Ann. Inst. Statist. Math.
30 (1978), Part A, 243-261

THE ASYMPTOTIC BEHAVIOUR OF MAXIMUM LIKELIHOOD ESTIMATORS FOR STATIONARY POINT PROCESSES

YOSIHIKO OGATA

(Received May 23, 1977; revised June 26, 1978)

1. Introduction

Let $P(\cdot)$ be an orderly stationary point process on the set of points $\omega=\{t_j;\ j=0, \pm 1, \pm 2, \cdots\} \in \Omega$ with no fixed atoms on the real line R. Here we take $\cdots<t_{-1}<0\leqq t_0<t_1<\cdots$, and it is assumed that the set ω has no limit point.

The counting measure $N(A)=N(A, \omega)$ is defined for each bounded Borel subset A of R to be the cardinal of the set $\omega \cap A$.

The complete intensity function and intensity function of the point process on the σ-algebra $H_{0,t}$ are defined respectively as follows:

$$(1.1)\qquad \begin{aligned} \lambda(t, \omega)&=\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\,[N\{[t, t+\delta)\}>0 \,|\, H_{-\infty,t}] \\ \lambda^*(t, \omega)&=\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\,[N\{[t, t+\delta)\}>0 \,|\, H_{0,t}]=\mathrm{E}\,\{\lambda(t, \omega) \,|\, H_{0,t}\} \end{aligned}$$

where $H_{s,t}$ denotes the σ-field generated by $\{N(u, t];\ s<u\leqq t\}$.

We consider a family of parametrized stationary complete intensity processes $\{\lambda_\theta(t, \omega);\ \theta \in \Theta \subset R^d\}$ which are assumed to correspond uniquely to the stationary point processes $\{P_\theta;\ \theta \in \Theta\}$. Thus we have the exact log-likelihood on the interval $[0, T]$ as follows:

$$(1.2)\qquad L_T^*(\theta)=-\int_0^T \lambda_\theta^*(t, \omega)dt+\int_0^T \log \lambda_\theta^*(t, \omega)dN(t)\ .$$

The maximum likelihood estimator $\hat{\theta}_T=\hat{\theta}(t_i;\ 0\leqq t_i\leqq T)$ is defined by the estimator of θ which maximizes the exact likelihood (1.2) under observations from the stationary point process P_{θ_0}. Several asymptotic properties of likelihood procedures for the point processes are suggested in [10]. In this paper we will give some proofs, and develop the asymptotic properties of the maximum likelihood estimator. For this purpose we consider theoretically a conditional log-likelihood under the information from the infinite past

$$(1.3)\qquad L_T(\theta)=-\int_0^T \lambda_\theta(t,\omega)dt+\int_0^T \log\lambda_\theta(t,\omega)dN(t)\ ,$$

and it will be seen later that we can identify $L_T^*(\theta)$ with $L_T(\theta)$ for sufficiently large T under the Assumptions C given in Section 2. In Section 2 assumptions are collected together, and examples which satisfy them are given. In Section 3 it will be proved that the maximum likelihood estimator is consistent, asymptotically normal, and efficient. In the last section Poisson processes will be characterised by a maximum likelihood estimator of parametrized renewal processes.

2. Assumptions and examples

Three groups of assumptions are given. Assumptions A are for observations. Assumptions B are the regularity conditions for the parametric family of complete intensity processes. Assumptions C are given for the relations between λ_θ^* and λ_θ in order that some limit theorems for some functional of λ_θ^* remain valid.

For convenience, the following notation will be introduced. Let $\partial\log\lambda_\theta(t,\omega)/\partial\theta_i$ and $\partial\lambda_\theta(t,\omega)/\partial\theta_i$ be denoted by $\partial\log\lambda/\partial\theta_i$ and $\partial\lambda/\partial\theta_i$ respectively, with similar notation being employed for second- and third-order derivatives. In addition, $\partial\log\lambda/\partial\lambda_i|_{\theta'}$ will be denoted the value of $\partial\log\lambda/\partial\theta_i$ at the point $\theta'\in\Theta$ with the same convention used for other functions. Instead of $\mathrm{E}_{\theta_0}(\cdot)$ and $P_{\theta_0}(\cdot)$, where θ_0 is the true value of the parameter, let us agree to write $\mathrm{E}(\cdot)$ and $P(\cdot)$.

ASSUMPTIONS A.

(A1) The point process is stationary, ergodic and absolutely continuous with respect to the standard Poisson process on any finite interval.

(A2) The point process is orderly; $\lim_{\delta\to 0}(1/\delta)\,\mathrm{P}[N\{[0,\delta)\}\geqq 2]=0$.

(A3) $\mathrm{E}[\sup_{0<\delta\leqq 1}(1/\delta)N([0,\delta])^2]<\infty$.

We say the process $\xi=\{\xi(t,\omega);\ t\geqq 0\}$ is adapted (with respect to the underlying point process $N(\cdot,\omega)$) if for fixed $t\geqq 0$ $\xi(t,\omega)$ is $H_{-\infty,t}$-measurable. Further, we say the process ξ is predictable if the mapping $\xi: R_+\times\Omega\to R$ is measurable with respect to the $P(\cdot)$-completed σ-algebra which is generated by left continuous functions from R_+ into R (see [7], p. 2 for example). It will be sufficient for our purpose to note that the adapted process ξ is predictable if the sample paths $\xi(t,\omega)$ are left continuous on $(0,\infty)$ for a.s. ω.

In Section 1 we have already used the stochastic Stieltjes integrals $\int_0^T \xi(t,\omega)dN(t)=\sum_{0\leqq t_i\leqq T}\xi(t_i(\omega),\omega)$ which are defined pathwise for the measurable process ξ (see [3], p. 89). It should be noted that for any finite predictable process ξ satisfying $\int_0^T \mathrm{E}\{\lambda_{\theta_0}(t, \omega)|\xi(t, \omega)|\}dt<\infty$ we are allowed to do the following calculation:

$$
\text{(2.1)} \qquad \mathrm{E}\left[\int_0^T \xi(t, \omega)dN(t)\right]=\mathrm{E}\left[\int_0^T \xi(t, \omega)\,\mathrm{E}\{dN(t)\,|\,H_{-\infty,t}\}\right] = \mathrm{E}\left[\int_0^T \xi(t, \omega)\lambda_{\theta_0}(t, \omega)dt\right] .
$$

Thus for integrands of finite predictable processes we can use the formal relation $\mathrm{E}\{dN(t)\,|\,H_{-\infty,t}\}=\lambda_{\theta_0}(t, \omega)dt$. Similarly, $\mathrm{E}\{dN(t)\,|\,H_{0,t}\}=\lambda^*_{\theta_0}(t, \omega)dt$.

The proof of (2.1) is directly derived from a theorem of [7], p. 23.

We next list a variety of regularity conditions which will be needed at different places in the sequel.

ASSUMPTIONS B.

(B1) Θ is a compact metric space with some metric ρ, and $\Theta\subset R^d$.

(B2) λ_θ is predictable for all Θ. $\lambda_\theta(t, \omega)$ is continuous in θ, and $\lambda_\theta(0, \omega)>0$ a.s. ω for any $\theta\in\Theta$.

(B3) $\lambda_{\theta_1}(0, \omega)=\lambda_{\theta_2}(0, \omega)$ a.s. if and only if $\theta_1=\theta_2$.

(B4) $\partial \log\lambda/\partial\theta_i$, $\partial^2\log\lambda/\partial\theta_i\partial\theta_j$ and $\partial^3\log\lambda/\partial\theta_i\partial\theta_j\partial\theta_k$ exist and are continuous in θ for all $i, j, k=1, 2,\cdots, d$, $t\in R_+$ and a.s. $\omega\in\Omega$. $\partial\lambda/\partial\theta_i$ and $\partial^2\lambda/\partial\theta_i\partial\theta_j$ have finite second moments for any $\theta\in\Theta$.

