CONTRIBUTIONS TO THE STATISTICAL ANALYSIS OF SERIES OF EVENTS BY THE LIKELIHOOD PROCEDURE

YOSIHIKO OGATA
THE INSTITUTE OF STATISTICAL MATHEMATICS
TOKYO

1980

CONTENTS

PREFACE

The book by Cox and Lewis (1966) and the volume editated by Lewis (1972) showed the development of statistical techniques for point processes. However the possibility of the maximum likelihood estimation procedure was not systematically developed.

In the last decade the likelihood analysis for point processes was developed to solve some problems in communication engeneering (see the recent book by Snyder (1975) for the extensive discussions of the field). A key to the likelihood theory is the conditional intensity function

$$\lambda(t|F_t) = E[N(dt)|F_t]/dt,$$

where F_t is the history of the process over the interval of observation $(0,t)$. Once the conditional intensity is known, the likelihood for the realization in $(0,T)$ can be written down in the form

$$[\prod_{i=1}^{N} \lambda(t_i|F_{t_i})]\exp[-\int_0^T \lambda(t|F_t)dt]$$

$$= \exp\{\int_0^T \log \lambda(t|F_t)dN_t - \int_0^T \lambda(t|F_t)dt\}.$$

The importance of the systematic approach to the likelihood procedure for statistical inference of general point processes was suggested by Vere-Jones (1975). It is now very important to obtain good parametric models of conditional intensity function $\lambda_\theta(t|F_t)$.

The present paper is concerned with the development of the maximum likelihood inference for point processes and consists of two chapters. Chapter 1 is devoted to the study of the asymptotic of the maximum likelihood procedure under the parameterized conditional intensities, and Chapter 2 to the application of an efficient simulation method to general point processes using their corresponding conditional intensities. Summaries of each chapter are briefly given at their biggining.

In 1976 Dr. H. Akaike organized the research project on statistical inference for point processes at the Institute of Statistical Mathematics, and invited Prof. D. Vere-Jones to join the project. I would wish to express my deep thanks to both of them for drawing my interest to the subject of the present paper. Their encouragements of my study and many valuable suggestions are greatly acknowledged. Thanks are also due to Miss N. Takenaka who typed this paper with great care and diligence. Finally I am most grateful to to Professors M. Okamoto and N. Inagaki of the Osaka University for their interests in my work, especially to Prof. Okamoto for his careful review of this paper and many critical comments.

REFERENCES

Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events, Methuen, London.

Lewis, P. A. W. (ed.) (1972). Stochastic Point Processes: Statistical Analysis, Theory and Applications, Wiley, New York.

Snyder, D. L. (1975). Random Point Processes, Wiley, New York.

Vere-Jones, D. (1975). "On updating algorithms and inference for stochastic point processes", Perspectives in Probability and Statistics, Gani, J. ed., Applied Probability trust, Sheffield.

Chapter 1

THE ASYMPTOTIC BEHAVIOR OF MAXIMUM LIKELIHOOD ESTIMATORS FOR STATIONARY POINT PROCESSES

(Published in Annals of the Institute of Statistical Mathematics, Vol.30, A, pp.243-261, 1978)

ABSTRACT

It is known that the likelihood of a point process which is absolutely continuous with respect to the standard Poisson process on a finite time interval can be written down in terms of the conditional intensity or generalized hazard function, which plays a similar role to the time-varying intensity of a non-homogeneous Poisson process.

The main object of the paper is to give regularity conditions for the parametrization of the complete intensity functions under which the standard statistical properties for the maximum likelihood procedure are satisfied. Furthermore some natural examples of the parametrized stationary point processes are given to check the regularity conditions for the practical uses.

I. INTRODUCTION

Let $P(\cdot)$ on the set of points $\omega = \{t_j;\ j=0,\pm1,\pm2,\cdots\}\varepsilon\Omega$ be an orderly stationary point process 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 points. 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 on the σ-algebra $H_{0,t}$ of the point process is defined respectively as follows:

$$(1.1)\qquad \lambda(t,\omega) = \lim_{\delta\to0}\frac{1}{\delta}P[N\{[t,t+\delta)\} > 0 \mid H_{-\infty,t}]$$

$$\lambda^*(t,\omega) = \lim_{\delta\to0}\frac{1}{\delta}P[N\{[t,t+\delta)\} > 0 \mid H_{0,t}] = E\{\lambda(t,\omega) \mid H_{0,t}\}$$

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 $\{\lambda_\theta(t,\omega);\ \theta\varepsilon\Theta\subset R^d\}$ which are assumed to correspond uniquely with the stationary point processes $\{P_\theta;\ \theta\varepsilon\Theta\}$. 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 theoretically consider 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 II. In section II assumptions are collected together, and examples which satisfy them are given. In section III it will be proved that the maximum likelihood estimator is consistent, asymptotically normal, and efficient. In the last section Poisson process will be characterized by a maximum likelihood estimator of parametrized renewal processes.

II. 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, $\partial \log \lambda/\partial\theta_i|_{\theta'}$ will be denoted the value of $\partial \log \lambda/\partial\theta_i$ at the point $\theta'\varepsilon\Theta$ with the same convention used for other functions. Instead of $E_{\theta_0}(\cdot)$, $P_{\theta_0}(\cdot)$, where θ_0 is the true value of the parameter, let us agree to write $E(\cdot)$ and $P(\cdot)$.

<u>Assumptions A.</u>

(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; $\lim_{\delta\to 0} \frac{1}{\delta} P[N\{[0,\delta)\} \geqq 2] = 0$.

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

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 $t \geq 0$ $\xi(t,\omega)$ is $H_{-\infty,t}$-measurable. Further, we say the process ξ is predictable if the mapping ξ: $R_+ \times \Omega \to R$ 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 path $\xi(t,\omega)$ is left continuous on $(0,\infty)$ for a.s.ω.

In section I we have already used the stochastic Stieltjes integrals $\int_0^T \xi(t,\omega)dN(t) = \sum_{0 \leq t_i \leq T} \xi(t_i(\omega),\omega)$ 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 $\int_0^T E\{\lambda_{\theta_0}(t,\omega)|\ \xi(t,\omega)|\}dt < \infty$ we are allowed to do the following calculation.

