} \right\} < \infty.$$

To see it, first expand the above integral by using (10), (11), and the relation

$$E_\xi^t \left[\frac{\|\nabla \xi(\theta)\|}{\xi(\theta)} \right] = \int_{\Theta_\xi} c_t \|\nabla \xi(\theta)\| \mathrm{e}^{\ell_t(\theta)} \,\mathrm{d}\theta.$$

Then (P1) and (P2) imply that the integrands involved are continuous and bounded over Θ_ξ so the desired result follows. $\square$

Thus, we have the simple consequence: *Suppose that* ξ *satisfies* (P1)–(P2). *Then, the necessary and sufficient condition for* (4) *to hold is*

$$E_\xi^t \left\{ [Uh(Z_t)]' \frac{\nabla f_t(Z_t)}{f_t(Z_t)} \right\} \to 0 \quad \text{in } P_{\theta_0}\text{-probability}.$$

4. Main results

We begin with some notation. Let D_t be the event defined in (13), $\xrightarrow{\mathrm{p}}$ denote convergence in P_{θ_0}-probability as $t \to \infty$, and

$$S_t = \{z : z = B_t(\theta - \hat{\theta}_t), \theta \in \Theta_\xi\}. \tag{14}$$

Next, for any $k \times k$ matrix A, the spectral norm is $\|A\|^2 = \lambda_{\max}(A'A)$. Finally, denote by $B(a; d)$ the k-dimensional open ball with radius d centered at a. Hereafter, we assume that the prior ξ satisfies the following condition:

(P3) There exist $\varepsilon_0 > 0$ and $\delta_0 > 0$ such that $\xi(\theta) > \varepsilon_0$ over $B(\theta_0; \delta_0)$.

The following three conditions are required for the likelihood. In brief, the first two conditions regard information growth and information continuity, respectively, and the last one concerns some integrability properties of $\exp[\ell_t(\theta) - \ell_t(\theta)]$ over S_t, which essentially involves the tail behavior of the likelihood.

(L1) $P_{\theta_0}(D_t^c) \to 0$, $\|B_t^{-1}\| \xrightarrow{\mathrm{p}} 0$, and $\hat{\theta}_t \xrightarrow{\mathrm{p}} \theta_0$ as $t \to \infty$.

(L2) There exists a sequence of positive constants $\{b_t\}$ increasing to ∞ such that $\sup_{\eta^{ij} \in \{\theta : \|z_t\| \leqslant b_t\}} \|I_k + (B_t')^{-1}(\partial^2 \ell_t / \partial\theta_i \partial\theta_j(\eta^{ij})) B_t^{-1}\| \xrightarrow{\mathrm{p}} 0$.

(L3) (i) Let b_t be as in (L2). There exist constants $r \geqslant 1$ and $c \geqslant 0$ such that for all $\theta \in \{\|z_t\| > b_t\} \cap \Theta_\xi$, $\|(B_t')^{-1} \nabla u_t(\theta)\| \leqslant c\|z_t\|^r$.

(ii) There exists a nonnegative function $g : \Re^+ \times \Re^k \to \Re$ for which, with P_{θ_0}-probability tending to 1 and $\forall \theta \in \Theta_\xi$, $[\ell_t(\hat{\theta}_t) - \ell_t(\theta)] \geqslant g(t, \theta)$, $m_t(\theta) \equiv (\det B_t)\|z_t\|^r \mathrm{e}^{-g(t,\theta)}$ are uniformly integrable in t, and $\int_{\Theta_\xi} m_t(\theta)\,\mathrm{d}\theta$ are uniformly bounded in t. Here the constant r is as in (i).

Note that the uniformly bounded condition of $\int_{\Theta_\xi} m_t(\theta)\,\mathrm{d}\theta$ in (L3)(ii) is guaranteed by the uniformly integrability, provided that Θ_ξ is bounded. The following facts related to conditions (L1)–(L3) will be used later.

(A) By rearranging (7) and (8) we obtain

$$\ell_t(\hat{\theta}_t) - \ell_t(\theta) = \tfrac{1}{2}\|z_t\|^2 - \tfrac{1}{2} z_t'[I_k + (B_t')^{-1}(\nabla^2 \ell_t(\theta_t^*)) B_t^{-1}] z_t.$$

So, if (L2) holds, there exist constants $c_1 > 0$ and $c_1' > 0$ such that

$$\sup_{\theta : \|z_t\| \leqslant c_1} [\ell_t(\hat{\theta}_t) - \ell_t(\theta)] \leqslant c_1' \quad \text{with } P_{\theta_0}\text{-probability tending to 1}. \tag{15}$$

(B) Condition (L3)(ii) ensures that for some $0 < M < \infty$,

$$\int_{S_t} \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z < M \quad \text{with } P_{\theta_0}\text{-probability tending to 1}, \tag{16}$$

$$\int_{S_t} \|z\| \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z < M \quad \text{with } P_{\theta_0}\text{-probability tending to 1}, \tag{17}$$

and

$$\int_{S_t\cap\{\|z\|>b_t\}} \|z\|^r e^{\ell_t(\theta)-\ell_t(\hat{\theta}_t)}\,dz \xrightarrow{p} 0, \tag{18}$$

where $b_t \to \infty$ as in (L2).

(C) As an alternative condition to (L3)(ii), (L3)(ii′) also implies (16)–(18).

(L3)(ii′) There exists a nonnegative function $g^* : \Re^k \to \Re$ for which, with P_{θ_0}-probability tending to 1 and $\forall\theta \in \Theta_\xi$, $[\ell_t(\hat{\theta}_t) - \ell_t(\theta)] \geqslant g^*(z_t)$ and $\int_{\Re^k} \|z\|^r e^{-g^*(z)} < \infty$.

Note that the results in the lemma and theorems below remain valid if the condition (L3)(ii) is replaced by (L3)(ii′), since we only need (16)–(18).

Lemma 4.2. *Let f_t be as in* (9) *and S_t as in* (14). *Suppose that ξ satisfies* (P1)–(P3). *We have* (a) *if conditions* (L1) *and* (L2) *hold, then there exists $C_1 > 0$ such that $\int_{S_t} \phi_k(z)f_t(z)\,dz > C_1$ with P_{θ_0}-probability tending to* 1; (b) *if* (L3)(ii) *holds, then there exists $C_2 > 0$ such that $\int_{S_t} \phi_k(z)f_t(z)\,dz < C_2$ with P_{θ_0}-probability tending to* 1.

