

## ON STEIN'S IDENTITY FOR POSTERIOR NORMALITY

Ruby C. Weng

*National Chengchi University*

*Abstract:* We propose a new method to derive posterior normality of stochastic processes. For a suitable parameter transformation  $Z_t$ , the likelihood function is converted to a form close to a standard normal density. Then we apply a version of Stein's Identity to obtain an expression for the posterior expectation. From this, posterior normality of  $Z_t$  can be established. Applications of this method are illustrated by the conditional exponential family and a nonhomogeneous Poisson process.

*Key words and phrases:* Maximum likelihood estimator, posterior distributions, posterior normality, Stein's identity, stochastic processes.

### 1. Introduction

Asymptotic posterior normality has been studied since the time of Laplace and has attracted the attention of many authors. See, for example, Le Cam (1953), Dawid (1970) and Johnson (1970) for independent and identically distributed (i.i.d.) observations; Heyde and Johnstone (1979), Basawa and Rao (1980), Chen (1985), and Sweeting and Adekola (1987) for stochastic processes. Walker (1969) presented a straightforward approach to posterior normality. Heyde and Johnstone (1979) simplified Walker's (1969) conditions and showed that asymptotic posterior normality holds under weaker conditions than those required for asymptotic normality of maximum likelihood estimator; Chen (1985) provided conditions with more operational flexibility for the asymptotic normality of limiting density functions, which can be applied to the problem of asymptotic posterior normality. Both Heyde and Johnstone (1979) and Chen (1985) used a fixed neighborhood on the condition for asymptotic continuity of information function. To be more precise, denoting  $\ell_t''(\theta)$  as the second derivative of the log-likelihood function with respect to  $\theta$ , they required that given  $\epsilon > 0$ , there exists  $\delta(\epsilon) > 0$  such that

$$P_{\theta_0} \left( \sup_{\{|\theta - \theta_0| < \delta(\epsilon)\}} \left| \frac{\ell_t''(\theta)}{\ell_t''(\theta_0)} - 1 \right| < \epsilon \right) \geq 1 - \epsilon.$$

However, this condition excludes certain processes of practical interest; for example, it fails for some nonhomogeneous Poisson processes which are of interest in reliability, see Sweeting and Adekola (1987). To attack this problem, Sweeting and Adekola adapted Dawid's (1970) method to a more flexible continuity condition on the information function when a shrinking neighborhood is used. But to generalize Dawid's approach, they needed a sequence to measure the order of the information function, and a condition such as their A3 seemed essential for the proof. It then appeared that the weakening of the continuity condition, in order to cover a broader range of applications, necessitated the introduction of other conditions to guarantee the asymptotic normality of the maximum likelihood estimator.

In this paper, we present a novel approach to posterior normality of stochastic processes. Let  $Z_t$  be a suitable parameter transformation and  $h$  be a measurable function. We modify the form of Stein's Identity (Woodroffe (1989, 1992)) and use the new version to write the posterior expectations of  $h(Z_t)$  in a form from which posterior normality can be easily established. Our main finding shows that a condition such as A3 of Sweeting and Adekola (1987) can be avoided. In addition, our choice of shrinking rate is more flexible. The advantage of using our shrinkage is remarked on at the end of Section 4 and illustrated by an example in Section 5.2. The cost we pay for using Stein's Identity is to impose a smoothness condition of  $\xi$  on its compact support.

The approach in this paper is related to Woodroffe (1989, 1992), Woodroffe and Coad (1997), and Weng and Woodroffe (2000), who applied Stein's Identity to obtain posterior expectations and employed a martingale structure to derive integrable posterior expansions. Their parameter transformations are based on the maximum likelihood estimator. The models they considered include linear models with i.i.d. normal errors and exponential families for the i.i.d. case. Although they came from a frequentist perspective, their results implied posterior normality,  $P_\xi^t(Z_t \leq z) \rightarrow \Phi(z)$  as  $t \rightarrow \infty$ , in  $P_{\theta_0}$ -probability, for *a.e.*  $\theta_0$ , where  $\Phi$  denotes the standard normal distribution function. From this point of view, the present paper can be viewed as an extension along this line to a general stochastic process.

The organization of the paper is as follows. In Section 2 we introduce the model. A key observation is that  $Z_t$  converts the likelihood function into a form close to a standard normal density. In Section 3 we derive a modified version of Stein's Identity and present its application to posterior distributions. The difference between the modified version and the original one is remarked on following Lemma 3.2. The conditions and the main results are given in Section 4. We use the conditional exponential family and a nonhomogeneous Poisson process in Section 5 to illustrate applications of this approach.