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

Thus for integrands finite predictable process we can use the formal relation $E\{dN(t)|\ H_{-\infty,t}\} = \lambda_{\theta_0}(t,\omega)dt$. Similarly, $E\{dN(t)|\ H_{0,t}\} = \lambda^*_{\theta_0}(t,\omega)dt$. The proof of (2.1) is directly derived from Theorem of [7], p.23.

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

<u>Assumptions B.</u>

(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 $\theta\varepsilon\Theta$.

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

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

$\theta'\varepsilon U$,

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

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

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

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

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

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

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

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 be always chosen. In (B2) we assume that the same is true for the complete intensity functions 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 $\theta\varepsilon U$ 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 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.

<u>Assumptions C.</u>

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

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

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

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

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

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

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

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

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

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

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

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

(ii) $\sup_{\theta'\varepsilon U} \frac{\partial^3\lambda^*}{\partial\theta_i\partial\theta_j\partial\theta_k}$ and $\sup_{\theta'\varepsilon U} \frac{\partial^3 \log\lambda^*}{\partial\theta_i\partial\theta_j\partial\theta_k}$ 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 $1 - F_\theta(t)$, $t \geq 0$, are given. In this case the complete intensity function coincide 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 as $t^*(\omega) \leq t$). The stationary joint distribution of forward and backward recurrence time is given by

$P(X \leq u, Y \leq v) = \mu_\theta^{-1} \int_0^u \{F_\theta(v+w) - F(w)\}dw$, where $\mu_\theta = \int_0^\infty t dF_\theta(t)$ (see [12]).

Since $P(X \leq v/Y > t) = \int_t^\infty F(v+w) - F(w)\}dw/\int_t^\infty \{1 - F(w)\}dw$, we see that

$\lambda^*(t,\omega) = E\{\lambda(t,\omega)/H_{0,t}\} = \{1 - F(t)\}/\int_t^\infty \{1 - F(w)\}dw$ 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 $0 \leq t_0 < \cdots < t_{n-1} \leq T$ 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,

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

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

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

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

where $\tau = t - t_{-1}$ and $\sigma = t_{-1} - t_{-2}$. It should be noted that the complete intensity function coincide 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

$$(2.4) \qquad \lambda_\theta(t,\omega) = \nu + \int_{-\infty}^{t}\gamma_\mu(t-u)dN(u), \quad \theta = (\nu,\mu)$$

$$= \nu + \sum_{t_i<t}\gamma_\mu(t - t_i)$$

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 integer 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-exciting 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 $\theta = (v,\alpha,\beta)$ such that $v>0$, $\alpha<\beta$, and obtained successfully maximum likelihood estimates from the simulation data. It is easily seen that Assumption C is always satisfied by the simple case above.

In general, we see for (C1) - (i) that $E\{\sup_{\theta'\varepsilon U} |\lambda^*_{\theta'}(t,\omega) - \lambda_{\theta'}(t,\omega)|\} \leqq 2E\{\lambda_{\theta_0}\}\times$ $\int_t^\infty \sup_{(v,\mu)\varepsilon U} \gamma_\mu(u)du$ with the rate of decrease of the left hand side depending on the rate of decrease of $\sup_{(v,\mu)\varepsilon U} \gamma_\mu(u)$. 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, $\frac{\partial\gamma}{\partial\theta_i}$, $\frac{\partial^2\gamma}{\partial\theta_i\partial\theta_j}$ and $\frac{\partial^3\gamma}{\partial\theta_i\partial\theta_j\partial\theta_k}$, and are certainly satisfied when γ_μ has exponential form. Finally, though $\lambda^*_\theta(t,\omega) = E\{\lambda_\theta(t,\omega)| H_{0,t}\}$ gives the best approximation of $\lambda_\theta(t,\omega)$, it is difficult to get the exact likelihood numerically. So, practically, we can use

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

This is predictable and satisfies the assumptions similarly.

III. ASYMPTOTIC PROPERTIES OF THE LIKELIHOOD PROCEDURE

Lemma 1. Under the assumptions A we have

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

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

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

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

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

$$\leqq \frac{1}{\delta} E\{N(t,t+\delta)^2 |\ H_{-\infty,t}\}.$$

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

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

Therefore with probability one

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

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

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

since with probability one each interval $[\frac{k}{n}, \frac{k+1}{n}]$ will ultimately have either zero or one event in it. Also we have

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

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

$$nE\{N(0, \frac{1}{n})^2\} \to E\{N(0,1)\} \quad \text{as } n \to \infty,$$

that is, for any $t \geqq 0$

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

Thus we have

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

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

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

Theorem 1. Under the assumptions A, (B2) and (B4)-(B6)

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

and

$$(3.2) \qquad E\{\frac{\partial L_T(\theta)}{\partial \theta_i} \frac{\partial L_T(\theta)}{\partial \theta_j}\}_{\theta=\theta_0} = -E\{\frac{\partial^2 L_T(\theta)}{\partial \theta_i \partial \theta_j}\}_{\theta=\theta_0}$$

$$= T\, E\{\frac{1}{\lambda} \frac{\partial \lambda}{\partial \theta_i} \frac{\partial \lambda}{\partial \theta_j}\}_{\theta=\theta_0}, \; i,j=1,2,\cdots,d.$$

Proof. By (ii) of Lemma 1

$$\lim_{\delta\to 0} \frac{1}{\delta} E\{N[t,t+\delta) | H_{-\infty,t}\} = \lim_{\delta\to 0} \frac{1}{\delta} P\{N[t,t+\delta) = 1 | H_{-\infty,t}\}$$

$$= \lambda_{\theta_0}(t,\omega).$$

This means formally that $E\{dN(t) | H_{-\infty,t}\} = \lambda_{\theta_0}(t,\omega)dt$. Thus we have by (B2), (B4), (B5) and (2.1) that

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

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

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

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

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

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

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

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

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

$$= E[\int\int_{\{0\leq s<t\leq T\}} + \int\int_{\{0\leq t<s\leq T\}} + \int\int_{\{0\leq s=t\leq T\}}]_{\theta=\theta_0}$$

$$= I_1 + I_2 + I_3.$$

From the relation (2.1) and

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

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

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

Thus we see from (2.1) that

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

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 $\xi = \{\xi(t,\omega);\ t\geq 0\}$ is a stationary predictable process with finite second-order moment. Then

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

with probability one, and

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

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

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

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

$\frac{1}{[T]}\sum_{i=1}^{[T]} Y_i \to 0$ as $T \to \infty$. Thus $\frac{1}{T}\,\eta(T,\omega) \to 0$ with probability one. This and (3.3) implies (3.4).