(B5) For any $\theta\in\Theta$ there exist a neighbourhood $U=U(\theta)$ of θ such that for all $\theta'\in U$,

$$
|\lambda_{\theta'}(0, \omega)|\leqq\Lambda_0(\omega) \quad\text{and}\quad |\log\lambda_{\theta'}(0, \omega)|\leqq\Lambda_1(\omega) ,
$$

where Λ_0 and Λ_1 are random variables with finite 2nd moments.

(B6) For every $\theta\in\Theta$, the matrix $I(\theta)=\{I_{ij}(\theta)\}_{i,j=1,\cdots,d}$ with $I_{ij}(\theta)=\mathrm{E}\{(1/\lambda)\cdot(\partial\lambda/\partial\theta_i)\cdot(\partial\lambda/\partial\theta_j)\}$ is nonsingular, and each element $(1/\lambda)\cdot(\partial\lambda/\partial\theta_i)\cdot(\partial\lambda/\partial\theta_j)$ has finite 2nd moment.

(B7) For any $\theta\in\Theta$, there exists a neighbourhood U of θ such that if

$$
\max_{1\leqq i,j,k\leqq d}\ \sup_{\theta'\in U}\left|\frac{\partial^3\lambda}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|=H(t, \omega) ,
$$

$$
\max_{1\leqq i,j,k\leqq d}\ \sup_{\theta'\in U}\left|\frac{\partial^3\log\lambda}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|=G(t, \omega)
$$

then $\mathrm{E}\{H(0, \omega)\}<\infty$ and $\mathrm{E}\{\lambda_{\theta_0}(0, \omega)^2G(0, \omega)^2\}<\infty$.

It is known by [3] that if the intensity function of the point process on the half-line exists, then a predictable version of the intensity function can always be chosen. In (B2) we assume that the same is true for the complete intensity function of a stationary process. By the continuity conditions (B4) all of the derivatives are separable with respect to θ. From this fact it can be shown that the processes in (B4) and their supremum with respect to $\theta \in U$ are also predictable.

A further set of assumptions, rather technical in character, are needed for the stochastic approximations of λ_θ by λ_θ^*. Condition (C1) is needed for the proof of consistency, (C2) and (C4) for the discussion of the Hessian, and (C3) for the proof of asymptotic normality, etc. Each assumption (ii) of (C1), (C2) and (C4) is for the uniform integrability conditions with respect to the true probability P_{θ_0}. We will make use of the theorems T20 and T21 in Meyer [14] Chapter 2.

ASSUMPTIONS C.

(C1) For any $\theta \in \Theta$ there is a neighbourhood U of θ such that

(i) $\sup_{\theta' \in U} |\lambda_{\theta'}(t, \omega) - \lambda_{\theta'}^*(t, \omega)| \to 0$ in probability as $t \to \infty$,

(ii) $\sup_{\theta' \in U} |\log \lambda_{\theta'}^*(t, \omega)|$ has, for some $\alpha > 0$, finite $(2+\alpha)$th moment uniform bounded with respect to t.

(C2) (i) For any $\theta \in \Theta$ and $i, j = 1, 2, \cdots, d$ the following tend to zero in probability as $t \to \infty$;

$$\lambda_\theta - \lambda_\theta^*, \quad \frac{\partial \lambda}{\partial \theta_i} - \frac{\partial \lambda^*}{\partial \theta_i} \quad \text{and} \quad \frac{\partial^2 \lambda}{\partial \theta_i \partial \theta_j} - \frac{\partial^2 \lambda^*}{\partial \theta_i \partial \theta_j}.$$

(ii) For any $\theta \in \Theta$ the following have, for some $\alpha > 0$, finite $(2+\alpha)$th moments uniformly bounded with respect to t,

$$\frac{\lambda_\theta}{\lambda_\theta^*}, \quad \frac{1}{\lambda^*} \frac{\partial \lambda^*}{\partial \theta_i} \frac{\partial \lambda^*}{\partial \theta_j} \quad \text{and} \quad \frac{\partial^2 \lambda^*}{\partial \theta_i \partial \theta_j}, \qquad i, j = 1, 2, \cdots, d.$$

(C3) For any $\theta \in \Theta$ and $i = 1, 2, \cdots, d$, as $T \to \infty$

$$\mathrm{E}\left\{\frac{1}{\sqrt{T}} \int_0^T \left| \frac{\partial \lambda}{\partial \theta_i} - \frac{\partial \lambda^*}{\partial \theta_i} \right| dt\right\} \to 0 \quad \text{and}$$

$$\mathrm{E}\left\{\frac{1}{\sqrt{T}} \int_0^T |\lambda_\theta - \lambda_\theta^*| \left| \frac{1}{\lambda_\theta^*} \frac{\partial \lambda^*}{\partial \theta_i} \right| dt\right\} \to 0.$$

(C4) For any $\theta \in \Theta$ and $i, j, k = 1, 2, \cdots, d$ there is a neighbourhood U of θ such that

(i) $\sup_{\theta' \in U} \left| \frac{\partial^3 \lambda}{\partial \theta_i \partial \theta_j \partial \theta_k} - \frac{\partial^3 \lambda^*}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \to 0$ in probability as $t \to \infty$,

(ii) $\sup_{\theta' \in U} \frac{\partial^3 \lambda^*}{\partial \theta_i \partial \theta_j \partial \theta_k}$ and $\sup_{\theta' \in U} \frac{\partial^3 \log \lambda^*}{\partial \theta_i \partial \theta_j \partial \theta_k}$ have finite $(2+\alpha)$th moments which are uniformly bounded with respect to t.

We now give some illustrative examples for our results.

Example 1. *Stationary Poisson process*

It follows directly from Theorem 2 of [6], for example, that the complete intensity process is deterministic and positive constant if and only if the corresponding point process is stationary Poisson. Put $\lambda_\theta(t, \omega)=\mu(\theta)$, then the log-likelihood function on the interval $[0, T]$ is given

$$L_T^*(\theta)=-\mu(\theta)T+N(0, T)\log\mu(\theta) \tag{2.2}$$

and the maximum likelihood estimator of $\mu(\theta)$ is given by $N(0, T)/T$.

Example 2. *Stationary delayed renewal process*

Suppose the parametrized survivor functions $1-F_\theta(t)$, $t\geqq 0$, are given. In this case the complete intensity function coincides with the hazard function $\lambda_\theta(t, \omega)=f_\theta(t-t^*(\omega))/\{1-F_\theta(t-t^*(\omega))\}$, where $f_\theta(\cdot)$ is the left continuous p.d.f. of $F_\theta(\cdot)$, and $t^*(\omega)$ is the last occurrence time such that $t^*(\omega)<t$. Then it is easily seen that $\lambda_\theta(t, \omega)$ is a predictable process (note $\lambda(t, \omega)$ is not a predictable process if $t^*(\omega)$ is defined such that $t^*(\omega)\leqq t$). The stationary joint distribution of forward and backward recurrence times is given by $\mathrm{P}(X\leqq u,\ Y\leqq v)=\mu_\theta^{-1}\int_0^u\{F_\theta(v+w)-F(w)\}dw$, where $\mu_\theta=\int_0^\infty tdF_\theta(t)$ (see [12]). Since $\mathrm{P}(X\leqq v\,|\,Y>t)=\int_t^\infty F(v+w)-F(w)\}\Big/\int_t^\infty\{1-F(w)\}dw$, we see that $\lambda^*(t, \omega)=\mathrm{E}\{\lambda(t, \omega)\,|\,H_{0,t}\}=\{1-F(t)\}\Big/\int_t^\infty\{1-F(w)\}dw$ if there are no points in $(0, t]$, otherwise $\lambda^*(t, \omega)=\lambda(t, \omega)$. Thus we have the exact log-likelihood function for the observation $0\leqq t_0<\cdots<t_{n-1}\leqq T$ on the interval $[0, T]$,

$$L_T^*(\theta)=\log\mu_\theta^{-1}\{1-F_\theta(t_0)\}+\sum_{i=1}^{n-1}\log f_\theta(t_i-t_{i-1})+\log\{1-F_\theta(T-t_{n-1})\}\ . \tag{2.3}$$

Let us briefly check the assumptions for renewal processes. (C1)-(i) and (C2)-(i) are automatically satisfied because, for example,

$$\mathrm{P}\{\sup_{\theta'\in U}|\lambda_{\theta'}^*(t, \omega)-\lambda_{\theta'}(t, \omega)|>\varepsilon\}\leqq\mathrm{P}\{\text{no events in }(0, t]\}$$

$=(1/\mu)\int_t^\infty\{1-F(s)\}ds=J(t)$ (say) by the facts above. By the similar idea and using Cauchy-Schwartz inequality we also see that the conditions in (C3) are satisfied if $(1/\sqrt{T})\int_0^T J(t)^{1/2}dt\to 0$, which in turn are satisfied if $F(t)$ has a finite variance. Integrability in Assumptions B and C depend on the decreasing rate of $J(t)$ or $1-F(t)$.