Proof. Throughout the proof, we need the following expression from (8) and (9):

$$\begin{aligned}\int_{S_t} \phi_k(z)f_t(z)\,dz &= (2\pi)^{-k/2}\int_{S_t} \xi(\theta(z))e^{-(1/2)\|z\|^2+u_t(\theta)}\,dz\\ &= (2\pi)^{-k/2}\int_{S_t} \xi(\theta(z))e^{\ell_t(\theta)-\ell_t(\hat{\theta}_t)}\,dz.\end{aligned}$$

Now consider (a). Let δ_0 and ε_0 be as in (P3), and recall that (L2) implies (15). By assumptions $\|B_t^{-1}\| \xrightarrow{p} 0$ and $\hat{\theta}_t \xrightarrow{p} \theta_0$ in (L1), we have $P_{\theta_0}(\{\theta : \|z_t\| \leqslant c_1\} \subset B(\hat{\theta}_t; \delta/2) \subset B(\theta_0; \delta)) \to 1$. Therefore, with P_{θ_0}-probability tending to 1, we have $\xi(\theta(z)) > \varepsilon_0$ over $\|z_t\| \leqslant c_1$. So, (a) follows by taking $C_1 = (2\pi)^{-k/2}\varepsilon_0 e^{-c_1'} \int_{\|z\|\leqslant c_1} dz$.

Next, (b) follows by (16) and the boundedness of ξ (by (P1) and (P2)). □

From (8) and (9),

$$E_\xi^t\left(\frac{\|\nabla\xi\|}{\xi}\right) = \frac{\int_{S_t} \|\nabla\xi\|\phi_k(z)e^{u_t(\theta)}\,dz}{\int_{S_t} \phi_k(z)f_t(z)\,dz} = \frac{(2\pi)^{-k/2}\int_{S_t} \|\nabla\xi\|e^{\ell_t(\theta)-\ell_t(\hat{\theta}_t)}\,dz}{\int_{S_t} \phi_k(z)f_t(z)\,dz}.$$

Therefore, by boundedness of $\nabla\xi$, (L3)(ii), and the arguments in Lemma 4.2, we have the following corollary.

Corollary 4.1. *Under the same conditions as in Lemma* 4.2, *there exists $0 < C < \infty$ such that $E_\xi^t(\|\nabla\xi\|/\xi) \leqslant C$ with P_{θ_0}-probability tending to* 1.

Theorem 4.1. *Let h be any bounded measurable function. Suppose that the prior ξ satisfies* (P1)–(P3) *and the likelihood satisfies* (L1)–(L3). *Then, $E_\xi^t[h(Z_t)] \xrightarrow{p} \Phi h$.*

Proof. Throughout the proof, we note that Uh is bounded by Lemma 3.1, and ξ is bounded by (P1) and (P2). From (10) and Proposition 3.2, for a.e. on D_t,

$$\begin{aligned}E_\xi^t[h(Z_t)] - \Phi h &= E_\xi^t\left\{[Uh(Z_t)]'B_t'^{-1}\frac{\nabla\xi}{\xi}\right\} + E_\xi^t\{[Uh(Z_t)]'B_t'^{-1}\nabla u_t(\theta)\}\\ &= I_t + II_t \quad \text{say.}\end{aligned} \tag{19}$$

Since $P_{\theta_0}(D_t^c) \to 0$ by (L1), it suffices to show $(I_t + II_t) \xrightarrow{p} 0$. First, $I_t \xrightarrow{p} 0$ follows from Corollary 4.1 and the assumption $\|B_t^{-1}\| \xrightarrow{p} 0$ under (L1).

Next, consider II_t. Write

$$II_t = \frac{\int_{S_t} [Uh(z)]'B_t'^{-1}\nabla u_t(\theta)\phi(z)f_t(z)\,dz}{\int_{S_t} \phi(z)f_t(z)\,dz},$$

where by Lemma 4.2(a) the denominator is bounded below by some $C_1 > 0$; therefore, we need only to show that the numerator $\xrightarrow{p} 0$. Now, decompose the numerator into integrations over $\|z\| \leqslant b_t$ and $\|z\| > b_t$, and denote the corresponding integrals as $II_t^{(1)}$ and $II_t^{(2)}$, respectively. By (12), (P1)–(P2), and Lemma 3.1, there exists $C > 0$ such that

$$|II_t^{(1)}| \leqslant \int_{\|z\| \leqslant b_t} |[Uh(z)]' B_t'^{-1} \nabla u_t(\theta)| \xi(\theta(z)) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz$$
$$\leqslant C \left\{ \sup_{\theta : \|z\| \leqslant b_t} \left\| I_k - (B_t')^{-1} \left[-\left(\frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j} \right) (\theta^{*ij}) \right] B_t^{-1} \right\| \right\} \int_{\|z\| \leqslant b_t} \|z\| e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz.$$

So, $II_t^{(1)} \xrightarrow{p} 0$ by (L2) and (17). Finally, by (P1)–(P2), (L3)(i), and Lemma 3.1, there exists $C > 0$ such that

$$|II_t^{(2)}| \leqslant C \int_{S_t \cap \{\|z\| > b_t\}} \|z\|^r e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz,$$

which $\xrightarrow{p} 0$ by (18). □

Proposition 3.2 requires (P1) and (P2), which exclude priors such as uniform (not continuously differentiable on $\Re^k$) and normal (not having compact support). The next two theorems consider more general priors. The following notation is needed:

$$S_t^* \equiv \{z : z = B_t(\theta - \hat{\theta}_t), \theta \in \Theta\}. \tag{20}$$

Theorem 4.2. *Let ξ be a prior on Θ. Suppose the following conditions hold*:

(i) *there exists a sequence of priors $\{\xi_n, n \geqslant 1\}$ with supports Θ_{ξ_n} such that* (P1)–(P2) *are satisfied*, (P3) *is satisfied with the same ε_0 and δ_0*,
(ii) $\sup_{\theta \in \Theta} |\xi_n(\theta) - \xi(\theta)| \to 0$ *as $n \to \infty$*,
(iii) *the likelihood satisfies* (L1)–(L3), *but with Θ_ξ in* (L3) *replaced by Θ_{ξ_n}*,
(iv) *there exists $0 < M_* < \infty$, not depending on t, such that the likelihood satisfies the following*:

$$\int_{S_t^*} e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz \leqslant M_* < \infty \quad \textit{with } P_{\theta_0}\textit{-probability tending to } 1. \tag{21}$$

Then, for any bounded measurable h we have $E_\xi^t[h(Z_t)] \xrightarrow{p} \Phi h$.

Note that if Θ_{ξ_n} in (iii) above is replaced by Θ, then (16) holds with S_t replaced by S_t^*, which is exactly (21).