The next lemma treat the Kullback-Leibler's information of the stationary point processes.

Lemma 3. For the likelihood ratio on the unit interval [0,1],

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

$E\{\Lambda_1(\theta_0;\theta)\} \geqq 0$ 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

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

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

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

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

$E\{\Lambda_T^*(\theta_0;\theta)\} \geqq 0$, 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 estimator $\hat{\theta}_T = \hat{\theta}(t_i;\ 0\leqq t_i\leqq T)$ converge to θ_0 in probability as $T \to \infty$.

Proof. By (B2) and (B5) we have

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

and

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

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 $E\{\Lambda_1(\theta_0;\theta)\} \geqq 3\varepsilon$ for any $\theta\in\Theta\setminus U_0$. Now for any $\theta\in\Theta\setminus U_0$, we can choose U small enough so that

$$E[\inf_{\theta'\in U}\lambda_{\theta'}(0,\omega) - \lambda_{\theta_0}(0,\omega) + \lambda_{\theta_0}(0,\omega)\log\{\lambda_{\theta_0}(0,\omega)/\sup_{\theta'\in U}\lambda_{\theta'}(0,\omega)\}]$$

$$\geqq E\{\Lambda_1(\theta_0;\theta)\} - \varepsilon.$$

Select a finite number of θ_s such that $U_s = U_{\theta_s}$, $1 \leqq s \leqq N$, cover $\Theta\setminus U_0$. Since

$\inf_{\theta' \varepsilon U} \lambda_{\theta'}(t,\omega)$ and $\sup_{\theta' \varepsilon U} \lambda_{\theta'}(t,\omega)$ are predictable processes, by Lemma 2 there exists, for any $\varepsilon > 0$, $T_0 = T_0(\varepsilon)$ depending on sample such that for any $T > T_0$ and $s = 1,2,\cdots,N$,

$$(3.5) \qquad \frac{1}{T} L_T(\theta_0) - \sup_{\theta \varepsilon U_s} \frac{1}{T} L_T(\theta)$$

$$\geqq \frac{1}{T}\int_0^T \{\inf_{\theta\varepsilon U_s} \lambda_\theta(t,\omega) - \lambda_{\theta_0}(t,\omega)\}dt + \frac{1}{T}\int_0^T \log \frac{\lambda_{\theta_0}(t,\omega)}{\sup_{\theta\varepsilon U_s} \lambda_\theta(t,\omega)} dN(t)$$

$$\geqq E\{\Lambda_1(\theta_0;\theta)\} - 2\varepsilon \geqq \varepsilon.$$

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

$$(3.6) \qquad \sup_{\theta\varepsilon U_0} L_T(\theta) \geqq \sup_{\theta\varepsilon\Theta\backslash U_0} L_T(\theta) + \varepsilon T.$$

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 $\hat{\theta}\varepsilon U_0$. This completes the proof.

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

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

$$\frac{1}{T}\frac{\partial^2 L_T(\theta)}{\partial\theta_i\partial\theta_j} = \frac{1}{T}\int_0^T \frac{\partial^2\lambda}{\partial\theta_i\partial\theta_j}\{\frac{dN(t)}{\lambda} - dt\}\Big|_{\theta_0} - \frac{1}{T}\int_0^T \frac{1}{\lambda}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}\Big|_{\theta_0} dN(t)$$

$$- \frac{1}{T}\int_0^T \alpha|\theta - \theta_0| H(t,\omega)dt + \frac{1}{T}\int_0^T \beta|\theta - \theta_0| G(t,\omega)dN(t),$$

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 $T \to \infty$

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

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

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

and

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

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 that if $T \geqq T_0$ then with probability exceeding $1 - \varepsilon$.

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

(3.7)

$$\Big|I_{ij}(\theta_0) + \frac{1}{T}\int_0^T \frac{-1}{\lambda^2}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}\Big|_{\theta_0}dN(t)\Big| < \delta$$

and

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

for $i,j = 1,\cdots,d$. Choose also δ so small that $\{\theta;\ |\theta - \theta_0| < \delta\} \subset U$ and so small 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 $|\theta - \theta_0| < \delta$

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

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 $|\theta - \theta_0| < \delta$.

Remark. In the following case the Hessian is always non-positive definite a.s. ω;

$$\lambda^*_\theta(t,\omega) = \sum_{i=1}^{d} \theta_i \xi^*_i(t,\omega) + \eta^*(t,\omega), \qquad \xi^*_i,\ \eta > 0 \quad \text{a.s.}\omega,$$

$$\lambda^*_\theta(t,\omega) = \exp\{\sum_{i=1}^{d} \theta_i \xi^*_i(t,\omega) + \eta^*(t,\omega)\},$$

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

$$\sum_{i,j=1}^{d} u_i u_j \frac{\partial^2 L^*_T(\theta)}{\partial\theta_i \partial\theta_j} = -\int_0^T \{\sum_{i=1}^{d} u_i \xi^*_i(t,\omega)/\lambda^*_\theta(t,\omega)\}^2 dN(t)$$

$$= -\int_0^T \{\sum_{i=1}^{d} u_i \xi^*_i(t,\omega)\}^2 \ \lambda^*_\theta(t,\omega)dt,$$

respectively.

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

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

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

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

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

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 the sequence $\{\Delta(k-1, k)\}_{k=1,2,\cdots}$ is a stationary ergodic martingale differences with $E\{\Delta(0,1)\Delta(0,1)'\} = I(\theta_0)$ 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 $E\{\Delta(0,1)\Delta(0,1)'\} = E[\frac{\mu}{\lambda^2}\frac{\partial\lambda}{\partial\theta_i}\frac{\partial\lambda}{\partial\theta_j}]$ if $E[dN(t)^2|H_{-\infty,t}] = \mu(t,\omega)dt$.

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 $T \to \infty$,

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

and

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

in law.