## 2. 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  $\mathcal{R}^1$ . Assume that the log-likelihood function, denoted by  $\ell_t(\theta)$ , is twice continuously differentiable with respect to  $\theta$ . Throughout let  $\hat{\theta}_t$  be a root of the likelihood equation satisfying  $\ell'_t(\hat{\theta}_t) = 0$ , differentiation is with respect to  $\theta$ . Whenever such a root exists and  $\ell''_t(\hat{\theta}_t) < 0$ , we define

$$\sigma_t = [-\ell''_t(\hat{\theta}_t)]^{-1/2}, \quad (1)$$

$$Z_t = (\theta - \hat{\theta}_t)/\sigma_t; \quad (2)$$

otherwise, define  $\sigma_t$  and  $Z_t$  arbitrarily (in a measurable way).

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

$$\zeta^t(z) \propto \xi^t(\theta(z)) \propto e^{\ell_t(\theta) - \ell_t(\hat{\theta}_t)}\xi(\theta), \quad (3)$$

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

$$\ell_t(\theta) = \ell_t(\hat{\theta}_t) + \frac{1}{2}(\theta - \hat{\theta}_t)^2 \ell''_t(\theta^*), \quad (4)$$

where  $\theta^*$  lies between  $\theta$  and  $\hat{\theta}_t$ . Letting

$$R_t(\theta) = \sigma_t^2[\ell''_t(\hat{\theta}_t) - \ell''_t(\theta^*)], \quad (5)$$

it follows that  $\ell_t(\theta) = \ell_t(\hat{\theta}_t) - z_t^2/2 - z_t^2 R_t(\theta)/2$ . So (3) can be rewritten as

$$\zeta^t(z) \propto \phi(z) f_t(z), \quad (6)$$

where  $f_t(z) = \xi(\theta(z))\exp[-z^2 R_t(\theta)/2]$  and  $\phi(z) = (1/\sqrt{2\pi})\exp[-z^2/2]$ .

Throughout this paper we denote the derivative of  $\xi$  with respect to  $\theta$  by  $\xi'$ , the probability measure and expectation under  $\theta \in \Theta$  by  $P_\theta$  and  $E_\theta$ , and the conditional probability and expectation given data  $x_t$  by  $P_\xi^t$  and  $E_\xi^t$ .

## 3. Modified Stein's Identity

In this section we derive a new version of Stein's Identity and apply it to posterior distributions. This forms the mathematical basis of our approach. Write  $\Phi h = \int h d\Phi$  for functions  $h$  for which the integral is finite. Next let  $\Gamma$  denote a finite signed measure of the form  $d\Gamma = f d\Phi$ , where  $f$  is a real-valued function defined on  $\mathcal{R}$  satisfying  $\Phi|f| = \int |f| d\Phi < \infty$ . For  $p \geq 0$ , denote by  $H_p$  the collection of all measurable functions  $h : \mathcal{R} \rightarrow \mathcal{R}$  for which  $|h(z)| \leq 1 + |z|^p$ , and define  $H = \cup_{p \geq 0} H_p$ . Let

$$Uh(z) = e^{\frac{1}{2}z^2} \int_z^\infty [h(y) - \Phi h] e^{-\frac{1}{2}y^2} dy, \quad (7)$$

for  $-\infty < z < \infty$ .

**Lemma 3.1.** *There are (finite) positive constants  $c_0, c_1, c_2, \dots$  for which  $UH_0 \subseteq c_0H_0$  and  $UH_p \subseteq c_pH_{p-1}$  for all  $p = 1, 2, \dots$*

**Proof.** See Woodroffe (1992, Lemma 1).

**Lemma 3.2.** *(Modified Stein's Identity) Let  $r$  be a nonnegative integer. Suppose  $d\Gamma = fd\Phi$ , where  $f$  is continuous on  $[l, u]$ , continuously differentiable on  $(l, u)$ , and zero outside  $[l, u]$  for some  $-\infty < l < u < \infty$ . If  $\int_l^u |f'(z)|dz < \infty$ ,*

$$\Gamma h = \Gamma 1 \cdot \Phi h + f(l)\phi(l)Uh(l) - f(u)\phi(u)Uh(u) + \int_l^u Uh(z)f'(z)\Phi(dz), \quad (8)$$

for all  $h \in H_r$ .