Proof. To start, notice that one can assume $|h| \leqslant 1$ without losing generality, and that by (i) and (iii) we can apply Theorem 4.1 to obtain that for each $n \geqslant 1$, $E_{\xi_n}^t h(Z_t) \xrightarrow{p} \Phi h$ as $t \to \infty$. Next, write

$$E_\xi^t h(Z_t) = \frac{\int_{S_t^*} \xi(\theta(z)) h(z) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz}{\int_{S_t^*} \xi(\theta(z)) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz}$$
$$= \frac{\int_{S_t^*} \xi_n(\theta(z)) h(z) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz + \int_{S_t^*} [\xi(\theta(z)) - \xi_n(\theta(z))] h(z) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz}{\int_{S_t^*} \xi_n(\theta(z)) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz + \int_{S_t^*} [\xi(\theta(z)) - \xi_n(\theta(z))] e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz}$$
$$= \frac{(\text{Nume})_1 + (\text{Nume})_2}{(\text{Denom})_1 + (\text{Denom})_2}. \tag{22}$$

By (i), (L1), (L2), and a similar proof of Lemma 4.2(a) we can find a $C_* > 0$, independent of n, such that for each $n \geqslant 1$

$$(\text{Denom})_1 = \int_{S_t^*} \xi_n(\theta(z)) e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz \geqslant C_* > 0 \quad \text{with } P_{\theta_0}\text{-probability tending to 1.}$$

Then, by (ii) and (iv) we have that $(\text{Nume})_2$ and $(\text{Denom})_2$ are bounded by

$$\sup_{\theta \in \Theta} |\xi_n(\theta) - \xi(\theta)| \int_{S_t^*} e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \, dz \leqslant \sup_{\theta \in \Theta} |\xi_n(\theta) - \xi(\theta)| M_* \equiv a_n \to 0.$$

Finally, dividing both the numerator and the denominator of (22) by $(\text{Denom})_1$, we have that for each fixed n

$$\frac{\Phi_k h - a_n C_*^{-1}}{1 + a_n C_*^{-1}} \leqslant \liminf_{t\to\infty} E_\xi^t h(Z_t) \leqslant \limsup_{t\to\infty} E_\xi^t h(Z_t) \leqslant \frac{\Phi_k h + a_n C_*^{-1}}{1 - a_n C_*^{-1}};$$

hence, the desired result follows since $a_n \to 0$ as $n \to \infty$. □

Conditions (i) and (ii) in the above theorem imply that ξ must be continuous and bounded; however, ξ need not have a compact support or be continuously differentiable on $\mathfrak{R}^k$. So, the theorem applies to priors such as normal, Gamma(p, λ) with a shape parameter $p \geqslant 1$, and Beta(r, s) with $(r, s) \in \{(1, 2) \times (1, \infty)\} \cup \{(1, \infty) \times (1, 2)\}$. For unbounded or non-differentiable (on $\mathfrak{R}^k$) priors such as Beta(r, s) with $r \leqslant 1$ or $s \leqslant 1$, and Gamma(p, λ) with a shape parameter $p < 1$, we provide the following theorem.

Theorem 4.3. *Let ξ be a prior on Θ. Suppose the following conditions hold*:

(i) *there exists a sequence of priors $\{\xi_n, n \geqslant 1\}$ with supports Θ_{ξ_n} such that* (P1)–(P2) *are satisfied*, (P3) *is satisfied with the same ε_0 and δ_0*;
(ii) *there exists a sequence of sets $\{A_n, n \geqslant 1\}$ such that $\theta_0 \in A_n$, $A_n \nearrow \Theta$, and the Lebesgue measure of $A_n^c \equiv \Theta \backslash A_n$ approaches* 0;
(iii) $\sup_{\theta \in A_n} |\xi_n(\theta) - \xi(\theta)| \to 0$;
(iv) $|\xi_n| \leqslant |\xi|$ *on A_n^c*;
(v) *the likelihood satisfies* (L1)–(L3), *but with Θ_ξ in* (L3) *replaced by Θ_{ξ_n}*;
(vi) *there exists $0 < M_* < \infty$, not depending on t, such that* (21) *holds*;
(vii) *there exists some $n_0 > 0$ such that $(\det B_t)\xi(\theta)\mathrm{e}^{-g(\theta,t)} 1_{A_{n_0}^c}$ is uniformly integrable in t, where $g(\theta, t)$ is as in* L3(ii).

Then, for any bounded measurable h we have $E_\xi^t[h(Z_t)] \xrightarrow{\mathrm{p}} \Phi h$.

Proof. We give only a sketch of the proof as it resembles that of Theorem 4.2. To begin, by (i) and (v) we can apply Theorem 4.1 for ξ_n, $n \geqslant 1$; and we can write $E_\xi^t h(Z_t)$ as in (22). A similar argument shows that by (i), (L1), and (L2) there exists a $C_* > 0$, independent of n, such that for each $n \geqslant 1$

$$(\text{Denom})_1 \geqslant C_* > 0 \quad \text{with } P_{\theta_0}\text{-probability tending to 1.}$$

Next, letting $A_{n,t} = \{z : z = B_t(\theta - \hat{\theta}_t), \theta \in A_n\}$ and S_t^* be as in (20), both $(\text{Nume})_2$ and $(\text{Denom})_2$ are bounded by

$$\int_{S_t^*} |\xi(\theta(z)) - \xi_n(\theta(z))| \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z$$
$$= \int_{A_{n,t}} |\xi(\theta(z)) - \xi_n(\theta(z))| \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z + \int_{S_t^* \backslash A_{n,t}} |\xi(\theta(z)) - \xi_n(\theta(z))| \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z,$$

where by (vi) the first integral is bounded by $\sup_{\theta \in A_n} |\xi(\theta) - \xi_n(\theta)| M_*$, which approaches 0 as $n \to \infty$ by assumption (iii); and by assumptions (ii), (iv), and (vii) the second integral is bounded by

$$\int_{S_t^* \backslash A_{n,t}} 2\xi(\theta(z)) \mathrm{e}^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)} \,\mathrm{d}z \leqslant \sup_t \int_{A_n^c} 2\xi(\theta)(\det B_t) \mathrm{e}^{-g(t,\theta)} \,\mathrm{d}\theta \equiv m_n^*,$$

which approaches 0 as $n \to \infty$. Then, the result follows. □

5. Comparison with earlier work

Since this work is closely related to Weng (2003) and Sweeting (1992), this section compares the present conditions with these two papers. To compare with Weng (2003), we first note that she only considered the one-dimensional case. In such case, one has the property that $\sup_z |zUh(z)| < \infty$ for bounded h; therefore, by (12),

$$|[Uh(Z_t)]'(B_t')^{-1} \nabla u_t(\theta)| \leqslant C \left| 1 - \frac{\ell_t''(\theta^*)}{\ell_t''(\hat{\theta}_t)} \right|, \tag{23}$$

for some $C > 0$, and the analysis of II_t in (19) is simpler. For multi-parameter cases, we have $\sup_z |z^{\mathrm{T}} Uh(z)| < \infty$ for any bounded h; however, it does not lead to an inequality like (23) because here $[Uh(Z_t)]'(B_t')^{-1} \nabla u_t(\theta)$ involves cross terms $z_i (Uh)_j(z)$, $i \neq j$, where $(Uh)_j$ denotes the jth component of Uh. Of course, if the Hessian matrix of the loglikelihood function is diagonal, we have a multivariate version of (23).