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

$$0 = \frac{1}{\sqrt{T}}\frac{\partial L_T^*(\theta_0)}{\partial\theta} + \frac{1}{T}\frac{\partial^2 L_T^*(\theta_0)}{\partial\theta\partial\theta'}\sqrt{T}(\hat{\theta} - \theta_0)$$

$$+\sqrt{T}(\hat{\theta}_T - \theta_0)'\{-\frac{\alpha}{T}\int_0^T H(t,\omega)dt + \frac{\beta}{T}\int_0^T G(t,\omega)dN(t)\}(\hat{\theta}_T - \theta_0)$$

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

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

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

$$2\{L_T^*(\hat{\theta}) - L_T^*(\theta_0)\} = 2\,\frac{\partial L_T^*(\theta_0)}{\partial\theta'}\,(\hat{\theta}_T - \theta_0) + (\hat{\theta}_T - \theta_0)'\,\frac{\partial^2 L_T^*(\theta_0)}{\partial\theta\partial\theta'}\,(\hat{\theta}_T - \theta_0)$$

$$+ |\hat{\theta}_T - \theta_0|^3\{\int_0^T \alpha H(t,\omega)dt + \int_0^T \beta G(t,\omega)dN(t)\}$$

for $\hat{\theta}_T \varepsilon U$, where α, β are some random variable such that $|\alpha|$, $|\beta| \leq d^3/3$. 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 $0 \leq t_0(\omega) \leq t_1(\omega) < ,\cdots, < t_{n-1}(\omega) \leq T$ be the observation on the interval [0,T] from the point process $P_\theta^*(\cdot)$ which has the Radon-Nikodym derivative $\rho_T^*(\omega,\theta) = \exp\{\int_0^T \log \lambda_\theta^*(t,\omega)dN(t) + \int_0^T (1 - \lambda_\theta^*(t,\omega))dt\}$ 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 $\theta\varepsilon\Theta\subset R^d$. Consider the following additional conditions.

Conditions D

(D1) $E_\theta\{\delta_T(\omega)^2\} < \infty$ for all $\theta\varepsilon\Theta$.

(D2) $E_\theta\{\int_0^T \frac{1}{\lambda^*(t,\omega)}\,\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta_i}\,\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta_j}\,dt\} < \infty$ for all $\theta\varepsilon\Theta$ 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 $\theta\varepsilon\Theta$

(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 $\theta\varepsilon\Theta$,

where Π is the probability measure of the standard Poisson process.

Theorem 6. If $E_\theta\{\delta_T(\omega)\} = \theta + b_T(\theta)$, and $b_T(\theta)$ is differentiable, then under the regularity conditions D

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

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

$$I_T^*(\theta) = E_\theta\{\int_0^T \frac{1}{\lambda_\theta^*(t,\omega)} \frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta} \frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta'} dt\}$$

Proof. By (1.2)

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

Thus by (D4) and (1.2)

$$\frac{\partial}{\partial\theta} E_\theta\{\delta_T(\omega)\} = \int\delta_T(\omega) \frac{\partial}{\partial\theta} \rho_T^*(\omega,\theta)\Pi(d\omega)$$

$$= \int\delta_T(\omega)\{\frac{\partial}{\partial\theta} \log \rho_T^*(\omega,\theta)\} \rho_T^*(\omega,\theta)\Pi(d\omega)$$

$$= E_\theta\{\delta_T(\omega)\frac{\partial}{\partial\theta} \log \rho_T^*(\omega,\theta)\}$$

By (2.1) and (D3)

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

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

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

$$\leqq E_\theta[t'\{\delta_T(\omega) - \theta - b_T(\theta)\}\{\delta_T(\omega) - \theta - b_T(\theta)\}'t]E_\theta\{s'\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)s\}.$$

Put

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

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

$$t'\{I + \frac{\partial}{\partial\theta} b_T(\theta)\}E_\theta\{\frac{\partial}{\partial\theta} \log \rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'} \log \rho_T^*(\omega,\theta)\}^{-1}\{I + \frac{\partial}{\partial\theta}b_T(\theta)\}'t$$

$$\leqq t'E_\theta[\{\delta_T(\omega) - \theta - b_T(\theta)\}\{\delta_T(\omega) - \theta - b_T(\theta)\}']t.$$

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

$$E_\theta\{\frac{\partial}{\partial\theta}\log\rho_T^*(\omega,\theta)\frac{\partial}{\partial\theta'}\log\rho_T^*(\omega,\theta)\}$$

$$= E_\theta\{\int_0^T \frac{1}{\lambda_\theta^*(t,\omega)}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta'}dt\}.$$

This completes the proof.

By the Assumption (C2) it is easily seen that

$$\frac{1}{T}E_\theta\{\int_0^T \frac{1}{\lambda_\theta^*(t,\omega)}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta}\frac{\partial\lambda_\theta^*(t,\omega)}{\partial\theta'}dt\} \to I(\theta).$$

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.

IV. CHARACTERIZATION OF POISSON PROCESSES BY A MAXIMUM LIKELIHOOD ESTIMATOR

Suppose a stationary delayed renewal process which 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

$$(4.1) \qquad \int_0^\infty t\,dF_\theta(t) = \int_0^\infty tf_\theta(t)dt = \mu(\theta).$$

Remember the definition of the maximum likelihood estimator given in Section 1, that is, $\hat{\theta}_T$ maximises 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 $\theta\varepsilon\Theta$ and for any $t > 0$ there are $T(\theta) > t$ and $\omega(\theta)\varepsilon\Omega$ 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 $\theta\varepsilon\Theta$.