**Proof.** The proof is a modification of Woodroffe (1989, Proposition 1). Without loss of generality, take  $\Gamma 1 = 1$ . Then

$$\begin{aligned} \Gamma h - \Phi h &= \int_l^u [h(z) - \Phi h]\phi(z)f(z)dz \\ &= \int_l^u [h(z) - \Phi h]\phi(z)\left\{\int_l^z f'(x)dx + f(l)\right\}dz \\ &= f(l)\int_l^u [h(z) - \Phi h]\phi(z)dz + \int_l^u \int_x^u [h(z) - \Phi h]\phi(z)f'(x)dzdx \\ &= f(u)\int_{-\infty}^u [h(z) - \Phi h]\phi(z)dz - f(l)\int_{-\infty}^l [h(z) - \Phi h]\phi(z)dz \\ &\quad - \int_l^u \int_{-\infty}^x [h(z) - \Phi h]\phi(z)f'(x)dzdx, \end{aligned}$$

where the third equality follows by interchanging the orders of integration (justified by the assumed integrability of  $|f'|$ ), the last equality by simple algebra. Then (8) follows by (7).

The major difference between the original version of Stein's Identity (Woodroffe (1989, 1992)) and the modified one is that the former requires  $f$  to be continuously differentiable on  $\mathcal{R}$ , but the latter allows  $f$  to have jump discontinuities at both  $l$  and  $u$ . There are two additional terms in the modified version,  $f(l)\phi(l)Uh(l)$  and  $f(u)\phi(u)Uh(u)$ , which vanish when  $l = -\infty$  and  $u = \infty$ . The reason for considering a restricted  $f$  is to keep  $\xi^{-1}$  bounded below by zero, needed for the proof of Theorem 4.1 below.

From (6), the posterior distribution of  $Z_t$  is of a form suitable for Stein's Identity. In the proposition below, we suppose that there is a measurable  $\hat{\theta}_t = \hat{\theta}_t(X_t)$  and let

$$D_t = \{\ell'_t(\hat{\theta}_t) = 0, \ell''_t(\hat{\theta}_t) < 0\}. \quad (9)$$

**Proposition 3.1.** *Let  $r$  be a nonnegative integer. Suppose that  $\xi$  has a compact support  $\Theta_1 = [a, b] \subseteq \Theta$ ,  $\xi$  is continuous on  $[a, b]$  and is continuously differentiable on  $(a, b)$ . Then for all  $h \in H_r$  a.e. on  $D_t$ ,*

$$E_\xi^t[h(Z_t)] - \Phi h = A_t(a_t) - A_t(b_t) + E_\xi^t[Uh(Z_t) \frac{f'_t(Z_t)}{f_t(Z_t)}], \tag{10}$$

where  $a_t = (a - \hat{\theta}_t)/\sigma_t$ ,  $b_t = (b - \hat{\theta}_t)/\sigma_t$ , and

$$A_t(x) = \frac{f_t(x)\phi(x)Uh(x)}{\int_{a_t}^{b_t} f_t(s)\phi(s)ds}.$$

**Proof.** By Lemma 3.2, it suffices to prove that

$$E_\xi^t \left[ \left| \frac{f'_t(Z_t)}{f_t(Z_t)} \right| \right] < \infty. \tag{11}$$

A straightforward calculation shows that

$$\begin{aligned} \frac{f'_t(Z_t)}{f_t(Z_t)} &= \frac{d\theta}{dz} \left[ \frac{d\xi/d\theta}{\xi}(\theta) + \frac{d[\exp(-z^2 R_t(\theta)/2)]/d\theta}{\exp(-z^2 R_t(\theta)/2)} \right] \\ &= \sigma_t \left[ \frac{\xi'(\theta)}{\xi(\theta)} + \ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t) \right] \\ &= \sigma_t \left[ \frac{\xi'(\theta)}{\xi(\theta)} \right] - Z_t \left[ \frac{\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)}{(\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)} \right]. \end{aligned} \tag{12}$$

For fixed  $x_t$ , the second term of (12) is a continuous function of  $\theta$  and hence bounded on  $\Theta_1$ . For the first term, write  $\xi^t(\theta) = C_t \xi(\theta) e^{\ell_t(\theta)}$ . Then  $C_t$  and  $e^{\ell_t(\theta)}$  are bounded on  $\Theta_1$  for fixed  $x_t$ . Similarly,  $\sigma_t$  is bounded for fixed  $x_t$ . Since  $\xi'$  is continuous, it follows that  $|\xi'(\theta)| < M$  on  $\Theta_1$  for some  $M > 0$ . So

$$\sigma_t E_\xi^t \left[ \left| \frac{\xi'(\theta)}{\xi(\theta)} \right| \right] = \sigma_t \int_{\Theta_1} C_t |\xi'(\theta)| e^{\ell_t(\theta)} d\theta < \infty,$$

establishing (11).

The following Lemma is needed in the proof of Theorem 4.1 below.