Next, the prior conditions in Weng (2003) are quite restrictive. It is assumed that the prior ξ is continuous and strictly positive on some closed interval $[a, b]$, and is continuously differentiable on (a, b). Here we allow priors to be more general.

To compare with Sweeting's (1992) conditions, note that both have similar conditions on information growth and information continuity, but our condition on the nonlocal behavior of the likelihood is different from his. More explicitly (using our notations) his (C6) states that

C6 (Nonlocal behavior).

$$(\det B_t)\int_{\theta\in N_t^c(\hat\theta_t,c_t)} e^{\ell_t(\theta)-\ell_t(\hat\theta_t)}\xi(\theta)\,d\theta \to 0,$$

where $N_t(\psi,c)\equiv\{\theta : |[(-\nabla^2\ell_t(\psi))^{1/2}]'(\theta-\psi)|<c\}$ and $c_t\to\infty$ such that

$$\sup_{\theta\in N_t(\hat\theta_t,c_t)} \|(-\nabla^2\ell_t(\hat\theta_t))^{-1/2}(\nabla^2\ell_t(\theta)-\nabla^2\ell_t(\hat\theta_t))[((-\nabla^2\ell_t(\hat\theta_t))^{-1/2}]'\|\to 0.$$

As he pointed out, condition C6 can be difficult to check in its present form; therefore, he proposed condition C6*(i)–(iii) that reduces the region over which it is necessary to check nonlocal behavior of the posterior distribution. Furthermore, when information on different parameters does not vary too wildly, he proposed Lemma 2.3, which is helpful to check C6*(iii). The stochastic versions of C6* and Lemma 2.3 are his D6 and Lemma 3.3. The condition D6 states that

D6 (Nonlocal behavior). For each $t\in\mathscr{T}$ there exists a nonrandom open convex set C_t containing θ_0 which satisfies
- (i) $P_{\theta_0}(-\nabla^2\ell_t(\theta)$ is positive definite on $C_t)\to 1$.
- (ii) $\xi(\theta)$ eventually bounded on C_t.
- (iii) $(\det B_t)\int_{\theta\notin C_t} e^{\ell_t(\theta)-\ell_t(\hat\theta_t)}\xi(\theta)\,d\theta \xrightarrow{p} 0$.

However, by the proof of his Lemma 2.2, for his C6 and C6* to be equivalent, it is necessary that $\xi(\theta)$ be uniformly bounded in C_t; that is, D6(ii) is indeed the following:

$$\sup_{\theta\in C_t}\xi(\theta)<M<\infty \quad \text{where } M \text{ does not involve } t.$$

So, D6 can be difficult to check when ξ is unbounded. In contrast, unlike his D6(iii), our L3(ii), (16), and (21) do not have ξ in the integrand. In Section 6.1 we provide an example where our conditions hold, but his Lemma 3.3 and D6 fail. On the other hand, our approach has some disadvantages too. Take the nonhomogeneous Poisson model in Section 6.3 as an example. With the choice of $g(t,\theta)$ in (34), when ξ has an unbounded support though it can be shown that

$$(\det B_t)\int_{\Theta} e^{\ell_t(\theta)-\ell_t(\hat\theta_t)}\xi(\theta)\,d\theta<M<\infty,$$

for some M not involving t, it is difficult to see whether (21) holds. Therefore, we have to assume that the prior has a compact support, which, however, is not required in Sweeting (1992).

6. Examples

Three examples are given to illustrate our method in this section. First, we give an example for which the eigenvalue condition for Lemma 3.3 in Sweeting (1992) does not hold but the conditions for our Theorem 4.3 do hold. Secondly, we consider an i.i.d. sequence from exponential families, where Theorems 4.1–4.3 all apply. Then, we revisit a nonhomogeneous Poisson model.

6.1. A bivariate normal model

Consider a sequence of independent bivariate normal variables

$$X_n=(X_{n1},X_{n2})'\sim N(\theta_1 n^{-1/2},\theta_2 n^{-1/4}\exp[\tfrac12 n^{1/2}],1,1,0),\quad n=1,\ldots,t.$$

It is easy to show that

$$\hat\theta_t=(\hat\theta_{t1},\hat\theta_{t2})'=\left(\frac{\sum_{n=1}^t n^{-1/2}X_{n1}}{\sum_{n=1}^t n^{-1}},\frac{\sum_{n=1}^t n^{-1/4}\exp[\frac12 n^{1/2}]X_{n2}}{\sum_{n=1}^t n^{-1/2}\exp(n^{1/2})}\right)'\xrightarrow{p}(\theta_1,\theta_2)'$$

and that

$$-\nabla^2\ell_t(\theta)=\begin{pmatrix}\sum_{n=1}^t n^{-1} & 0\\ 0 & \sum_{n=1}^t n^{-1/2}\exp(n^{1/2})\end{pmatrix},$$

not depending on the parameters. So,

$$B_t=\begin{pmatrix}(\sum_{n=1}^t n^{-1})^{1/2} & 0\\ 0 & (\sum_{n=1}^t n^{-1/2}\exp(n^{1/2}))^{1/2}\end{pmatrix}=\begin{pmatrix}b_{t1} & 0\\ 0 & b_{t2}\end{pmatrix}$$

and

$$(B_t')^{-1}\left(\frac{\partial^2 \ell_t}{\partial\theta_i\,\partial\theta_j}(\eta^{ij})\right)B_t^{-1} = -I_k.$$

Therefore, (L1) holds easily; (L2) holds with arbitrary $b_t \to \infty$; by (12), (L3)(i) holds for any $r \geqslant 1$ and $c \geqslant 0$. Moreover, (L3)(ii) is satisfied since $\ell_t(\hat{\theta}_t) - \ell_t(\theta) = \|z_t\|^2/2$ and $\int_{\mathfrak{R}^k} \|z\|^r \mathrm{e}^{-\|z\|^2/2}\,\mathrm{d}z < \infty$ for any $r \geqslant 1$. Hence, the asymptotic posterior normality holds if ξ satisfies (P1)–(P3) or conditions (i)–(ii) in Theorem 4.2.