Example 3. *Wold process* (*Markov-dependent intervals*)

The 0-memory Wold process is defined formally with a complete intensity function which is independent of any of the past occurrence times, and m-memory is defined with a complete intensity function $\lambda(t,\omega)=\lambda(t-t_{-1},\cdots,t-t_{-m})$ which depends only on the m most recent occurrence times $t_{-1},\cdots,t_{-m}$. These processes are extensions of the renewal process with the hazard function, and exist as finite order Markov processes. The relation between the conditional hazard functions and survivor functions is given in the last chapter of [9]. For example, if the process is 2-memory and $P_{10}(t-t_{-1},t_{-1}-t_{-2})$ is the probability of no points in the interval (t_{-2},t_{-1}) and one point in the interval $(t_{-1},t]$ under the condition that we have just two most recent points t_{-2}, t_{-1}, then

$$\lambda(t,\omega)=h(\tau,\sigma)=-\frac{\partial}{\partial\tau}\left[\log\left\{\frac{\partial}{\partial\sigma}\left(\frac{\partial}{\partial\sigma}-\frac{\partial}{\partial\tau}\right)P_{10}(\tau,\sigma)\right\}\right]$$

where $\tau=t-t_{-1}$ and $\sigma=t_{-1}-t_{-2}$. It should be noted that the complete intensity function coincides with the conditional hazard function in the case of the finite memory processes.

Example 4. *Hawkes' self-exciting process*

Consider the point process which is formally defined with a complete intensity function of the form

$$(2.4)\qquad \begin{aligned}\lambda_\theta(t,\omega)&=\nu+\int_{-\infty}^{t}\gamma_\mu(t-u)dN(u)\,,\qquad \theta=(\nu,\mu)\\ &=\nu+\sum_{t_i<t}\gamma_\mu(t-t_i)\end{aligned}$$

where $\nu>0$, $\gamma_\mu(u)\geqq 0$, γ_μ is left continuous for $u\geqq 0$, and $\int_0^\infty\gamma_\mu(u)du<1$. Note that the range of the integral in (2.4) is $(-\infty,t)$, in other words, the sum is taken for all integers i such that $t_i<t$; this guarantees the predictability of $\lambda_\theta(t,\omega)$. If we take the integral on the range $(-\infty,t]$, then $\lambda_\theta(t,\omega)$ is no longer predictable. It is easily seen that this difference between $(-\infty,t)$ and $(-\infty,t]$ also appears significantly when we calculate the likelihood under given data. It was shown in [5] that the stationary self-exiting point process exists uniquely as a generalized Poisson cluster process, in which the cluster structure is that of a birth process. For a simple special case $\gamma(u)=\alpha e^{-\beta u}$ $(\alpha<\beta)$, Ozaki [8] performed simulations for given parameters $\theta=(\nu,\alpha,\beta)$ such that $\nu>0$, $\alpha<\beta$, and successfully obtained maximum likelihood estimates from the simulated data. It is easily seen that Assumption C is always satisfied by the simple case above.

In general, we see for (C1)-(i) that $E\{\sup_{\theta'\in U}|\lambda_{\theta'}^*(t,\omega)-\lambda_{\theta'}(t,\omega)|\}\leqq 2E\{\lambda_{\theta_0}\}\cdot\int_t^\infty\sup_{(\nu,\mu)\in U}\gamma_\mu(u)du$ with the rate of decrease of the left-hand side depending on the rate of decrease of $\sup_{(\nu,\mu)\in U} \gamma_\mu(u)$. Assumption (C2)-(i) and (C3) are satisfied similarly. For the integrability conditions it should be noted that $\lambda_\theta(t, \omega)$ and $\lambda_\theta^*(t, \omega)$ in this example are uniformly bounded away from 0. Thus we can see that integrability conditions depend on the rate of decrease of the tail of γ_μ, $\partial\gamma_\mu/\partial\theta_i$, $\partial^2\gamma_\mu/\partial\theta_i\partial\theta_j$ and $\partial^3\gamma_\mu/\partial\theta_i\partial\theta_j\partial\theta_k$, and are certainly satisfied when γ_μ has exponential form. Finally, though $\lambda_\theta^*(t, \omega)=\mathrm{E}\{\lambda_\theta(t, \omega)\,|\,H_{0,t}\}$ gives the best approximation of $\lambda_\theta(t, \omega)$, it is difficult to get the exact likelihood numerically. So, practically, we can use

$$\lambda_\theta^{**}(t, \omega)=\nu+\int_0^t \gamma_\mu(t-u)dN(u)=\nu+\sum_{0\leqq t_i<t} \gamma_\mu(t-t_i) .$$

This is predictable and satisfies the assumptions similarly.

3. Asymptotic properties of the likelihood procedure

LEMMA 1. *Under the Assumptions* A *we have*

(i) $\mathrm{E}\{N(0, 1)^2\}<\infty$,

(ii) $\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\{N(t, t+\delta)\geqq 2\,|\,H_{-\infty,t}\}=0$,

(iii) $\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{E}\{N(t, t+\delta)^2\,|\,H_{-\infty,t}\}=\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{E}\{N(t, t+\delta)\,|\,H_{-\infty,t}\}$.

PROOF. (i) is obtained directly from (A3). For the proof of (ii) note that

$$\frac{1}{\delta} \mathrm{P}\{N(t, t+\delta)\geqq 2\,|\,H_{-\infty,t}\}\leqq\frac{1}{\delta} \sum_{i=1}^{\infty} i^2\, \mathrm{P}\{N(t, t+\delta)=i\,|\,H_{-\infty,t}\}$$
$$\leqq\frac{1}{\delta} \mathrm{E}\{N(t, t+\delta)^2\,|\,H_{-\infty,t}\} .$$

Then by (A2), (A3) and the dominated convergence theorem

$$\mathrm{E}\left[\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\{N(t, t+\delta)\geqq 2\,|\,H_{-\infty,t}\}\right]=\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\{N(t, t+\delta)\geqq 2\}=0 .$$

Therefore with probability one

$$\lim_{\delta\to 0} \frac{1}{\delta} \mathrm{P}\{N(t, t+\delta)\geqq 2\,|\,H_{-\infty,t}\}=0 .$$

Proof of (iii). By (A3) we have

$$\sum_{k=0}^{n-1} N\left(\frac{k}{n}, \frac{k+1}{n}\right)^2-\sum_{k=0}^{n-1} N\left(\frac{k}{n}, \frac{k+1}{n}\right)\to 0 \qquad \text{a.s. as } n\to 0 ,$$

since with probability one each interval $[k/n, (k+1)/n]$ will ultimately have either zero or one event in it. Also we have

$$N(0, 1)=\sum_{k=0}^{n-1} N\left(\frac{k}{n}, \frac{k+1}{n}\right) \quad \text{and}$$

$$\sum_{k=0}^{n-1} N\left(\frac{k}{n}, \frac{k+1}{n}\right)^2 \leqq \left\{\sum_{k=0}^{n-1} N\left(\frac{k}{n}, \frac{k+1}{n}\right)\right\}^2 = N(0, 1)^2 ,$$

since all terms are non-negative. It therefore follows from the dominated convergence theorem and stationarity that

$$n\,\mathrm{E}\left\{N\left(0, \frac{1}{n}\right)^2\right\} \to \mathrm{E}\{N(0, 1)\} \qquad \text{as } n\to\infty ,$$

that is, for any $t \geqq 0$

$$\lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta)^2\} = \lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta)\} = \mathrm{E}\{N(0, 1)\} .$$

Thus we have

$$\mathrm{E}\left[\lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta)^2 \mid H_{-\infty,t}\} - \lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta) \mid H_{-\infty,t}\}\right]$$
$$= \lim_{\delta\to 0} \left[\frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta)^2\} - \frac{1}{\delta}\,\mathrm{E}\{N(t, t+\delta)\}\right] = 0 ,$$

and the integrand above is always non-negative since $N(t, t+\delta)$ is non-negative integer-valued. This completes the proof.