**Lemma 3.3.** *Let  $h$  be any bounded measurable function. Then  $\sup_z |zUh(z)| < \infty$ .*

**Proof.** See Stein (1987, Chapter 2).

#### 4. Main Results

In this section we establish the main theorem and remark on the verification of our conditions. Throughout this section,  $\xi$  is assumed to have a compact

support  $\Theta_1 = [a, b] \subseteq \Theta$ ,  $D_t$  is the event defined as in (9), and  $\xrightarrow{P}$  denotes convergence in  $P_{\theta_0}$ -probability as  $t \rightarrow \infty$ , where  $\theta_0 \in [a, b]$ . The following conditions are required.

- (B1)  $\xi$  is continuous on  $[a, b]$ , strictly positive on  $[a, b]$ , and continuously differentiable on  $(a, b)$ .
- (B2)  $P_{\theta_0}(D_t^c) \rightarrow 0$ ,  $\sigma_t \xrightarrow{P} 0$ , and  $\hat{\theta}_t \xrightarrow{P} \theta_0$  as  $t \rightarrow \infty$ .
- (B3) Let  $R_t(\theta)$  be as in (5). There exist some constants  $c > 0$  and  $c' > 0$  such that  $\sup_{|z_t| \leq c} |R_t(\theta)| \leq c'$ .
- (B4) For any  $\theta_1 \in [a, b]$  and  $\theta_1 \neq \theta_0$ ,  $\ell_t(\hat{\theta}_t) - \ell_t(\theta_1) \xrightarrow{P} \infty$ .
- (B5)  $E_\xi^t [ |(\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)) / (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)| ] \xrightarrow{P} 0$ .

**Lemma 4.4.** *Let  $f_t$  be as in (6) and  $a_t$  and  $b_t$  be as in Proposition 3.1. Suppose that (B1)–(B3) hold. Then for each  $\theta_0 \in (a, b)$ , there exists some  $C > 0$  such that  $\int_{a_t}^{b_t} \phi(z)f_t(z)dz > C$  with  $P_{\theta_0}$ -probability tending to 1.*

**Proof.** Write

$$\int_{a_t}^{b_t} \phi(z)f_t(z)dz = \frac{1}{\sqrt{2\pi}} \int_{a_t}^{b_t} \xi(\theta(z))e^{-(z^2/2)[R_t(\theta)+1]}dz,$$

where  $\xi$  is strictly positive on  $[a, b]$  by (B1), and  $e^{-(z^2/2)[R_t(\theta)+1]}$  is bounded below by some positive constant over the interval  $\{|z| \leq c\}$  by (B3). Then from (B2),  $\sigma_t \xrightarrow{P} 0$  and  $\hat{\theta}_t \xrightarrow{P} \theta_0$ , and then for  $\theta_0 \in (a, b)$ ,  $a_t \xrightarrow{P} -\infty$  and  $b_t \xrightarrow{P} \infty$ . The result follows.

**Theorem 4.1.** *Let  $h$  be any bounded measurable function. Suppose that (B1)–(B5) hold. Then for  $\theta_0 \in (a, b)$ ,  $E_\xi^t[h(Z_t)] \xrightarrow{P} \Phi h$ .*

**Proof.** From (10) and (12),

$$\begin{aligned} E_\xi^t[h(Z_t)] - \Phi h &= A_t(a_t) - A_t(b_t) + \sigma_t E_\xi^t \left[ Uh(Z_t) \frac{\xi'(\theta)}{\xi(\theta)} \right] \\ &\quad + E_\xi^t \left\{ Uh(Z_t) Z_t \left[ \frac{\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)}{(\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)} \right] \right\}, \end{aligned} \tag{13}$$

*a.e.* on  $D_t$ . Since  $P_{\theta_0}(D_t^c) \rightarrow 0$  by (B2), it suffices to show that the right side of (13) approaches zero in  $P_{\theta_0}$ -probability. First, note that  $\xi'/\xi$  is bounded on  $[a, b]$  under (B1),  $\sigma_t \xrightarrow{P} 0$  under (B2), and  $|Uh|$  is bounded by Lemma 3.1. So  $\sigma_t E_\xi^t [Uh(Z_t) \frac{\xi'(\theta)}{\xi(\theta)}] \xrightarrow{P} 0$ . Next, (B5) and Lemma 3.3 together imply that the last

term of (13) approaches zero in  $P_{\theta_0}$ -robability. For  $A_t(a_t) - A_t(b_t)$ , we show only that  $A_t(a_t) \xrightarrow{P} 0$ . The result for  $A_t(b_t)$  can be obtained similarly. Observe from Proposition 3.1 that

$$\begin{aligned} A_t(a_t) &= \left\{ \int_{a_t}^{b_t} \phi(x)f(x)dx \right\}^{-1} f_t(a_t)\phi(a_t)Uh(a_t) \\ &= \left\{ \sqrt{2\pi} \int_{a_t}^{b_t} \phi(x)f(x)dx \right\}^{-1} \xi(a)Uh(a_t)e^{-(a_t^2/2)[R_t(a)+1]}. \end{aligned}$$