To illustrate how Theorem 4.3 works, below we consider the case where the prior distributions of θ_i, $i = 1, 2$ are independent Gamma(p_i, λ_i) with $p_i < 1$. Note that the prior densities approach ∞ at 0. It is easily seen that the likelihood satisfies (L1)–(L3) with Θ_ξ in (L3) replaced by Θ, and (i)–(vi) in Theorem 4.3 are satisfied. So, it suffices to check condition (vii). Take $A_n = \{(\theta_1, \theta_2) : \theta_1 \geqslant 1/n \text{ or } \theta_2 \geqslant 1/n\}$. Denote the true parameter as $\theta_0 = (\theta_{01}, \theta_{02})'$. So, $\theta_{0i} > 0$, $i = 1, 2$. Let n_0 be large enough such that $1/n_0 < \theta_{01}/2$ and $1/n_0 < \theta_{02}/2$. Write

$$\int \xi(\theta_1, \theta_2)\mathrm{e}^{-\|z\|^2/2}(\det B_t)1_{A^c_{n_0}}\,\mathrm{d}\theta = \int_0^{1/n_0}\int_0^{1/n_0} \xi(\theta_1, \theta_2)\mathrm{e}^{-(\sum_{i=1}^2 b_{ti}^2(\theta_1 - \hat{\theta}_{t1})^2)/2} b_{t1}b_{t2}\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2. \tag{24}$$

Note that if for $i = 1, 2$ $\theta_i \in (0, 1/n_0)$, then $\{|\theta_{0i} - \hat{\theta}_{ti}| < \theta_{0i}/4\} \subseteq \{|\theta_i - \hat{\theta}_{ti}| > \theta_{0i}/4\}$; moreover, since $\hat{\theta}_t \xrightarrow{\mathrm{p}} \theta_0$, we have that with P_{θ_0}-probability tending to 1, $|\theta_i - \hat{\theta}_{ti}| > \theta_{0i}/4$; and therefore, with P_{θ_0}-probability tending to 1,

$$\exp\left[-\frac{\sum_{i=1}^2 b_{ti}^2(\theta_i - \hat{\theta}_{ti})^2}{2}\right] b_{t1}b_{t2} \leqslant C \quad \forall t,$$

for some $0 < C < \infty$. Then, together with (ii) in Theorem 4.3, condition (vii) follows since the integrand in (24) is bounded by $C\xi(\theta)$, which is integrable on $A^c_{n_0}$.

To compare with the conditions D1–D6 in Sweeting (1992), below we shall check his D6 and Lemma 3.3 (recall the discussion at the end of Section 5). To start, we note that his Lemma 3.3 requires that

$$\frac{\log \lambda_{\max}(B_t)}{\lambda_{\min}(B_t)} \quad \text{is stochastically bounded,}$$

which does not hold here because

$$\lambda_{\max}(B_t) = \left\{\sum_{n=1}^t n^{-1/2}\exp(n^{1/2})\right\}^{1/2} \sim [2(\mathrm{e}^{t^{1/2}} - e)]^{1/2},$$

$$\lambda_{\min}(B_t) = \left\{\sum_{n=1}^t n^{-1}\right\}^{1/2} \sim (\log t)^{1/2}.$$

Next, in this case his D6(iii) becomes

$$\int_{\theta \notin C_t} \xi(\theta_1, \theta_2)\mathrm{e}^{-(\sum_{i=1}^2 b_{ti}^2(\theta_i - \hat{\theta}_{ti})^2)/2} b_{t1}b_{t2}\,\mathrm{d}\theta \xrightarrow{\mathrm{p}} 0,$$

which is much stronger than the integrability condition (vii) in our Theorem 4.3 and does not hold here because by D6(ii) the set $\{\theta \notin C_t\}$ must contain a fixed region around the origin.

6.2. Exponential families

Consider an i.i.d. sample $X_1, \ldots, X_t$ from a k-parameter standard exponential family defined by probability densities of the form

$$p_\theta(x) = \mathrm{e}^{\theta' x - \psi(\theta)},$$

where ψ is strictly convex and $\theta \in \Theta$, the natural parameter space, assumed to be open. For references of exponential families, see Brown (1986) and Lehmann (1983, 1986). So, $\ell_t(\theta) = \theta' Y_t - t\psi(\theta)$ and $\nabla^2\ell_t(\theta) = -t\nabla^2\psi(\theta)$ where $Y_t = \sum_{i=1}^t X_i$. It is easily seen that (L1) holds. Next, since $B_t = \sqrt{t}(\nabla^2\psi(\hat{\theta}_t))^{1/2}$, (L2) holds with $b_t = t^{1/4}$. Then, by (12) and the assumption that ξ has a compact support, (L3)(i) holds with $r = 1$. Finally, we claim that L3(ii$'$) holds with

$$g^*(z_t) = \begin{cases} c\|z_t\| & \text{if } \|z_t\| > b, \\ 0 & \text{otherwise,} \end{cases}$$

where c is some positive constant and $b=2[\lambda_{\max}(\nabla^2\psi(\theta_0))]^{1/2}$. To see this, first observe that

$$\int_{\mathfrak{R}^k}\|z\|\mathrm{e}^{-g^*(z)}\,\mathrm{d}z<M<\infty.$$

Now it suffices to show that $\ell_t(\hat\theta_t)-\ell_t(\theta)\geqslant g^*(z_t)$. Apparently $\ell_t(\hat\theta_t)-\ell_t(\theta)\geqslant 0$ for all $\theta\in\Theta$. Next, note that

$$\|z_t\|\geqslant b\quad\text{implies } \|\theta-\hat\theta_t\|>2t^{-1/2}[\lambda_{\max}(\nabla^2\psi(\theta_0))/\lambda_{\max}(\nabla^2\psi(\hat\theta_t))]^{1/2},$$

which exceeds $t^{-1/2}$ when t is sufficiently large; and moreover, for $\|\theta-\hat\theta_t\|>t^{-1/2}$, by concavity of ℓ_t and Lemma 6.3 below we have

$$\begin{aligned}\ell_t(\hat\theta_t)-\ell_t(\theta)&\geqslant\frac{1}{2}\inf_{\|u\|=1}\left\{\left|\left(\nabla\ell_t\left(\hat\theta_t+\frac{t^{-1/2}u}{2}\right)\right)'u\right|\right\}\|\theta-\hat\theta_t\|\\&=\frac{1}{2}\inf_{\|u\|=1}\left\{\left|\left(\nabla\psi\left(\hat\theta_t+\frac{t^{-1/2}u}{2}\right)-\nabla\psi(\hat\theta_t)\right)'u\right|\right\}t\|\theta-\hat\theta_t\|.\end{aligned}$$

Then, by mean value theorem the above line is bounded below by $c\|z_t\|$, for some $c>0$. So, the desired result follows.