THEOREM 1. *Under the Assumptions* A, (B2) *and* (B4)–(B6)

$$\mathrm{E}\left\{\frac{\partial L_T(\theta)}{\partial \theta_i}\right\}_{\theta=\theta_0} = 0 , \qquad i=1, 2, \cdots, d \tag{3.1}$$

and

$$\mathrm{E}\left\{\frac{\partial L_T(\theta)}{\partial \theta_i}\frac{\partial L_T(\theta)}{\partial \theta_j}\right\}_{\theta=\theta_0} = -\mathrm{E}\left\{\frac{\partial^2 L_T(\theta)}{\partial \theta_i \partial \theta_j}\right\}_{\theta=\theta_0} = T\,\mathrm{E}\left\{\frac{1}{\lambda}\frac{\partial \lambda}{\partial \theta_i}\frac{\partial \lambda}{\partial \theta_j}\right\}_{\theta=\theta_0} , \qquad i, j=1, 2, \cdots, d . \tag{3.2}$$

PROOF. By (ii) of Lemma 1

$$\lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{E}\{N[t, t+\delta) \mid H_{-\infty,t}\} = \lim_{\delta\to 0} \frac{1}{\delta}\,\mathrm{P}\{N[t, t+\delta)=1 \mid H_{-\infty,t}\}$$
$$= \lambda_{\theta_0}(t, \omega) .$$

This means formally that $\mathrm{E}\{dN(t) \mid H_{-\infty,t}\} = \lambda_{\theta_0}(t, \omega)dt$. Thus we have by

(B2), (B4), (B5) and (2.1) that

$$\mathrm{E}\left\{\int_0^T \frac{\partial \log \lambda}{\partial \theta_i} dN(t)\right\}_{\theta=\theta_0} = \mathrm{E}\left[\int_0^T \frac{1}{\lambda}\frac{\partial \lambda}{\partial \theta_i}\,\mathrm{E}\{dN(t)\,|\,H_{-\infty,t}\}\right]_{\theta=\theta_0}$$

$$= \mathrm{E}\left\{\int_0^T \frac{\partial \lambda}{\partial \theta_i} dt\right\}_{\theta=\theta_0} = T\,\mathrm{E}\left\{\frac{\partial \lambda}{\partial \theta_i}\right\}_{\theta=\theta_0}$$

for $i=1, 2, \cdots, d$. This implies (3.1). Similarly we have for $i, j=1, 2, \cdots, d$,

$$\mathrm{E}\left\{\frac{\partial^2 L_T(\theta)}{\partial\theta_i\partial\theta_j}\right\}_{\theta=\theta_0}$$

$$= \mathrm{E}\left\{-\int_0^T \frac{\partial^2\lambda}{\partial\theta_i\partial\theta_j}dt + \int_0^T \frac{1}{\lambda}\frac{\partial^2\lambda}{\partial\theta_i\partial\theta_j}dN(t) - \int_0^T \frac{1}{\lambda^2}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}dN(t)\right\}_{\theta=\theta_0}$$

$$= \mathrm{E}\left\{-\int_0^T \frac{1}{\lambda}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}dt\right\}_{\theta=\theta_0} = -T\,\mathrm{E}\left\{\frac{1}{\lambda}\frac{\partial\lambda\partial\lambda}{\partial\theta_i\partial\theta_j}\right\}_{\theta=\theta_0}.$$

On the other hand from (B5) and (B6) each of the following terms exists.

$$\mathrm{E}\left\{\frac{\partial L_T(\theta)}{\partial\theta_i}\frac{\partial L_T(\theta)}{\partial\theta_j}\right\}_{\theta=\theta_0}$$

$$= \mathrm{E}\left[\int_0^T\int_0^T \frac{\partial\lambda(s)}{\partial\theta_i}\frac{\partial\lambda(t)}{\partial\theta_j}\left\{dsdt - \frac{dN(s)dt}{\lambda(s)} - \frac{dsdN(t)}{\lambda(t)} + \frac{dN(s)dN(t)}{\lambda(s)\lambda(t)}\right\}\right]_{\theta=\theta_0}$$

$$= \mathrm{E}\left\{\iint_{\{0\le s<t\le T\}} + \iint_{\{0\le t<s\le T\}} + \iint_{\{0\le s=t\le T\}}\right]_{\theta=\theta_0}$$

$$= I_1 + I_2 + I_3\,.$$

From the relation (2.1) and

$$\mathrm{E}\{dN(s)dN(t)\,|\,H_{-\infty,t}\} = dN(s)\,\mathrm{E}\{dN(t)\,|\,H_{-\infty,t}\} = \lambda(t, \omega)dN(s)dt$$

for $s<t$, we have $I_1=0$. Similarly $I_2=0$. For the third term I_3 note that (iii) of Lemma 1 means formally that

$$\mathrm{E}[\{dN(t)\}^2\,|\,H_{-\infty,t}] = \mathrm{E}\{dN(t)\,|\,H_{-\infty,t}\} = \lambda(t, \omega)dt\,.$$

Thus we see from (2.1) that

$$I_3 = \mathrm{E}\left\{\int_0^T \frac{1}{\lambda}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}dt\right\}_{\theta=\theta_0} = T\,\mathrm{E}\left\{\frac{1}{\lambda}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}\right\}_{\theta=\theta_0}.$$

This completes the proof.

Remark. We can get the same results as Theorem 1 except the last equality for λ^* and L_T^*, if we replace $H_{-\infty,t}$ with $H_{0,t}$ in (2.1) et al.

In order to carry through the further argument we need the following lemma which is a version of the ergodic theorem.

LEMMA 2. *Suppose* (A1) *holds. If* $\xi=\{\xi(t, \omega);\ t\geqq 0\}$ *is a stationary predictable process with finite second-order moment. Then*

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \xi(t, \omega)dt = \mathrm{E}\{\xi(0, \omega)\} \tag{3.3}$$

with probability one, and

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \xi(t, \omega)\frac{dN(t)}{\lambda(t)} = \mathrm{E}\{\xi(0, \omega)\} \tag{3.4}$$

with probability one, where $\lambda(t)=\lambda_{\theta_0}(t, \omega)$.

PROOF. Since for each t, $\xi(t, \omega)$ is a measurable functional of the point process $\{N(s, t);\ s<t\}$, the stationary process ξ satisfies the ergodic theorem (3.3). The proof of (3.4) is not so simple. Consider

$$\begin{aligned}\eta(T, \omega) &= \int_0^T \xi(t, \omega)\frac{dN(t)}{\lambda(t)} - \int_0^T \xi(t, \omega)dt \\ &= \sum_{i=1}^{[T]} Y_i = \{\eta(T, \omega) - \eta([T], \omega)\}\end{aligned}$$

where $Y_i=\eta(i, \omega)-\eta(i-1, \omega)$, $i=1, 2, \cdots, [T]$. Then we see that $\mathrm{E}\{Y_i | Y_1, \cdots, Y_{i-1}\}=0$ by the same way as (3.1). This implies Kolmogorov's inequality (see [4], p. 235 for example). Since by the same way as (3.2) we see that $\mathrm{E}\{Y_i^2\}$ are finite and independent of i, we get with probability one that $(1/[T])\sum_{i=1}^{[T]} Y_i \to 0$ as $T\to\infty$. Thus $(1/T)\eta(T, \omega)\to 0$ with probability one, which, with (3.3), implies (3.4).