By (4) and (5), the exponent in above expression can be written as  $-(a_t^2/2)[R_t(a) + 1] = \ell_t(a) - \ell_t(\hat{\theta}_t)$ , which approaches  $-\infty$  in  $P_{\theta_0}$ -probability for  $\theta_0 \in (a, b)$  by (B4); by Lemmas 3.1 and 4.4, there exists some  $c > 0$  such that

$$\left\{ \int_{a_t}^{b_t} \phi(x)f(x)dx \right\}^{-1} |Uh(a_t)| \leq c$$

with  $P_{\theta_0}$ -probability tending to 1. The desired result follows.

Of course if the convergence assumptions in (B2), (B4) and (B5) hold uniformly in compact subsets of  $\Theta_1$ , then the convergence result in Theorem 4.1 also holds uniformly in compact subsets of  $\Theta_1$ . Now applying Theorem 4.1 to  $1_{(a,b]}$ , we obtain the following Corollary.

**Corollary 4.1.**  $P_{\xi}^t(\hat{\theta}_t + a\sigma_t \leq \theta \leq \hat{\theta}_t + b\sigma_t) - [\Phi(b) - \Phi(a)] \xrightarrow{P} 0$ .

Note that it is straightforward to determine whether (B1)-(B4) are satisfied. For (B5), we consider a Taylor's expansion of  $\ell'_t(\theta)$  at  $\hat{\theta}_t$ ,

$$\ell'_t(\theta) = \ell''_t(\theta_t^{**})(\theta - \hat{\theta}_t), \tag{14}$$

where  $\theta_t^{**}$  lies between  $\theta$  and  $\hat{\theta}_t$ . So the integrand of (B5) can be rewritten as

$$\left| \frac{\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)}{(\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)} \right| = \left| 1 - \frac{\ell''_t(\theta_t^{**})}{\ell''_t(\hat{\theta}_t)} \right|. \tag{15}$$

The following condition and Theorem 4.2 are useful in the verification of (B5).

- (S) There exists a sequence of nonnegative functions  $C_t(\theta)$ , possibly tending to infinity as  $t \rightarrow \infty$ , such that

$$\left| \frac{\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)}{(\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)} \right| \leq C_t(\theta)|\hat{\theta}_t - \theta|,$$

with  $P_{\theta_0}$ -probability tending to 1.

**Theorem 4.2.** *Suppose that (B2)–(B4) and (S) hold. If there exists a sequence  $\delta_t > 0$  such that*

$$\sup_{\{|\theta-\theta_0|\leq\delta_t\}} C_t(\theta)|\hat{\theta}_t - \theta| \xrightarrow{P} 0, \tag{16}$$

$$\sup_{\{|\theta-\theta_0|>\delta_t\}} \{\ell_t(\theta) - \ell_t(\hat{\theta}_t) + \log C_t(\theta) - \log \sigma_t\} \xrightarrow{P} -\infty, \tag{17}$$

as  $t \rightarrow \infty$ , then (B5) holds.

**Proof.** Under (S),

$$\begin{aligned} E_\xi^t \left[ \left| \frac{\ell'_t(\theta) - (\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)}{(\theta - \hat{\theta}_t)\ell''_t(\hat{\theta}_t)} \right| \right] &\leq E_\xi^t [C_t(\theta)|\hat{\theta}_t - \theta| 1_{\{|\theta-\theta_0|\leq\delta_t\}}] \\ &\quad + (b - a)E_\xi^t [C_t(\theta) 1_{\{|\theta-\theta_0|>\delta_t\}}] \\ &= I + II, \text{ say,} \end{aligned}$$

where  $I \xrightarrow{P} 0$  by (16). So it suffices to show that  $II \xrightarrow{P} 0$ . Write

$$\begin{aligned} \frac{II}{b - a} &= \frac{\int_{\{|\theta-\theta_0|>\delta_t\}} C_t(\theta)\xi(\theta)e^{\ell_t(\theta)} d\theta}{\int_{\theta \in \Theta_1} \xi(\theta)e^{\ell_t(\theta)} d\theta} \\ &= \frac{\int_{\{|\theta-\theta_0|>\delta_t\}} \xi(\theta)\exp\{\ell_t(\theta) - \ell_t(\hat{\theta}_t) + \log C_t(\theta) - \log \sigma_t\} d\theta}{\sigma_t^{-1} \int_{\theta \in \Theta_1} \xi(\theta)\exp\{\ell_t(\theta) - \ell_t(\hat{\theta}_t)\} d\theta}. \end{aligned}$$