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^* maximises (2.3). Now fix $n^* = n(\theta^*)$ and $T^* = T(\theta^*)$. Then

$$(4.2) \qquad 0 = \frac{\partial}{\partial\theta} L^*_{T^*}(\theta^*)$$

$$= \frac{\partial}{\partial\theta} \log \mu^{-1}_{(\theta^*)}(1-F_{\theta^*}(t_0)) + \sum_{i=1}^{n^*-1} \frac{\partial}{\partial\theta} \log f_{\theta^*}(t_i - t_{i-1}) + \frac{\partial}{\partial\theta} \log(1-F_{\theta^*}(T^*-t_{n^*-1})).$$

Put

$$(4.3) \qquad G_\theta(s) = \frac{\partial}{\partial\theta} \log f_\theta(s),\ H_\theta(s) = \frac{\partial}{\partial\theta} \log \mu^{-1}_{(\theta)}(1 - F_\theta(s)),$$

$$K_\theta(s) = \frac{\partial}{\partial\theta} \log(1 - F_\theta(T^* - s)) \text{ and } s_i = t_i - t_{i-1},\ i=1,2,\cdots,n^*-1,\ s_0 = t_0.$$

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

$$(4.4) \qquad H_{\theta^*}(s_0) + \sum_{i=1}^{n^*-1} G_{\theta^*}(s_i) = K_{\theta^*}\left(\sum_{i=0}^{n^*-1} s_i\right).$$

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 function equation (see [11] Chapter 3 for example), and we have the general solution for $s > 0$

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

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

Since θ^* is arbitary and by (E2), $a(\theta)$, $b(\theta)$ are indefinitely integrable, 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

$$\lim_{T\to\infty} TE_{\theta,\sigma}\{N(0,T)/T - \theta\}^2 = \sigma^2\theta^3,$$

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 again 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.

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., Vol. 12, 788-792.

[3] Dellacherie, C. (1972). Capacites 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. & Oakes, D. (1974). "A cluster process representation of a self-exiting point process, J. Appl. Prob., 11, 493-503.

[6] Kabanov, Yu. M., Lipster, R. Sh. & 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 Memo., 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.

Chapter 2

ON LEWIS' SIMULATION METHOD FOR POINT PROCESSES

(Published in IEEE Transactions on Information Theory, Vol.IT-27, No.1, Jan., pp.23-31, 1981)

ABSTRACT

A simple and efficient method of simulation is discussed for point processes which are specified by their conditional intensities. The method is based on the thinning algorithm which was introduced recently by Lewis and Shedler for the simulation of non-homogeneous Poisson processes. A proof and algorithms are given for past dependent point processes containing multivariate case. The simulations are performed for some parametric conditional intensity functions, and accuracy of the simulated data are discussed and demonstrated by the likelihood ratio test and the minimum AIC procedure.

I. INTRODUCTION

Any point process (N_t, F_t, P) on a finite interval $(0,T]$ is a submartingale and therefore by the Doob-Meyer decomposition we get $N_t = m_t + A_t$, where m_t is (F_t,P)-martingale and A_t is the natural increasing process. It is known that there is a predictable process (λ_t, F_t) such that $A_t = \int_0^t \lambda_s ds$ if and only if P is absolutely continuous with respect to the standard Poisson process P_0; furthermore $\lambda = \{\lambda_t,\ 0<t\leq T\}$ corresponds uniquely to the process P, and the Radon-Nikodym derivative is given by

$$\frac{dP}{dP_0} = \exp\{\int_0^T \log \lambda_t dN_t + \int_0^T (1-\lambda_t)dt\}.$$

Similar result is obtained for multivariate or marked point processes ([7], [8]).

The main object of this paper is to discuss the applications of Lewis' thinning simulation algorithm to any point process which is absolutely continuous with respect to the standard Poisson process. Recently, Ozaki [11] generated simulation data for Hawkes' self-exciting processes by making use of a recursive structure. However his method is not fast enough unless the process has a simple structure, because given a past history of the process $t_1, t_2, \cdots, t_n$ and a uniform random number U_{n+1} from the interval $(0,1)$ we have to solve the equation $U_{n+1} = S(t_{n+1}|t_1,\cdots,t_n)$ by Newton's iteration method to get the next point t_{n+1}, where S is the conditional survivor function $S(t|t_1,\cdots,t_n) = \exp\{-\int_{t_n}^t \lambda(s|t_1,\cdots,t_n)ds\}$. We do not solve the equation to get the next point. The idea of simulating these point processes by thinning is developed using algorithms due to Lewis and Shedler [9] for the simulations of non-homogeneous Poisson processes.

In section 2 we will give the simulation method and a proof for past dependent point processes containing multivariate cases. Also some typical algorithms will be given. In section 3 we will give some examples of parametric intensity functions for the simulations, and obtain their maximum likelihood estimates from the simulated data. Also accuracy of the simulated data will be discussed by the likelihood ratio test or the minimum AIC procedure. Numerical results are listed in Appendix.

II. SIMULATION OF POINT PROCESSES

Consider a point process $(N,F,P) = \{N_t, F_t, 0<t\leq T, P\}$ on a fixed interval with its F-predictable intensity process $\lambda = \{\lambda_t\}$, where $F = \{F_t\}$ is a family of right-continuously increasing σ-fields. Suppose we get a positive F-predictable piece-wise constant process $\lambda^* = \{\lambda^*_t\}$ which is constructed pathwise in such a way that $\lambda_t \leqq \lambda^*_t$, a.s., $0<t\leqq T$. Then λ^* can be an intensity process of a locally homogeneous Poisson process $(N^*,F,P) = \{N^*_t, F_t, P\}$ with piecewise constant intensity changing its rate according to the past history F_t. The main result, which is formally similar to the one given in [9], is described as follows:

Let $t^*_1 < t^*_2 < ,\cdots, < t^*_{N^*_T}$ be the points in $(0,T]$ of the process (N^*,F,P). Delete the points t^*_j with probability $1 - \lambda_{t_j}/\lambda^*_{t_j}$ for $j = 1,2,\cdots,N^*_T$. Then the remaining points $\{t_i\}$ form a point process (N,F,P) with the conditional intensity $\lambda = \{\lambda_t\}$ in the interval $(0,T]$.

It is readily seen from the predictability of λ^* that the constructions of λ^*, t^*_j and t_i should be performed sequentially in the following manner:

1. Suppose that the last point before time t have just been obtained. Then construct λ^*_t which is F_{t_i}-measurable, piecewise constant and $\lambda^*_t \geqq \lambda_t$, for $t \geqq t_i$.

2. Simulate homogeneous Poisson points t^*_j $(>t_i)$ according to the intensity λ^*_t.

3. For each of the points $\{t^*_j\}$, the probability $\lambda_{t^*_j}/\lambda^*_{t^*_j}$ is given conditionally independent of t^*_j under the past history $F_{t^*_j}$.