The next lemma treats Kullback-Leibler's information for stationary point processes.

LEMMA 3. *For the likelihood ratio on the unit interval* $[0, 1]$,

$$\Lambda_1(\theta_0; \theta) = \int_0^1 \{\lambda_\theta(t, \omega) - \lambda_{\theta_0}(t, \omega)\}dt + \int_0^1 \log\frac{\lambda_{\theta_0}(t, \omega)}{\lambda_\theta(t, \omega)}dN(t)\ , \qquad \theta\in\Theta\ ,$$

$\mathrm{E}\{\Lambda_1(\theta_0; \theta\}\geqq 0$ *always holds, and the equality holds if and only if* $\lambda_\theta(0, \omega) = \lambda_{\theta_0}(0, \omega)$ *a.s.*

PROOF. From (2.1) we have

$$\mathrm{E}\{\Lambda_1(\theta_0; \theta)\} = \mathrm{E}\left[\lambda_{\theta_0}(0, \omega)\left\{\frac{\lambda_\theta(0, \omega)}{\lambda_{\theta_0}(0, \omega)} - 1 + \log\frac{\lambda_{\theta_0}(0, \omega)}{\lambda_\theta(0, \omega)}\right\}\right]\ .$$

Thus the lemma is immediately obtained by the following elementary fact; for positive x, $\log x-1+(1/x)\geqq 0$ always holds and equality holds if and only if $x=1$.

Remark. It is easily seen by the preceding proof that a similar result is valid for the non-stationary case, i.e. for

$$\Lambda_T^*(\theta_0;\theta)=\int_0^T \{\lambda_\theta^*(t,\omega)-\lambda_{\theta_0}^*(t,\omega)\}dt+\int_0^T \log\frac{\lambda_{\theta_0}^*(t,\omega)}{\lambda_\theta(t,\omega)}dN(t)\ , \qquad \theta\in\Theta$$

$\mathrm{E}\{\Lambda_T^*(\theta_0;\theta)\}\geqq 0$, and the equality holds if and only if $\lambda_{\theta_0}^*(t,\omega)=\lambda_\theta^*(t,\omega)$ for a.s. (t,ω).

THEOREM 2. *Under the Assumptions* (A1), (B1)–(B3), (B5) *and* (C1), *the maximum likelihood estimators* $\hat\theta_T=\hat\theta(t_i;\ 0\leqq t_i\leqq T)$ *converge to* θ_0 *in probability as* $T\to\infty$.

PROOF. By (B2), and (B5) we have

$$\mathrm{E}\{\inf_{\theta'\in U}\lambda_{\theta'}(0,\omega)\}\to\mathrm{E}\{\lambda_\theta(0,\omega)\}\ ,$$

and

$$\mathrm{E}[\lambda_{\theta_0}(0,\omega)\log\{\lambda_{\theta_0}(0,\omega)/\sup_{\theta'\in U}\lambda_{\theta'}(0,\omega)\}]\to\mathrm{E}\left\{\lambda_{\theta_0}(0,\omega)\log\frac{\lambda_{\theta_0}(0,\omega)}{\lambda_\theta(0,\omega)}\right\}$$

as the neighbourhood U of θ shrinks to $\{\theta\}$. Let U_0 be an open neighbourhood of θ_0. Then by Lemma 3 and (B3) there is a positive ε such that $\mathrm{E}\{\Lambda_1(\theta_0;\theta)\}\geqq 3\varepsilon$ for any $\theta\in\Theta\backslash U_0$. Now for any $\theta\in\Theta\backslash U_0$, we can choose U small enough so that

$$\begin{aligned}&\mathrm{E}[\inf_{\theta'\in U}\lambda_{\theta'}(0,\omega)-\lambda_{\theta_0}(0,\omega)+\lambda_{\theta_0}(0,\omega)\log\{\lambda_{\theta_0}(0,\omega)/\sup_{\theta'\in U}\lambda_{\theta'}(0,\omega)\}]\\&\quad\geqq\mathrm{E}\{\Lambda_1(\theta_0;\theta)\}-\varepsilon\ .\end{aligned}$$

Select a finite number of θ_s such that $U_s=U_{\theta_s}$, $1\leqq s\leqq N$, cover $\Theta\backslash U_0$. Since $\inf_{\theta'\in U}\lambda_{\theta'}(t,\omega)$ and $\sup_{\theta'\in U}\lambda_{\theta'}(t,\omega)$ are predictable processes, by Lemma 2 there exists, for any $\varepsilon>0$, $T_0=T_0(\varepsilon)$ depending on the sample such that for any $T>T_0$ and $s=1,2,\cdots,N$,

$$\begin{aligned}(3.5)\qquad &\frac{1}{T}L_T(\theta_0)-\sup_{\theta\in U_s}\frac{1}{T}L_T(\theta)\\&\geqq\frac{1}{T}\int_0^T\{\inf_{\theta\in U_s}\lambda_\theta(t,\omega)-\lambda_{\theta_0}(t,\omega)\}dt+\frac{1}{T}\int_0^T\log\frac{\lambda_{\theta_0}(t,\omega)}{\sup_{\theta\in U_s}\lambda_\theta(t,\omega)}dN(t)\\&\geqq\mathrm{E}\{\Lambda_1(\theta_0;\theta)\}-2\varepsilon\geqq\varepsilon\ .\end{aligned}$$

It follows that there exists $T_1=T_1(\varepsilon,U_0)>T_0$ such that for all $T>T_1$

$$\sup_{\theta \in U_0} L_T(\theta) \geqq \sup_{\theta \in \Theta \setminus U_0} L_T(\theta) + \varepsilon T . \tag{3.6}$$

From (C1) we easily see that the inequality (3.5) and (3.6) remain valid for the case of $\lambda_\theta^*(t, \omega)$ and $L_T^*(\theta)$ with probability going to one as $T \to \infty$. But (3.6) means $\hat{\theta} \in U_0$. This completes the proof.

THEOREM 3. *Under Assumptions* A, (B2, 4, 6, 7) *and* (C2) *the Hessian matrix* $\{(1/T)(\partial^2 L_T^*(\theta)/\partial\theta_i \partial\theta_j)\}_{i,j=1,2,\cdots,d}$ *is asymptotically negative-definite in some neighbourhood* U *of* θ_0.

PROOF. Let U be some neighbourhood of θ_0. If $\theta \in U$, then by (B4), (B7) and the mean value theorem we get for $i, j = 1, 2, \cdots, d$

$$\frac{1}{T} \frac{\partial^2 L_T(\theta)}{\partial\theta_i \partial\theta_j} = \frac{1}{T} \int_0^T \frac{\partial^2 \lambda}{\partial\theta_i \partial\theta_j} \left\{ \frac{dN(t)}{\lambda} - dt \right\} \bigg|_{\theta_0} - \frac{1}{T} \int_0^T \frac{1}{\lambda} \frac{\partial\lambda}{\partial\theta_i} \frac{\partial\lambda}{\partial\theta_j} \bigg|_{\theta_0} dN(t)$$
$$- \frac{1}{T} \int_0^T \alpha |\theta - \theta_0| H(t, \omega) dt + \frac{1}{T} \int_0^T \beta |\theta - \theta_0| G(t, \omega) dN(t) ,$$

where $|\theta - \theta_0|$ denotes length in R^d, and α, β are random variables such that $|\alpha|, |\beta| < d$. From (B4, 6, 7) and Lemma 2 we have as $T \to \infty$

$$- \frac{1}{T} \int_0^T \frac{\partial^2 \lambda}{\partial\theta_i \partial\theta_j} \bigg|_{\theta_0} dt + \frac{1}{T} \int_0^T \frac{1}{\lambda} \frac{\partial^2 \lambda}{\partial\theta_i \partial\theta_j} \bigg|_{\theta_0} dN(t) \to 0 ,$$