Then the result follows because the numerator approaches zero in  $P_{\theta_0}$ -probability by (17), and the denominator can be rewritten as  $\int_{\{z:\sigma_t z + \hat{\theta}_t \in \Theta_1\}} \xi(\theta)e^{-\frac{z^2}{2}[1+R_t(\theta)]} dz$ , which is bounded below by zero by (B2)-(B4) and Lemma 4.4.

The following proposition will be used in Section 5.1.

**Proposition 4.2.** *Let  $\theta_t^{**}$  be as in (14) and define  $\bar{A}_t(x) = \frac{f_t(x)\phi(x)x}{\int_{a_t}^{b_t} f_t(s)\phi(s)ds}$ . Then*

$$E_\xi^t [(\theta - \hat{\theta}_t)^2 \ell''_t(\theta_t^{**})] = -\left\{ 1 + \bar{A}_t(a_t) - \bar{A}_t(b_t) + E_\xi^t \left[ (\theta - \hat{\theta}_t) \frac{\xi'}{\xi}(\theta) \right] \right\}. \tag{18}$$

**Proof.** Let  $h(x) = x^2$ . Then  $Uh(x) = x$  and  $\Phi h = 1$ . Applying (13) to this  $h$  yields

$$E_\xi^t (Z_t^2) = 1 + \bar{A}_t(a_t) - \bar{A}_t(b_t) + \sigma_t E_\xi^t \left[ Z_t \frac{\xi'(\theta)}{\xi(\theta)} \right] + E_\xi^t \left\{ Z_t^2 \left[ 1 - \frac{\ell''_t(\theta_t^{**})}{\ell''_t(\hat{\theta}_t)} \right] \right\}.$$

Then (18) follows easily from (1) and (2).



It is desirable to compare our conditions with those of Sweeting and Adekola (1987). First of all, a condition such as their *information growth and stability* A3 is avoided here. Note that their A3 corresponds to C1 of Sweeting (1980), and their continuity condition A4 is slightly stronger than C2 there. These two conditions together with twice differentiability of  $\ell(\theta)$  imply the asymptotic normality of the maximum likelihood estimator. See Sweeting (1980). Secondly, our  $\delta_t$  is quite flexible: satisfy (16) and (17). Sweeting and Adekola (1987) used the shrinking rate  $\alpha_t(\theta) = K\{\log J_t(\theta)/J_t(\theta)\}^{1/2}$ , where  $K$  is any positive constant and  $J_t(\theta)$  measures the order of the information function  $-\ell_t''(\theta)$ . Obviously, if the shrinking rate is faster, then it is easier to satisfy (16) but more difficult for (17). Our approach allows us to choose a proper shrinking rate, according to the complexity of (16) and (17).

In Section 5.2 below, we revisit the nonhomogeneous Poisson process discussed by Sweeting and Adekola (1987). The shrinking rate they used is  $\alpha_t(\theta) = Kt^{-1}(2 \log t + \theta t)^{1/2}e^{-\theta t/2}$ . Here we show that (16) and (17) hold with  $\delta_t = 1/t^2$ .

### 5. Examples

We use both homogeneous and nonhomogeneous examples as applications of our results.

#### 5.1. Conditional exponential family

Let  $X_n = \{Y_1, \dots, Y_n\}$  be a sample from the time-homogeneous Markov process whose conditional density function of  $Y_n$  given  $Y_{n-1}$ ,  $f(Y_n|Y_{n-1}, \theta)$ , satisfies

$$\frac{d}{d\theta} \log f(x|y, \theta) = \psi(\theta)H(y)[m(x, y) - \theta] \tag{19}$$

for some functions  $\psi$ ,  $m$  and  $H$ , where  $\psi$  does not involve the  $Y_i$ , and  $m$  and  $H$  do not involve  $\theta$ . Equation (19) defines the class of *conditional exponential family*. Examples include various branching processes and the first order autoregression model. See Heyde and Feigin (1975) and Hall and Heyde (1980).