Lemma 6.3. *Suppose that $f(x)$ is strictly convex (or concave), x_m is its unique minimum (or maximum), and c is some fixed positive constant. Then, for all x such that $\|x-x_m\|\geqslant c$, we have*

$$|f(x)-f(x_m)|\geqslant\left(\frac{1}{2}\right)\inf_{\|u\|=1}\left\{\left|\left(\nabla f\left(x_m+\frac{cu}{2}\right)\right)'u\right|\right\}\|x-x_m\|.$$

Proof. For the univariate case, we have

$$\begin{aligned}|f(x)-f(x_m)|&\geqslant\left|f(x)-f\left(x_m+\frac{x-x_m}{2}\right)\right|\\&\geqslant\left|\left(\frac{\mathrm{d}}{\mathrm{d}x}f\left(x_m+\frac{x-x_m}{2}\right)\right)\left(x-x_m-\frac{c}{2}\right)\right|\\&\geqslant\left|\left(\frac{\mathrm{d}}{\mathrm{d}x}f\left(x_m+\frac{c}{2}\,\mathrm{sgn}(x-x_m)\right)\right)\frac{1}{2}(x-x_m)\right|.\end{aligned}$$

The result for multivariate cases follows analogously. □

Note that the arguments above remain valid for any b_t satisfying $b_t=o(t^{1/2})$ and $b_t\to\infty$. Note also that in the preceding paragraph the compactness of Θ_ξ is only needed for checking (L3)(i), and (21) holds in this case; hence, we can extend the results to general priors by Theorems 4.2 and 4.3.

6.3. Nonhomogeneous Poisson process

Here we review the two-parameter nonhomogeneous Poisson process in Sweeting (1992). To begin, assume that the intensity function is $\Lambda(t)=\lambda\mathrm{e}^{\mu+\lambda t}$ over the time interval, where $\lambda>0$ and μ are two unknown parameters, and that, over the time period $(0,t)$, N_t events are observed by times $x_1,\ldots,x_{N_t}$. So, for each fixed t, N_t is a Poisson with mean $\int_0^t\Lambda(s)\,\mathrm{d}s$. Letting $\theta=(\lambda,\mu)'$, then the loglikelihood function is $\ell_t(\theta)=N_t(\log\lambda+\mu)+\lambda\sum_{i=1}^{N_t}x_i-\mathrm{e}^{\mu}(\mathrm{e}^{\lambda t}-1)$, with derivatives

$$\nabla\ell_t(\theta)=\left(N_t/\lambda+\sum_{i=1}^{N_t}x_i-t\mathrm{e}^{\mu+\lambda t},N_t-\mathrm{e}^{\mu}(\mathrm{e}^{\lambda t}-1)\right)',\tag{25}$$

$$-\nabla^2\ell_t(\theta)=\begin{pmatrix}N_t/\lambda^2+t^2\mathrm{e}^{\mu+\lambda t} & t\mathrm{e}^{\mu+\lambda t}\\ t\mathrm{e}^{\mu+\lambda t} & \mathrm{e}^{\mu}(\mathrm{e}^{\lambda t}-1)\end{pmatrix}.\tag{26}$$

Suppose that θ satisfies P and P^*, and Θ_ξ is compact. In the following we shall verify our conditions (L1)–(L3). First, choose B_t in (1) as

$$B_t=\begin{pmatrix}b_{11} & b_{12}\\ 0 & b_{22}\end{pmatrix}=\begin{pmatrix}(N_t/\hat\lambda_t^2+t^2\mathrm{e}^{\hat\mu_t+\hat\lambda_t t})^{1/2} & t\mathrm{e}^{\hat\mu_t+\hat\lambda_t t}/b_{11}\\ 0 & [\det(-\nabla^2\ell_t(\hat\theta_t))]^{1/2}/b_{11}\end{pmatrix}.\tag{27}$$

Setting $\partial\ell_t/\partial\mu=0$, we obtain

$$\mathrm{e}^{\mu}=N_t/(\mathrm{e}^{\lambda t}-1).\tag{28}$$

Plugging this into $\partial\ell_t/\partial\lambda$ leads to

$$\frac{\partial\ell_t}{\partial\lambda}=N_t/\lambda+\sum_{i=1}^{N_t}x_i-te^{\lambda t}N_t/(e^{\lambda t}-1). \tag{29}$$

Given data by time t, we have that $\lim_{\lambda\to\infty}\partial\ell_t/\partial\lambda=\sum_{i=1}^{N_t}x_i-tN_t$, which is negative because $x_{N_t}\leqslant t$ and $x_i<t\forall i<N_t$; and that $\lim_{\lambda\to 0+}\partial\ell_t/\partial\lambda=\sum_{i=1}^{N_t}x_i-tN_t/2$, which tends to ∞ w.p.1 (P_{θ_0}). Moreover, since (29) is strictly decreasing in λ, it has a unique root $\hat\lambda_t>0$ w.p.1 under P_{θ_0} for t sufficiently large. Then from (28), $\hat\mu_t=\log\{N_t/(e^{\hat\lambda_t t}-1)\}$. From (26) one can verify that, for t sufficiently large, $-\nabla^2\ell_t$ is positive definite at $\hat\theta_t=(\hat\lambda_t,\hat\mu_t)'$; and therefore, $\hat\theta_t$ is the unique MLE. The fact that $\|B_t^{-1}\|\xrightarrow{p}0$ can be easily obtained by (26). For consistency of the MLE, we refer to Keiding (1974). Hence, (L1) holds.

Next, for (L2) it suffices to show that the supremum (over $\{\theta:\|z_t\|\leqslant b_t\}$) for each component of the matrix $I_k+(B_t')^{-1}((\partial^2\ell_t/\partial\theta_i\partial\theta_j)(\eta^{ij}))B_t^{-1}$ converges to zero in P_{θ_0}-probability. Define U as a 2×2 matrix with (i,j)-component

$$U_{ij}\equiv\frac{\partial^2\ell_t(\eta^{ij})}{\partial\theta_i\partial\theta_j}-\frac{\partial^2\ell_t(\hat\theta_t)}{\partial\theta_i\partial\theta_j}.$$

So,f

$$I_k+(B_t')^{-1}\left(\frac{\partial^2\ell_t}{\partial\theta_i\theta_j}(\eta^{ij})\right)B_t^{-1}=(B_t')^{-1}UB_t^{-1}. \tag{30}$$