$$\frac{1}{T} \int_0^T \frac{1}{\lambda^2} \frac{\partial\lambda}{\partial\theta_i} \frac{\partial\lambda}{\partial\theta_j} \bigg|_{\theta_0} dN(t) \to I_{ij}(\theta_0) ,$$

$$\frac{1}{T} \int_0^T H(t, \omega) dt \to \mathrm{E}\{H(0, \omega)\} ,$$

and

$$\frac{1}{T} \int_0^T G(t, \omega) dN(t) \to \mathrm{E}\{\lambda(0, \omega) G(0, \omega)\}$$

with probability one. Suppose $\varepsilon > 0$ is given. Choose $\delta = \delta(\varepsilon)$ in such a way that $\delta < \varepsilon$ and $\{\theta; |\theta - \theta_0| < \delta\} \subset U$. Having chosen δ, choose $T_0 = T_0(\varepsilon)$ large enough so that if $T \geqq T_0$, then with probability exceeding $1 - \varepsilon$

$$\left| - \frac{1}{T} \int_0^T \frac{\partial^2 \lambda}{\partial\theta_i \partial\theta_j} \bigg|_{\theta_0} dt + \frac{1}{T} \int_0^T \frac{1}{\lambda} \frac{\partial^2 \lambda}{\partial\theta_i \partial\theta_j} \bigg|_{\theta_0} dN(t) \right| < \delta ,$$
$$\left| I_{ij}(\theta_0) + \frac{1}{T} \int_0^T \frac{-1}{\lambda^2} \frac{\partial\lambda}{\partial\theta_i} \frac{\partial\lambda}{\partial\theta_j} \bigg|_{\theta_0} dN(t) \right| < \delta \tag{3.7}$$

and

$$\left| \frac{1}{T} \int_0^T -\alpha H(t, \omega) dt + \frac{1}{T} \int_0^T \beta G(t, \omega) dN(t) \right| < 2d^3 M^3 \tag{3.8}$$

for $i, j=1,\cdots,d$. Also choose δ so small that $\{\theta;\ |\theta-\theta_0|<\delta\}\subset U$ and that if (σ_{ij}) is any $d\times d$-symmetric matrix with $|\sigma_{ij}-I_{ij}|<2\delta(1+d^3M^3)$ for $i, j=1, 2,\cdots,d$ then (σ_{ij}) is positive definite. Thus from (3.7) and (3.8), for any θ such that $|\theta-\theta_0|<\delta$

$$\left|\frac{1}{T}\frac{\partial^2 L_T(\theta)}{\partial\theta_i\partial\theta_j}+I_{ij}(\theta_0)\right|<\delta+\delta+2d^3M^3\delta$$

holds with probability going to one as $T\to\infty$. Therefore from (C2) with probability going to one, the matrix $\{\partial^2 L_T^*(\theta)/\partial\theta_i\partial\theta_j\}$ is negative-definite for every θ such that $|\theta-\theta_0|<\delta$.

Example. In the following cases the Hessians are always non-positive definite a.s. ω;

$$\lambda_\theta^*(t,\omega)=\sum_{i=1}^d \theta_i\xi_i^*(t,\omega)+\eta^*(t,\omega)\ ,\qquad \xi_i^*,\ \eta>0 \text{ a.s. } \omega\ ,$$

$$\lambda_\theta^*(t,\omega)=\exp\left\{\sum_{i=1}^d \theta_i\xi_i^*(t,\omega)+\eta^*(t,\omega)\right\}\ ,$$

where ξ, η are some predictable processes with respect to the corresponding point processes respectively. In fact, we have for any real u_i, $i=1, 2,\cdots,d$

$$\sum_{i,j=1}^d u_iu_j\frac{\partial^2 L_T^*(\theta)}{\partial\theta_i\partial\theta_j}=-\int_0^T\left\{\sum_{i=1}^d u_i\xi_i^*(t,\omega)/\lambda_\theta^*(t,\omega)\right\}^2 dN(t)$$

$$=-\int_0^T\left\{\sum_{i=1}^d u_i\xi_i^*(t,\omega)\right\}^2\lambda_\theta^*(t,\omega)dt\ ,$$

respectively.

THEOREM 4. *Under Assumptions* A, (B2, 4, 6) *and* (C3) $(1/\sqrt{T})\cdot(\partial L_T^*(\theta_0)/\partial\theta)$ *converges in law to* $N(0, I(\theta_0))$ *as* $T\to\infty$.

PROOF. Since for $0\leqq S\leqq T$ and $i=1, 2,\cdots,d$

$$\mathrm{E}\left\{\frac{\partial L_T(\theta_0)}{\partial\theta_i}\,\middle|\,H_S\right\}=\frac{\partial L_S(\theta_0)}{\partial\theta_i}+\mathrm{E}\,\{\Delta(S,T)\,|\,H_S\}$$

where $H_S=H_{-\infty,S}$, and using (2.1) and the definition of conditional expectation

$$\mathrm{E}\,\{\Delta(S,T)\,|\,H_S\}=\mathrm{E}\left[\int_S^T\frac{\partial\lambda}{\partial\theta}\left\{\frac{dN(t)}{\lambda}-dt\right\}\bigg|_{\theta_0}\,\bigg|\,H_S\right]=0\ ,$$

we see $\partial L_T(\theta_0)/\partial\theta$ is a martingale. If we put

$$\frac{\partial L_T(\theta_0)}{\partial\theta}=\sum_{k=1}^{[T]}\Delta(k-1,k)+\Delta([T],T)\ ,$$

then $\{\Delta(k-1, k)\}_{k=1,2,\dots}$ is a sequence of stationary ergodic martingale differences with $\mathrm{E}\{\Delta(0,1)\Delta(0,1)'\}=I(\theta_0)$ by Lemma 1. Thus $[T]^{-1/2}\sum_{k=1}^{[T]}\Delta(k-1, k)$ converges in law to $N(0, I(\theta_0))$ as $T\to\infty$ by the central limit theorem for martingale differences (see [2]). On the other hand we see that $T^{-1/2}\Delta([T], T)\to 0$ in probability as $T\to\infty$. Assumption (C3) completes the proof.

Remark. If (A2) or (A3) do not hold, the covariance matrix is different from $I(\theta_0)$ as defined in (B6). For example $\mathrm{E}\{\Delta(0,1)\Delta(0,1)'\}=\mathrm{E}[(\mu/\lambda^2)(\partial\lambda/\partial\theta_i)(\partial\lambda/\partial\theta_j)]$ if $\mathrm{E}[dN(t)^2|H_{-\infty,t}]=\mu(t,\omega)dt$.

THEOREM 5. *Suppose the maximum likelihood estimator $\hat{\theta}_T$ satisfies the equation $\partial L_T^*(\theta)/\partial\theta=0$. Then under the Assumptions* A, (B2, 4, 6, 7) *and* (C2, 3, 4), *as* $T\to\infty$,

$$\sqrt{T}(\hat{\theta}_T-\theta_0)\to N(0, I(\theta_0)^{-1}) \tag{3.9}$$

and

$$2\{L_T^*(\hat{\theta}_T)-L_T^*(\theta_0)\}\to\chi_d^2 \tag{3.10}$$

in law.