From (19),  $\ell_n'(\theta) = \sum_{i=1}^n u_i(\theta)$ , where  $u_i(\theta) = \psi(\theta)H(Y_{i-1})[m(Y_i, Y_{i-1}) - \theta]$ . It can be verified that

$$\begin{aligned} \ell_n'(\theta) &= \psi(\theta) \sum_{i=1}^n H(Y_{i-1})(\hat{\theta}_n - \theta) \\ &= \sum_{i=1}^n E_{\theta}(u_i^2(\theta)|\mathcal{F}_{i-1})(\hat{\theta}_n - \theta), \end{aligned} \tag{20}$$

where  $\mathcal{F}_{i-1}$  is the  $\sigma$ -field generated by  $X_{n-1}$  and

$$\hat{\theta}_n = \left[ \sum_{i=1}^n H(Y_{i-1}) \right]^{-1} \sum_{i=1}^n H(Y_{i-1})m(Y_i, Y_{i-1}).$$

From (20),

$$\ell_n''(\theta) = [\psi'(\theta)(\hat{\theta}_n - \theta) - \psi(\theta)] \sum_{i=1}^n H(Y_{i-1}). \quad (21)$$

Suppose that  $\psi(\theta) > 0$  on the parameter space. Then  $\ell_n''(\hat{\theta}_n) < 0$  and  $\hat{\theta}_n$  is clearly the maximum likelihood estimator. Note also that  $\hat{\theta}_n$  is strongly consistent for  $\theta$  provided  $\sum_{i=1}^{\infty} H(Y_{i-1})$  diverges *a.e.* See Heyde and Feigin (1975). Hence (B2) is satisfied. Next, (B3) can be verified by (5) and (21), (B4) by (4) and (21). Then from (14) and (20),

$$\ell_n''(\theta_n^{**}) = -\psi(\theta) \sum_{i=1}^n H(Y_{i-1}). \quad (22)$$

Together with Proposition 4.2 we have  $E_{\xi}^t[(\theta - \hat{\theta}_t)^2] \xrightarrow{P} 0$ . From this, (15), (21) and (22), we can verify (B5).

## 5.2. NHPP model

We revisit a nonhomogeneous Poisson process discussed by Sweeting and Adekola (1987). Let  $N_t$ , the number of events observed by time  $t$ , follow a nonhomogeneous Poisson process with time-dependent intensity function  $\lambda(t)$  over the time interval. So, for each fixed  $t$ ,  $N_t$  is a Poisson with mean  $\int_0^t \lambda(s) ds$ . Suppose that  $\lambda(t) = \theta e^{\theta t}$ , where  $\theta > 0$  is the unknown parameter. Then the log-likelihood function is  $\ell_t(\theta) = N_t \log \theta + \theta \sum_{i=1}^{N_t} x_i - (e^{\theta t} - 1)$ , with derivatives

$$\ell_t'(\theta) = N_t/\theta + \sum_{i=1}^{N_t} x_i - te^{\theta t}, \quad (23)$$

$$\ell_t''(\theta) = -[N_t/\theta^2 + t^2 e^{\theta t}]. \quad (24)$$

It is easily seen from (23) and (24) that  $\ell_t'(\theta) \rightarrow -\infty$  as  $\theta$  approaches  $\infty$ ,  $\ell_t'(\theta) \rightarrow \infty$  as  $\theta$  approaches zero, and  $\ell_t''(\theta) < 0$  for each  $\theta > 0$ . Then there is a maximum likelihood estimator,  $\hat{\theta}_t > 0$ , which uniquely solves the likelihood equation.

Next we show that

$$t|\theta_0 - \hat{\theta}_t| \xrightarrow{P} 0. \quad (25)$$

Note that if we replace  $\theta$  in (23) by  $\hat{\theta}_t$ , then multiply the factor  $e^{-t\theta_0}$  on both sides,

$$e^{t(\hat{\theta}_t - \theta_0)} = \frac{\sum_{i=1}^{N_t} x_i}{te^{t\theta_0}} + \frac{N_t}{t\hat{\theta}_t e^{t\theta_0}}.$$

Since  $N_t t^{-1} e^{-t\theta_0} \rightarrow 0$  *a.e.* ( $P_{\theta_0}$ ) and  $\sum_{i=1}^{N_t} x_i t^{-1} N_t^{-1} \rightarrow 1$  *a.e.* ( $P_{\theta_0}$ ), (25) follows.

From (24), the observed Fisher information is

$$-\ell_t''(\hat{\theta}_t) = N_t/\hat{\theta}_t^2 + t^2 e^{\hat{\theta}_t t}, \quad (26)$$

so that (B2) is satisfied.