By Taylor expansions,

$$U_{ij}=\ell_{t,ij1}(\omega^{ij})(\eta_1^{ij}-\hat\lambda_t)+\ell_{t,ij2}(\omega^{ij})(\eta_2^{ij}-\hat\mu_t), \tag{31}$$

where $\ell_{t,ijk}$ denote the third derivatives of ℓ_t and $\omega^{ij}=(\omega_1^{ij},\omega_2^{ij})'$ lies between $\eta^{ij}=(\eta_1^{ij},\eta_2^{ij})'$ and $\hat\theta_t$. Simple calculations show that $(\ell_{t,111},\ell_{t,112})=(t\ell_{t,11}+2N_t/\lambda^3+tN_t/\lambda^2,\ell_{t,11}+N_t/\lambda^2)$, and for $(i,j)\neq(1,1)$ the relation between $(\ell_{t,ij1},\ell_{t,ij2})$ and $(\ell_{t,ij},\ell_{t,ij})$ can be easily obtained. Then, some algebra yields the following expression for the (1,1)-component of the matrix $(B_t')^{-1}UB_t^{-1}$:

$$\begin{aligned}[(B_t')^{-1}UB_t^{-1}]_{11}&=U_{11}/\ell_{t,11}(\hat\theta_t)\\&=[\ell_{t,11}(\hat\theta_t)]^{-1}\{(\eta_1^{ij}-\hat\lambda_t)[t\ell_{t,11}(\omega^{11})+2N_t/(\omega_1^{11})^3+tN_t/(\omega_1^{11})^2]\\&\quad+(\eta_2^{ij}-\hat\mu_t)[\ell_{t,11}(\omega^{11})+N_t/(\omega_2^{11})^2]\},\end{aligned} \tag{32}$$

where $|\eta_1^{ij}-\hat\lambda_t|\leqslant|\lambda-\hat\lambda_t|$ and $|\eta_2^{ij}-\hat\mu_t|\leqslant|\mu-\hat\mu_t|$. Since $N_t/e^{\mu_0+\lambda_0 t}\to 1$ a.e. P_{θ_0},

$$\{\|z_t\|\leqslant a_t\}\text{ implies }|\lambda-\hat\lambda_t|\leqslant a_t\hat\lambda_t e^{-(\hat\mu_t+\hat\lambda_t t)/2}\text{ and }|\mu-\hat\mu_t|\leqslant a_t(1+\hat\lambda_t t)e^{-(\hat\mu_t+\hat\lambda_t t)/2}. \tag{33}$$

So, together with (26), the right-hand side of (32) approaches 0 in P_{θ_0}-probability over the supremum of the set $\{\theta:\|z_t\|\leqslant t\}$. Similar arguments apply to the other components of the matrix. So, (L2) holds with $b_t=t$.

Now consider (L3)(i). From (12) and (30), it suffices to show that each component of the matrix $(B_t')^{-1}UB_t^{-1}$ is bounded by $\|z_t\|^s$ for some $s\geqslant 1$ over the set $\{\|z_t\|\geqslant b_t\}$. Here we consider the (1,2)-component of the matrix. Write

$$\begin{aligned}[(B_t')^{-1}UB_t^{-1}]_{12}&=(U_{12}b_{11}-U_{11}b_{21})/(b_{11}^2b_{22})\\&=[U_{12}\ell_{t,11}(\hat\theta_t)-U_{11}\ell_{t,21}(\hat\theta_t)]/\left[\ell_{t,11}(\hat\theta_t)\sqrt{\ell_{t,11}(\hat\theta_t)\ell_{t,22}(\hat\theta_t)-\ell_{t,12}^2(\hat\theta_t)}\right],\end{aligned}$$

where U_{ij} are as in (31). Direct calculations yield the desired result, noting that for t sufficiently large, if $t=b_t<\|z_t\|\leqslant t^{-2}e^{\lambda_0 t/2}$, then from (33) there exist $c_1>0$ and $c_2>0$ such that $t|\lambda-\hat\lambda_t|\leqslant c_1t^{-1}$ and $|\mu-\hat\mu_t|\leqslant c_2t^{-1}$; and hence $te^{(\eta_1^{ij}-\hat\lambda_t)t}\leqslant\|z_t\|$. Moreover, if $\|z_t\|>t^{-2}e^{\lambda_0 t/2}$, then by the assumption that ξ has a bounded support, $te^{(\eta_1^{ij}-\hat\lambda_t)t}\leqslant\|z_t\|^s$ for some $s\geqslant 1$. The same arguments can be applied to the other components of the matrix. Hence, (L3)(i) holds with $r=s$.

For (L3)(ii), first we let $0<\lambda_1<\lambda_0$; and define the regions $R_{1t}=\{\theta:\lambda\leqslant\lambda_1\}$, $R_{2t}=\{\theta:\lambda>\lambda_1$ and $(\mu-\mu_0)+(\lambda-\lambda_0)t<ct\}$, and $R_{3t}=\{\theta:\lambda>\lambda_1$ and $(\mu-\mu_0)+(\lambda-\lambda_0)t\geqslant ct\}$, where c can be chosen so that ℓ_t is concave over R_{2t}. Actually, the idea of choosing such regions is from Sweeting (1992). Since $\ell_t(\hat\theta_t)-\ell_t(\theta)\geqslant\ell_t(\theta_0)-\ell_t(\theta)$, it suffices to find suitable lower bounds on either $\ell_t(\hat\theta_t)-\ell_t(\theta)$ or $\ell_t(\theta_0)-\ell_t(\theta)$. Then we claim that (L3)(ii) holds with

$$g(t,\theta)=\begin{cases}C\|B_t(\theta-\hat\theta_t)\| & \text{if }\theta\in R_{2t}\cup R_{3t},\\ e^{\lambda_0 t/2} & \text{if }\theta\in R_{1t},\end{cases} \tag{34}$$

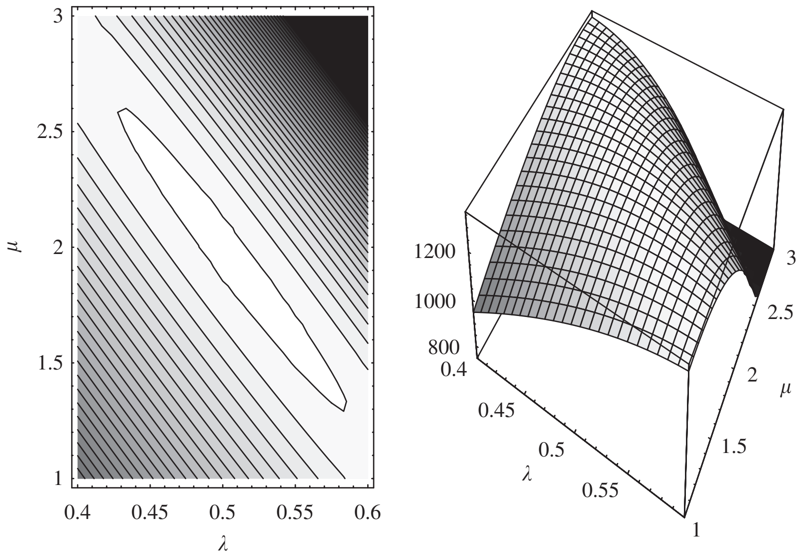

Fig. 1. Contour and 3D plots of a loglikelihood function.