PROOF. From (C2) and (C4) we have the following with probability going to one as $T\to\infty$,

$$0=\frac{1}{\sqrt{T}}\frac{\partial L_T^*(\theta_0)}{\partial\theta}+\frac{1}{T}\frac{\partial^2 L_T^*(\theta_0)}{\partial\theta\partial\theta'}\sqrt{T}(\hat{\theta}-\theta_0)$$
$$+\sqrt{T}(\hat{\theta}_T-\theta_0)'\left\{-\frac{\alpha}{T}\int_0^T H(t,\omega)dt+\frac{\beta}{T}\int_0^T G(t,\omega)dN(t)\right\}(\hat{\theta}_T-\theta_0)$$

where H, G are given in (B7), and α, β are random matrices with $|\alpha_{ij}|$, $|\beta_{ij}|\leqq d^2/2$. Since $|\hat{\theta}_T-\theta_0|\to 0$ as $T\to\infty$, we get from the last part in the proof of Theorem 3

$$\left|\frac{1}{\sqrt{T}}\frac{\partial L_T^*(\theta_0)}{\partial\theta}-\sqrt{T}I(\theta_0)(\hat{\theta}_T-\theta_0)\right|\leqq\varepsilon_T|\sqrt{T}(\hat{\theta}_T-\theta_0)| \tag{3.11}$$

for some ε_T such that $\varepsilon_T\to 0$ in probability as $T\to\infty$. Hence we have (3.9) by Theorem 3 and by Theorem 10.1 of [2]. Now we have from (C2) and (C4) with probability going to one as $T\to\infty$ that

$$2\{L_T^*(\hat{\theta})-L_T^*(\theta_0)\}=2\frac{\partial L_T^*(\theta_0)}{\partial\theta'}(\hat{\theta}_T-\theta_0)+(\hat{\theta}_T-\theta_0)'\frac{\partial^2 L_T^*(\theta_0)}{\partial\theta\partial\theta'}(\hat{\theta}_T-\theta_0)$$
$$+|\hat{\theta}_T-\theta_0|^3\left\{\int_0^T\alpha H(t,\omega)dt+\int_0^T\beta G(t,\omega)dN(t)\right\}$$

for $\hat{\theta}_T \in U$, where α, β are some random variable such that $|\alpha|, |\beta| \leqq d^3/3$. Since the last term tends to zero in probability, we get (3.10) by Theorem 3 and Theorem 4 and (3.11).

We now state and prove the Crámer-Rao inequality for finding a lower bound of the variance of the estimate. Let $0 \leqq t_0(\omega) \leqq t_1(\omega) < \cdots < t_{n-1}(\omega) \leqq T$ be the observation on the interval $[0, T]$ from the point process $P_\theta^*(\cdot)$ which has the Radon-Nykodim derivative $\rho_T^*(\omega, \theta) = \exp \left\{ \int_0^T \log \lambda_\theta^*(t, \omega) dN(t) + \int_0^T (1 - \lambda_\theta^*(t, \omega)) dt \right\}$ with respect to the standard Poisson process, (see Theorem 13 in [6] for example). An estimate $\delta_T(\omega) = \delta_T(t_0, \cdots, t_{n-1})$, not necessarily unbiased, is wanted for the vector parameter $\theta \in \Theta \subset R^d$. Consider the following additional conditions.

CONDITIONS D.

(D1) $\mathrm{E}_\theta \{\delta_T(\omega)^2\} < \infty$ for all $\theta \in \Theta$.

(D2) $\mathrm{E}_\theta \left\{ \int_0^T \frac{1}{\lambda_\theta^*(t, \omega)} \frac{\partial \lambda_\theta^*(t, \omega)}{\partial \theta_i} \frac{\partial \lambda_\theta^*(t, \omega)}{\partial \theta_j} dt \right\} < \infty$ for all $\theta \in \Theta$ and $i, j = 1, \cdots, d$.

(D3) $\frac{\partial}{\partial \theta} \int_0^T \lambda_\theta^*(t, \omega) dt = \int_0^T \frac{\partial \lambda_\theta^*(t, \omega)}{\partial \theta} dt$ and

$\frac{\partial}{\partial \theta} \int_0^T \log \lambda_\theta^*(t, \omega) dN(t) = \int_0^T \frac{\partial}{\partial \theta} \log \lambda_\theta^*(t, \omega) dN(t)$ for all $\theta \in \Theta$.

(D4) $\frac{\partial}{\partial \theta} \int \delta_T(\omega) \rho_T^*(\omega, \theta) \Pi(d\omega) = \int \delta_T(\omega) \frac{\partial}{\partial \theta} \rho_T^*(\omega, \theta) \Pi(d\omega)$ for all $\theta \in \Theta$, where Π is the probability measure of the standard Poisson process.

THEOREM 6. *If* $\mathrm{E}_\theta \{\delta_T(\omega)\} = \theta + b_T(\theta)$, *and* $b_T(\theta)$ *is differentiable, then under the regularity Conditions* D

$$\mathrm{E}_\theta [\{\delta_T(\omega) - \theta - b_T(\theta)\}' \{\delta_T(\omega) - \theta - b_T(\theta)\}] \\ \geqq \left\{ I + \frac{\partial}{\partial \theta} b_T(\theta) \right\} I_T^*(\theta)^{-1} \left\{ I + \frac{\partial}{\partial \theta} b_T(\theta) \right\}',$$

where I *is* $d \times d$*-identity matrix,* $A \geqq B$ *means that the matrix* $A - B$ *is non-negative definite, and*

$$I_T^*(\theta) = \mathrm{E}_\theta \left\{ \int_0^T \frac{1}{\lambda_\theta^*(t, \omega)} \frac{\partial \lambda_\theta^*(t, \omega)}{\partial \theta} \frac{\partial \lambda_\theta^*(t, \omega)}{\partial \theta'} dt \right\}.$$

PROOF. By (1.2)

$$\mathrm{E}_\theta \{\delta_T(\omega)\} = \int \delta_T(\omega) P_\theta^*(d\omega) = \int \delta_T(\omega) \rho_T^*(\omega, \theta) \Pi(d\omega).$$

Thus by (D4) and (1.2)

$$\frac{\partial}{\partial\theta}\mathrm{E}_\theta\{\delta_T(\omega)\}=\int\delta_T(\omega)\frac{\partial}{\delta\theta}\rho_T^*(\omega,\theta)\Pi(d\omega)$$
$$=\int\delta_T(\omega)\left\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\right\}\rho_T^*(\omega,\theta)\Pi(d\omega)$$
$$=\mathrm{E}_\theta\left\{\delta_T(\omega)\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\right\}.$$

By (2.1) and (D3)

$$\mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\right\}=\mathrm{E}_\theta\left\{\int_0^T\frac{1}{\lambda_\theta^*(t,\omega)}\frac{\lambda_\theta^*(t,\omega)}{\partial\theta}dN(t)-\int_0^T\frac{\partial\lambda^*(t,\omega)}{\partial\theta}dt\right\}=0.$$

Thus by Schwartz's inequality we have for any vectors $s,t\in R^d$

$$\left[t'\left\{I+\frac{\partial}{\partial\theta}b_T(\theta)\right\}s\right]^2$$
$$=\left[t'\,\mathrm{Cov}_\theta\left\{\delta_T(\omega)\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\right\}s\right]^2$$
$$\leqq\mathrm{E}_\theta\,[t'\{\delta_T(\omega)-\theta-b_T(\theta)\}\{\delta_T(\omega)-\theta-b_T(\theta)\}'t]$$
$$\cdot\,\mathrm{E}_\theta\left\{s'\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)s\right\}.$$

Put

$$s=\mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)\right\}^{-1}\left\{I+\frac{\partial}{\partial\theta'}b_T(\theta)\right\}'t.$$

Then we have for any vector $t\in R^d$,

$$t'\left\{I+\frac{\partial}{\partial\theta}b_T(\theta)\right\}\mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)\right\}^{-1}\left\{I+\frac{\partial}{\partial\theta}b_T(\theta)\right\}'t$$
$$\leqq t'\,\mathrm{E}_\theta\,[\{\delta_T(\omega)-\theta-b_T(\theta)\}\{\delta_T(\omega)-\theta-b_T(\theta)\}']t.$$

Note that by the similar method to the proof of (3.2), we have

$$\mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)\right\}$$
$$=\mathrm{E}_\theta\left\{\int_0^T\frac{1}{\lambda_\theta^*(t,\omega)}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta'}dt\right\}.$$

This completes the proof.

By Assumption (C2) it is easily seen that

$$\frac{1}{T}\mathrm{E}_\theta\left\{\int_0^T\frac{1}{\lambda_\theta^*(t,\omega)}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta'}dt\right\}\to I(\theta).$$

Therefore together with Theorem 4 we have the following.