From (14) and (15)

$$\left| \frac{\ell_t''(\hat{\theta}_t) - \ell_t''(\theta_t^{**})}{\ell_t''(\hat{\theta}_t)} \right| \leq \frac{\sup |\ell_t'''(\nu)|}{|\ell_t''(\hat{\theta}_t)|} |\hat{\theta}_t - \theta|, \tag{27}$$

where the supremum is over  $\{\nu : \nu \text{ lies between } \theta \text{ and } \hat{\theta}_t\}$ . Now simple algebra on (24) shows that the right side of (27) is bounded by  $\gamma t e^{t|\hat{\theta}_t - \theta|} |\hat{\theta}_t - \theta|$ , where  $\gamma > 0$  is some constant. Thus (S) is satisfied with  $C_t(\theta) = \gamma t e^{t|\hat{\theta}_t - \theta|}$ .

Recall that  $R_t(\theta) = \sigma_t^2 [\ell_t'(\hat{\theta}_t) - \ell_t'(\theta_t^*)]$  as in (5). So, similar to (27), we have  $|R_t(\theta)| \leq \gamma t e^{t|\hat{\theta}_t - \theta|} |\hat{\theta}_t - \theta|$ . Observe that the right side is bounded by  $\gamma |Z_t| e^{|Z_t|}$  because  $Z_t = (\theta - \hat{\theta}_t)/\sigma_t$  with  $\sigma_t = [N_t/\hat{\theta}_t^2 + t^2 e^{\hat{\theta}_t t}]^{-1/2}$ . It is then easily seen that (B3) is satisfied with some  $c > 0$  and  $c' > 0$ .

The verification of (B4) is straightforward. For (B5), we need only show that (16) and (17) hold and then apply Theorem 4.2. Setting  $\delta_t = 1/t^2$ ,

$$\begin{aligned} \sup_{\{|\theta - \theta_0| \leq \delta_t\}} C_t(\theta) |\hat{\theta}_t - \theta| &\leq \sup_{\{|\theta - \theta_0| \leq \delta_t\}} \gamma t (|\theta - \theta_0| + |\theta_0 - \hat{\theta}_t|) e^{t|\theta - \theta_0| + t|\theta_0 - \hat{\theta}_t|} \\ &\leq \gamma \left( \frac{1}{t} + t|\theta_0 - \hat{\theta}_t| \right) e^{1/t + t|\theta_0 - \hat{\theta}_t|}, \end{aligned}$$

which approaches zero by (25). So (16) is satisfied. Next, from (27),

$$\begin{aligned} &\sup_{\{|\theta - \theta_0| > \delta_t\}} \{ \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \log C_t(\theta) - \log \sigma_t \} \\ &\leq \sup_{\{|\theta - \theta_0| > \delta_t\}} \{ \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \log(\gamma t) + t|\hat{\theta}_t - \theta| - \log \sigma_t \}. \end{aligned}$$

Since  $\ell(\theta)$  is concave, the supremum occurs at  $\theta = \theta_0 \pm \delta_t$  and it is straightforward to verify (17).

**Acknowledgements**

The author would like to thank the referee and an associate editor for a careful report, which helped improve the quality of the paper. The author is partially supported by the National Science Council of Taiwan, 90-2118-M-004-012.

**References**

Basawa, I. V. and Rao, B. L. S. P. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, London.  
 Chen, C. F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47**, 540-546.

- Dawid, A. P. (1970). On the limiting normality of posterior distributions. *Proc. Camb. Phil. Soc.* **67**, 625-633.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
- Heyde, C. C. and Feigin, P. D. (1975). On efficiency and exponential families in stochastic process estimation. In *Statistical Distributions in Scientific Work 1* (Edited by G. P. Patil and S. Kotz and J. K. Ord), 227-240. Reidel, Dordrecht.
- Heyde, C. C. and 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.
- Le Cam, L. (1953). On some asymptotic properties of maximum likelihood and related Bayes estimates. *University of California Publications in Statist.* **1**, 277-330.
- Stein, C. (1987). *Approximate Computation of Expectations*. IMS, Hayward, Calif.
- Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375-1381.
- Sweeting, T. J. and 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. and Woodroffe, M. (2000). Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels. *Statist. Sinica* **10**, 693-713.
- Woodroffe, M. (1989). Very weak expansions for sequentially designed experiments: linear models. *Ann. Statist.* **17**, 1087-1102.
- Woodroffe, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2**, 91-111.
- Woodroffe, M. and Coad, D. S. (1997). Corrected confidence sets for sequentially designed experiments. *Statist. Sinica* **7**, 53-74.

Department of Statistics, National Chengchi University, Taipei, 11623, Taiwan.

E-mail: chweng@nccu.edu.tw

(Received June 2001; accepted December 2002)