4. t_{i+1} is the first accepted point among t^*_j $(>t_i)$.

Details of algorithms will be given later.

Now by generalizing the result to a multivariate point process, we have the following proposition:

Proposition 1. Consider a multivariate point process (N^p,F,P), $p=1,2,\cdots,m$ on an interval $(0,T]$ with their joint intensity $(\underline{\lambda},F) = \{\lambda_t^p,F_t\}, p=1.\cdots,m$. Suppose we can find a one-dimensional F-predictable process λ_t^* which is defined pathwise satisfying

$$\sum_{p=1}^{m} \lambda_t^p \leqq \lambda_t^*,\ 0 < t \leqq T,\ \text{P-a.s.},$$

and then define

$$\lambda_t^0 = \lambda_t^* - \sum_{p=1}^{m} \lambda_t^p.$$

Let $t_1^*, t_2^*,\cdots,t_{N_T^*}^* \in (0,T]$ be the points of the process (N^*,F,P) with intensity process λ_t^*. For each of the points, attach a mark $p = 0,1,\cdots,m$ with probability $\lambda_{t_j^*}^p/\lambda_{t_j^*}^*$. Then the points with marks $p = 1,2,\cdots,m$, provide the multivariate point process which is the same as given above.

Proof. Define a random measure for the finite marked process (using the notation in [7].)

$$M(dt,p) = N^*(dt)I(t,p),\quad p = 1,\cdots,m,$$

where $I(t,p)$ is the transition random measure of the marks under the condition that there is a point at t. As a result of the conditions of the proposition, $I(t,p)$ has the following properties:

(i) $I(t,q) = \delta_p(q)$ with probability λ_t^p/λ_t^* for $p,q=0,1,2,\cdots,m$, where $\delta_p(q)$ is a Dirac's delta function.

(ii) For fixed t, $N^*(dt)$ and $\{I(t,p);\ p=0,1,\cdots,m\}$ are conditionally independent given F_t.

Then, for each mark $p=1,2,\cdots,m$, the intensity measure of the marked point process is given by

$$\nu(dt,p) = E[M(dt,p)|\ F_t]$$
$$= E[N^*(dt)I(t,p)|\ F_t]$$
$$= E[N^*(dt)|\ F_t]E[I(t,p)|\ F_t].$$

Now by definition

$$E[N^*(dt)|\ F_t] = \lambda_t^* dt,$$

and also

$$E[I(t,p)|\ F_t] = 1\cdot P\{I(t,p)=1|\ F_t\} + 0\cdot P\{I(t,p)=0|\ F_t\}$$
$$= \lambda_t^p/\lambda_t^*.$$

Therefore for each $p = 1,2,\cdots,m$, we get

$$\nu(dt,p) = \lambda_t^p dt.$$

Since the predictable random measure correspond uniquely to the multivariate process (see[7], [8]), this completes the proof.

We can now give some cases of typical algorithms according to the proposition.

Algorithm 1. Bivariate (doubly) Poisson process with intensity process $\{X_t^p(\omega),F_t\}$, $p=1,2$.

1. Get a path function of the process $\omega_p(t) = X_t^p(\omega)$, $0<t\leq T$, $p=1,2$.

2. Take a piecewise constant function $\omega^*(t)$ such that $\omega_1(t) + \omega_2(t) \leqq \omega^*(t)$. For efficiency of simulation we should take $\omega^*(t)$ as close as possible to $\omega_1(t) + \omega_2(t)$.

3. Simulate stationary Poisson processes for each interval of constant intensity. Denote the points by $t_1^*, t_2^*,\cdots, t_{N_T^*}^*$.

4. Set $k = 1$, $i = 0$ and $j = 0$.

5. Independently generate a uniform random number U_k on $(0,1)$.

6. If $U_k \leqq \omega_1(t_k^*)/\omega^*(t_k^*)$, set i equal to $i + 1$ and $t_i^{(1)} = t_k^*$.

7. If $U_k \leqq \{\omega_1(t_k^*) + \omega_2(t_k^*)\}/\omega^*(t_k^*)$, set j equal to $j + 1$ and $t_j^{(2)} = t_k^*$.

8. Set k equal to $k + 1$. If $N_T^* < k$, then stop. Otherwise go to 5.

Now consider the case of univariate self-exciting processes. Since $F_t = \sigma\{N_s, 0<s\leqq t\}$, the intensity of the process is given by a function of t and the points t_i before t, i.e. $\lambda_t = \lambda(t|\ t_1,\cdots,t_n)$. Here we have two types of the intensity process.

Consider the case where the path function of a intensity process is decreasing if there are no more points occur. We note that the predictability of λ_t implies left-continuity of path functions. We assume here that the minimum value of the intensity function is μ and the jump size at each point is not larger than α, and Λ_i^*'s are value of the piecewise constant function, such that $\lambda(t|\ t_1,\cdots,t_n) \leqq \Lambda_i^*$ for $t_n \leqq s_i \leqq t < s_{i+1} \leqq t_{n+1}$, say.

Algorithm 2.

1. Set $\Lambda_0^* = \mu$ and put $s_0 = 0$.
2. Generate U_0 and put $u_0 = -\log U_0/\Lambda_0^*$.
3. If $u_0 \leqq T$ then put $t_1 = u_0$. Otherwise stop.
4. Set $i = j = k = 0$ and $n = 1$.
5. Set k equal to k + 1 and put $\Lambda_k^* = \lambda(t_n|\ t_1,\cdots,t_{n-1}) + \alpha$.
6. Set j equal to j + 1 and generate U_j.
7. Set i equal to i + 1 and put $u_i = -\log U_j/\Lambda_k^*$.
8. Put $s_i = s_{i-1} + u_i$. If $s_i > T$ then stop.
9. Set j equal to j + 1, and generate U_j.
10. If $U_j \leqq \lambda(s_i|\ t_1,\cdots,t_{n-1})/\Lambda_k^*$ then set n equal to n + 1, put $t_n = s_i$ and go to 5.
11. Set k equal to k + 1, put $\Lambda_k^* = \lambda(s_i|\ t_1,\cdots,t_{n-1})$ and go to 6.