THEOREM 7. *If the maximum likelihood estimator* $\hat{\theta}_T$ *satisfies the conditions of Theorem* 6, *then it is an asymptotically efficient estimator, that is,* $\hat{\theta}_T$ *asymptotically attains the lower bound of the variance of estimates.*

4. Characterization of Poisson processes by a maximum likelihood estimator

Suppose a stationary delayed renewal process has a survivor function $1-F_\theta(t)$, $0\leqq t\leqq\infty$, where $F_\theta(t)$ is a probability distribution function with density function $f_\theta(t)$ such that

$$\int_0^\infty t dF_\theta(t)=\int_0^\infty t f_\theta(t)dt=\mu(\theta) . \tag{4.1}$$

Remember the definition of the maximum likelihood estimator given in Section 1, that is, $\hat{\theta}_T$ maximizes the likelihood (1.2) under the observation from the stationary delayed renewal process. Assume the following conditions.

CONDITIONS E.

(E1) For any $t>0$ the maximum likelihood estimator $\hat{\theta}=\{\hat{\theta}_T(\omega)\}$ is a measurable function from $(t,\infty)\times\Omega$ onto Θ, that is, for any $\theta\in\Theta$ and for any $t>0$ there are $T(\theta)>t$ and $\omega(\theta)\in\Omega$ such that $\theta=\theta_{T(\theta)}(\omega(\theta))$ maximises the log-likelihood (2.3).

(E2) $\log(1-F_\theta(s))$ and $\log f_\theta(s)$ are differentiable in θ.

Then we have the following.

THEOREM 8. *The maximum likelihood estimator depends only on* T *and* $n=N(0,T)$ *if and only if* $F_\theta(t)=1-e^{-\mu(\theta)t}$ *where* $\mu(\theta)>0$ *for all* $\theta\in\Theta$.

PROOF. See Example 1 for the proof of "if" part. Now consider the "only if" part. Suppose θ^* is given. Then by (E1) there are $T(\theta^*)$, $n(\theta^*)=N(0,T(\theta^*))$ and $t_0, t_1,\cdots,t_{n(\theta^*)-1}$ such that θ^* maximizes (2.3). Now fix $n^*=n(\theta^*)$ and $T^*=T(\theta^*)$. Then

$$\begin{aligned} 0&=\frac{\partial}{\partial\theta}L^*_{T^*}(\theta^*) \\ &=\frac{\partial}{\partial\theta}\log\mu^{-1}_{(\theta^*)}(1-F_{\theta^*}(t_0))+\sum_{i=1}^{n^*-1}\frac{\partial}{\partial\theta}\log f_{\theta^*}(t_i-t_{i-1}) \\ &\quad+\frac{\partial}{\partial\theta}\log(1-F_{\theta^*}(T^*-t_{n^*-1})) . \end{aligned} \tag{4.2}$$

Put

$$G_\theta(s)=\frac{\partial}{\partial\theta}\log f_\theta(s)\ ,\qquad H_\theta(s)=\frac{\partial}{\partial\theta}\log \mu_{(\theta)}^{-1}(1-F_\theta(s))\ ,$$

(4.3)

$$K_\theta(s)=\frac{\partial}{\partial\theta}\log(1-F_\theta(T^*-s))\qquad \text{and}\qquad s_i=t_i-t_{i-1}\ ,$$

$$i=1,2,\cdots,n^*-1,\ s_0=t_0\ .$$

Then $G_\theta(\cdot)$, $H_\theta(\cdot)$ and $K_\theta(\cdot)$ are measurable functions such that

$$H_{\theta^*}(s_0)+\sum_{i=1}^{n^*-1}G_{\theta^*}(s_i)=K_{\theta^*}\left(\sum_{i=1}^{n^*-1}s_i\right)\ . \tag{4.4}$$

By (E1) and the assumption of the theorem s_i, $i=0,1,\cdots,n^*-1$, are arbitrary positive numbers independent of θ^* and T^*. Therefore the equation (4.3) is a kind of Pexider's functional equation (see [11], Chapter 3 for example), and we have the general solution for $s>0$

$$G_{\theta^*}(s)=a(\theta^*)s+b(\theta^*)\ ,\qquad H_{\theta^*}(s)=a(\theta^*)s+c(\theta^*)\ ,$$

$$K_{\theta^*}(\theta)=a(\theta^*)s+(n^*-1)b(\theta^*)+c(\theta^*)\ .$$

Since θ^* is arbitrary and $a(\theta)$, $b(\theta)$ are indefinitely integrable by (E2), solving the first differential equation (4.3) above we have

$$f_\theta(s)=B(\theta)e^{A(\theta)s}\ .$$

Since $f_\theta(t)$ is a probability density function satisfying (4.1) we get

$$f_\theta(t)=\mu(\theta)e^{-\mu(\theta)t}\ .$$

Therefore

$$F_\theta(t)=1-e^{-\mu(\theta)t}\ .$$

This completes the proof.

COROLLARY. *If $F_\theta(t)=F(\theta t)$, $\theta>0$, then $N(0,T)/T$ is a maximum likelihood estimator of θ if and only if $F(t)=1-e^{-t}$.*

Remark. It is easily seen by the Palm-Khinchine theory that $N(0, T)/T$ is an unbiased estimator of θ, that is, θ is equal to the intensity of the stationary delayed renewal process.

Remark. It is easily seen also that the estimator $N(0,T)/T$ is not always asymptotically efficient estimator of the intensity θ of the renewal process. In fact we see that

$$\lim_{T\to\infty} T\,\mathrm{E}_{\theta,\sigma}[\{N(0,T)/T-\theta\}^2]=\sigma^2\theta^3\ ,$$

where σ^2 is the variance of the failure-time distribution.

Acknowledgement

The author is grateful to Professor D. Vere-Jones and Mr. T. Ozaki for stimulating discussions of the subject, and also to Professor D. Vere-Jones for his help and careful reading of earlier drafts of this paper.

This research was partly supported by a grant from the Ministry of Education, Science and Culture.

Finally, the author thanks to Victoria University of Wellington for the visiting lectureship which enables him to complete this paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS AND VICTORIA UNIVERSITY OF WELLINGTON

REFERENCES

[1] Billingsley, P. (1961). *Statistical Inference for Markov Processes*, The University of Chicago Press, Chicago.

[2] Billingsley, P. (1961). The Lindeberg-Lévy Theorem for martingales, *Proc. Amer. Math. Soc.*, **12**, 788-792.

[3] Dellacherie, C. (1972). *Capacities et Processus Stochastiques*, Springer-Verlag, Heidelberg.

[4] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. 2, John Wiley & Sons, New York.

[5] Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exiting point process, *J. Appl. Prob.*, **11**, 493-503.

[6] Kavanov, Yu. M., Lipster, R. Sh. and Shiryaev, A. N. (1975). Martingale method in the theory of point processes (in Russian), *Proceeding of Vilnius Symposium U.S.S.R.*

[7] Meyer, P. A. (1972). *Martingales and Stochastic Integrals* I, Lecture Notes in Mathematics 284, Springer, Berlin.

[8] Ozaki, T. (1977). Maximum likelihood estimation of Hawkes' self-exciting point process, *Research Memorandom*, No. 115, The Institute of Statistical Mathematics, Tokyo.

[9] Vere-Jones, D. (1973). *Lectures on Point Processes*, Department of Statistics, University of California, Berkeley.

[10] Vere-Jones, D. (1975). On updating algorithms and inference for stochastic point processes, *Perspectives in probability and statistics*, Gani, J. ed., Applied Probability Trust.

[11] Aczel, J. (1966). *Lectures on Functional Equations and their Applications*, Academic Press, New York.

[12] Daley, D. J. and Vere-Jones, D. (1972). A summary of the theory of point processes, *Stochastic point processes: statistical analysis, theory and applications*, Lewis, A. W. ed., Wiley, New York.

[13] Huber, P. J. (1967). The behaviour of maximum likelihood estimates under nonstandard conditions, *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, **1**, 221-233.

[14] Meyer, P. A. (1966). *Probability and Potentials*, Blaisdell Publishing Co., Waltham.