where C is some positive constant. Since $\det B_t$ is of order $\exp(\lambda_0 t)$ and Θ_ξ is compact, it is easily seen that $(\det B_t)\|z_t\| \exp\{-g(t,\theta)\}$ is uniformly integrable in t and $\int_{\Theta_\xi}(\det B_t)\|z_t\|^s \exp\{-g(t,\theta)\}\,d\theta < M < \infty$. So, it now suffices to show that $\ell_t(\hat{\theta}_t) - \ell_t(\theta) \geqslant g(t,\theta)$. For R_{1t}, with N_t replaced by its asymptotic expectation $e^{\mu_0+\lambda_0 t}$, it can be shown that $\ell_t(\theta) - \ell_t(\theta_0) \leqslant -c_0 e^{\lambda_0 t}$ over R_{1t}. Next, consider R_{2t}. Using the arguments in Lemma 2.2 of Sweeting (1992) but with c_t there replaced by a constant c_0', we obtain

$$\ell_t(\hat{\theta}_t) - \ell_t(\theta) \geqslant \frac{c_0''}{4}\|B_t(\theta - \hat{\theta}_t)\| = \frac{c_0''}{4}\|Z_t\| \tag{35}$$

for $\theta \in R_{2t} - \{z_t : \|z_t\| \leqslant c_0''\}$. Finally, we consider the region R_{3t}. Given $\theta^\diamond = (\lambda^\diamond, \mu^\diamond)' \in R_{3t}$, denote $\theta^b = (\lambda^b, \mu^b)'$ as the intersection of $\lambda = \lambda^\diamond$ and $(\mu - \mu_0) + (\lambda - \lambda_0)t = ct$; therefore, $\mu^\diamond > \mu^b$. Now, in view of

$$\ell_t(\hat{\theta}_t) - \ell_t(\theta^\diamond) = [\ell_t(\hat{\theta}_t) - \ell_t(\theta^b)] + [\ell_t(\theta^b) - \ell_t(\theta^\diamond)],$$

(35), and the triangular inequality, it suffices to show that

$$\ell_t(\theta^b) - \ell_t(\theta^\diamond) \geqslant c_0'''\|B_t(\theta^b - \theta^\diamond)\|.$$

To prove the above inequality, we first observe from (25) that

$$-\frac{\partial \ell_t}{\partial \mu}(\theta) \geqslant e^{\mu_0+\lambda_0 t + ct}(1 - e^{-\lambda_1 t}) - N_t,$$

for $\theta \in R_{3t}$, where the right-hand side exceeds $[\mathrm{tr}(B_t' B_t)]^{1/2}$ with P_{θ_0}-probability tending to 1. Therefore,

$$\ell_t(\theta^b) - \ell_t(\theta^\diamond) = \frac{\partial \ell_t}{\partial \mu}(\theta^\sharp)(\theta^b - \theta^\diamond) \geqslant [\mathrm{tr}(B_t' B_t)]^{1/2}\|\theta^b - \theta^\diamond\| \geqslant \|B_t\|\|\theta^b - \theta^\diamond\|.$$

Thus, from Theorem 4.1, if the prior satisfies conditions (P1)–(P3) then asymptotic posterior normality holds. We note, however, that with the choice of $g(t,\theta)$ in (34), one cannot verify (21), so conditions (iv) and (vi) of Theorems 4.2 and 4.3, respectively, will fail to hold.

To gain more idea on the behavior of the loglikelihood function, we take $(\lambda_0, \mu_0) = (0.5, 2.0)$ and generate a nonhomogeneous Poisson process over the time period $(0, 8)$. The resulting loglikelihood function based on simulated data is

$$397(\mu + \log\lambda) + 2446.3\lambda - e^{\mu}(e^{8\lambda} - 1),$$

and the maximum likelihood estimate is $(\hat{\lambda}_t, \hat{\mu}_t) = (0.504, 1.965)$. Fig. 1 shows the contour and three-dimensional plots of the loglikelihood function. These plots give some insight for the selection of regions R_{it} in verifying L3(ii).

Acknowledgments

The first author is partially supported by the National Science Council of Taiwan.

References

Brown, L.D., 1986. Fundamentals of Statistical Exponential Families. IMS, Hayward, CA.
Cam, L.L., 1953. On some asymptotic properties of maximum likelihood and related Bayes estimates. Univ. Calif. Publ. Statist. 1, 277–330.
Chen, C.F., 1985. On asymptotic normality of limiting density functions with Bayesian implications. J. Roy. Statist. Soc. Ser. B 47, 540–546.
Heyde, C.C., Johnstone, I.M., 1979. On asymptotic posterior normality for stochastic processes. J. Roy. Statist. Soc. Ser. B 41, 184–189.
Johnson, R., 1970. Asymptotic expansions associated with posterior distributions. Ann. Math. Statist. 41, 851–864.
Keiding, N., 1974. Estimation in the birth process. Biometrika 61, 71–80.
Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.
Lehmann, E.L., 1986. Testing Statistical Hypotheses. Wiley, New York.
Sweeting, T.J., 1992. On asymptotic posterior normality in the multiparameter case. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics. Oxford University Press, Oxford, pp. 825–835.
Sweeting, T.J., Adekola, A.O., 1987. Asymptotic posterior normality for stochastic processes revisited. J. Roy. Statist. Soc. Ser. B 49, 215–222.
Walker, A.M., 1969. On the asymptotic behaviour of posterior distributions. J. Roy. Statist. Soc. Ser. B 31, 80–88.
Weng, R.C., 2003. On Stein's identity for posterior normality. Statist. Sinica 13, 495–506.
Weng, R.C., Woodroofe, M., 2000. Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels. Statist. Sinica 10, 693–713.
Woodroofe, M., 1989. Very weak expansions for sequentially designed experiments: linear models. Ann. Statist. 17, 1087–1102.
Woodroofe, M., 1992. Integrable expansions for posterior distributions for one-parameter exponential families. Statist. Sinica 2, 91–111.
Woodroofe, M., Coad, D.S., 1997. Corrected confidence sets for sequentially designed experiments. Statist. Sinica 7, 53–74.