If a sample function of the intensity function $\lambda(t|\ t_1,\cdots,t_n)$ is not always decreasing, but only has a decreasing tail, then we can define a process $\lambda^{**}(t|\ t_1,\cdots,t_n)$ which is always decreasing and satisfies $\lambda(t|\ t_1,\cdots,t_n) \leqq \lambda^{**}(t|\ t_1,\cdots,t_n)$ for $t_n \leqq t$ (see Example 1). Replace

$\lambda(t_n | t_1,\cdots,t_{n-1})$ by $\lambda^{**}(t_n | t_1,\cdots,t_{n-1})$ in 5. and $\lambda(s_i | t_1,\cdots,t_{n-1})$ by $\lambda^{**}(s_i | t_1,\cdots,t_{n-1})$ in 11. Then we can have the point process with intensity λ_t by using a modified Algorithm 2.

Assume now that the intensity function $\lambda_t(\omega) = \lambda(t | t_1,\cdots,t_n)$ is monotonically increasing unless any more points occur. In the following algorithm the interval $(0,T]$ is divided equally into subintervals $(kr,(k+1)r]$, where the length r should be determined suitably in accordance with the efficiency of the algorithm.

Algorithm 3.

1. Set $i = n = 1$.
2. Put $\lambda_i^* = \lambda((i+1)r | t_1,\cdots,t_n)$.
3. Generate a homogeneous Poisson process with intensity λ_i^* on the interval $(kr,(k+1)r]$.
4. If the number of the points on the interval, say N_i^*, is zero, then go to 11.
5. Denote the ordered points on the interval $(ir,(i+1)r]$ by $s_1^*, s_2^*,\cdots,s_{N_i^*}^*$.
6. Set $j = 1$.
7. Generate U_j, uniformly distributed between 0 and 1.
8. If $U_j > \lambda(s_j^* | t_1,\cdots,t_n)/\lambda_i^*$, then go to 7.
9. Put $t_n = s_j^*$ and set n equal to $n + 1$.
10. Set j equal to $j + 1$. If $j \leq N_i^*$, then go to 7.
11. Set i equal to $i + 1$. If $(i+1)r \leq T$, then go to 2.
12. Stop.

Thus $t_1,t_2,\cdots$ are the data which are required. It is recommended for the numerical accuracy in 4 to adopt the method which generates a Poisson random number N_i^* and then uniform random numbers from $(ir,(i+1)]$ according to N_i^*.

If $\lambda_t = \lambda_t(t | t_1,\cdots,t_n)$ is only eventually increasing (unless any more points occur), a modification similar to that for Algorithm 2 is possible. That is to say, construct a function $\lambda^{**}(t|\ t_1,\cdots,t_n)$ which is increasing in $(ir,(i+1)r]$, and satisfy $\lambda(t|\ t_1,\cdots,t_n) \leqq \lambda^{**}(t|\ t_1,\cdots,t_n)$ for $ir < t \leqq (i+1)r$. Then change 2. of Algorithm 2 as follows;

2*. $\lambda_i^* = \lambda^{**}((i+1)r|\ t_1,\cdots,t_n)$.

It is not difficult to construct simulation algorithms for multivariate mutually-exciting point processes, or mixed doubly Poisson and self-exciting point processes in terms of the above algorithms.

III. SOME EXAMPLES AND DISCUSSIONS

Hawkes' Self-Exciting Process

The intensity function is given by

$$\lambda(t) = \lambda(t \mid t_1, \cdots, t_n) = \mu + \int_0^t \nu(t-s)dN(s),$$

where $\mu > 0$, $\nu(s) \geqq 0$ and $\int_0^\infty \nu(s)ds < 1$ for the asymptotic stationarity of the process. Hawkes and Oakes [5] first gave the author the idea that the simulation for the process may be constructed through the simulations of non-homogeneous Poisson processes. Indeed it says that a Hawkes' self-exciting process is nothing but an immigrant-birth process which is composed of homogeneous Poisson immigrant with rate μ and non-homogeneous Poisson decendants with rate $\nu(s)$.

As a parametrization of $\nu(s)$ Hawkes [4] used an exponential, $\nu(s) = \alpha e^{-\beta s}$. In this case we can adopt Algorithm 2 for the simulation. Ozaki and Akaike [12] suggested a generalized parametrization $\nu(s) = \sum_{j=0}^{p} \alpha_j s^j e^{-\beta s}$, where α's and β (>0) are suitably restricted to satisfy $\nu(s) \geqq 0$ and $\int_0^\infty \nu(s)ds < 1$. This is a decreasing function for sufficiently large s. Thus we have a function $\nu^{**}(s)$ which is always decreasing and $\nu(s) \leqq \nu^{**}(s)$ for $s \geqq 0$, say $\nu^{**}(s) = \sum_{j=0}^{p} \alpha_j^+ (j/\beta)^j e^{-j}$

where $\alpha_j^+ = \max(\alpha_j, 0)$. Therefore making use of a predictable intensity $\nu^{**}(s) = \mu + \int_0^t \nu^{**}(s)dN(s)$, where $N(s)$ is the point process generated by the intensity function $\lambda(t) = \mu + \int_0^t \nu(t-s)dN(s)$, the modified Algorithm 2 can be applied for the simulation. The jump size for this case is $\sum_{j=0}^{p} \alpha_j^+ (j/\beta)^j e^{-j}$.

Suppose $t_1, t_2, \cdots, t_n$ in $(0,T]$ are the simulated data. Then the log-likelihood function is given by

$$L_T(\alpha_0, \cdots, \alpha_p, \beta) = \sum_{i=1}^{n} \log\{\mu + \sum_{j=0}^{p} \alpha_j R_j(i)\} - \mu T - \sum_{i=1}^{n} \sum_{j=0}^{p} \alpha_j S_j(T-t_i),$$

where $R_j(i)$ and $S_j(t)$ are given recursively in the following way:

Set $t_0 = 0$, $R_0(1) = 0$, $S_0(t) = (1 - e^{-\beta t})/\beta$ and $A_k(t) = t^k e^{-\beta t}$.

Then for $j = 0,1,2,\cdots$ and $i = 2,3,\cdots$,

$$R_j(i) = A_j(t_i - t_{i-1}) + \sum_{k=0}^{j} {}_jC_k \quad A_k(t_i - t_{i-1})R_k(i-1),$$

and

$$S_{j+1}(t) = \{(j+1)S_j(t) - A_{j+1}(t)\}/\beta,$$

where the notation ${}_jC_k$ means the combination factorial.

The gradient vector and Hessian matrix of the log-likelihood function are also written recursively using the function above. It is worthwhile to note that the simulation can also be much faster if we make use of the recursive structure of the intensity function $\lambda(t| t_1,\cdots,t_n)$. That is,

$$\lambda(t_{n+1}| t_1,\cdots,t_n) = \mu + \sum_{j=0}^{p} \alpha_j R_j(n+1).$$

Linear Wold Process

The intensity function is given by

$$\lambda(t) = \mu + \alpha_1(t-t_{(1)}) + \sum_{k=2}^{p} \alpha_k(t_{(k-1)}-t_{(k)}),$$

where μ and the α's are non-negative parameters and $t_{(k)}$ is the k-th last point before t. The point process with this intensity is always asymptotically stationary (see[3]). It is easily seen that the simulation of this point process is performed by the simple relation $t_{n+1} = t_n + \frac{1}{\alpha_1}\{-\beta_n + (\beta_n^2 - 2\alpha_1 \log U_{n+1})^{1/2}\}$, where $\beta_n = \mu + \sum_{k=2}^{p} \alpha_k(t_{n-k+2} - t_{n-k+1})$ and U_{n+1} is a uniform random number from (0,1).

However we would like to adopt here the modified Algorithm 3 with setting

$$\lambda^{**}(t| \quad t_1,\cdots,t_n) = \mu + \max_{0<i\leq k} (\alpha_i)(t - t_{(k)}).$$

Suppose $t_1, t_2, \ldots, t_n$ are the simulation data on the interval (0,T]. Then setting $t_0 = 0$ and $t_{n+1} = T$, the log -likelihood function is given by

$$L_T(\mu, \alpha_1, \ldots, \alpha_p) = \sum_{i=1}^{n} \log\{\mu + \sum_{k=1}^{p} \alpha_k(t_{i-k+1}-t_{i-k})\}$$
$$- \mu T - \sum_{i=1}^{n+1} \{\alpha_1(t_i-t_{i-1})^2/2 + \sum_{k=2}^{p} \alpha_k(t_i-t_{i-1})(t_{i-k+1}-t_{i-k})\}.$$

One of the nice properties of the model is that Hessian matrix of the log-likelihood function is negative definite everywhere with respect to the parameters (see [10] p.255 for example).

Stress-Release Process

Vere-Jones [14] has suggested models for a series of big earthquakes. One of these models is defined by the intensity function

$$\lambda(t) = e^{\alpha+\beta t-\gamma N(t-)}, \text{ where } N(t-) = N[0,t).$$

(A similar process is discussed by Isham and Westcott [6]). This is asymptotically stationary and mean intensity rate is γ/β. Although this process is obtained by the simple relation

$$t_{n+1} = t_n + \frac{1}{\beta} \log \{1 - \beta e^{-\alpha-\beta t_n+\gamma n} \cdot \log U_{n+1}\},$$

we apply here Algorithm 3 for the simulation.

Given simulated data $t_1, t_2, \ldots, t_n$ on the interval (0,T], we have the log-likelihood function (setting $t_0=0$ and $t_{n+1}=T$)

$$L_T(\alpha,\beta,\gamma) = \sum_{i=1}^{n} \log\{\alpha+ \beta t_i-\gamma(i-1)\} + \sum_{i=1}^{n+1} e^{\alpha-\gamma(i-1)}\cdot\{e^{\beta t_{i-1}}-e^{\beta t_i}\}/\beta.$$

Nonlinear Hawkes' Type Point Process

Consider an intensity function of the form

$$\lambda(t) = \mu + \int_0^t \nu(t-s)dN(s) + \int_0^t\int_0^s \pi(t-s,s-u)dN(s)dN(u),$$

where $\mu>0$, $\nu(s)\geqq 0$ and $\pi(s,u)\geqq 0$. It is necessary for the asymptotic stationarity that

$$\nu + \pi_0 < 1 \quad \text{and} \quad (1 - \nu - \pi_0)^2 \geqq 4\pi(\mu + \pi_C),$$

where $\nu = \int_0^\infty \nu(s)ds$, $\pi_0 = \int_0^\infty \pi(s,0)ds$, $\pi = \int_0^\infty\int_0^\infty \pi(s,u)dsdu$ and $\pi_C = \int_0^\infty\int_0^\infty \pi(s,u)C(u)dsdu$ ($C(u)$ is an auto-covariance of the process). Unfortunately we cannot evaluate π_C nor $C(u)$. We can only hypothesize the domain. For example, the noise level μ of the Poisson must be small enough for the asymptotic stationarity.

A parametric example of $\nu(s)$ and $\pi(s,u)$ is

$$\nu(s) = \alpha e^{-\beta s} \quad \text{and} \quad \pi(s,u) = \gamma e^{-\beta(s+u)}.$$

Algorithm 2 is applied for this case. Now suppose the simulation data $t_1, t_2, \ldots, t_n$ on the interval $(0,T]$ are obtained. Then the log-likelihood function for the case is

$$L_T(\mu, \alpha, \beta,\gamma) = \sum_{i=1}^{n} \log\{\mu+\alpha R_1(i)+\gamma R_3(1)\} - \mu T$$
$$- \frac{\alpha}{\beta}\sum_{i=1}^{n}\{1-e^{-\beta(T-t_i)}\} - \frac{\gamma}{\beta}\sum_{i=1}^{n}\{1-e^{-\beta(T-t_i)}\}R_2(i),$$

where $R_1(i)$, $R_2(i)$ and $R_3(i)$ are given recursively as follows;

$$R_1(1) = 0,\ R_2(1) = 1,\ R_3(1) = 0,$$
$$R_1(i) = e^{-\beta(t_i-t_{i-1})}\{R_1(i-1)+1\},$$
$$R_2(i) = e^{-\beta(t_i-t_{i-1})}\cdot R_2(i-1) + 1$$

and

$$R_3(i) = e^{-\beta(t_i-t_{i-1})}\cdot\{R_2(i-1) + R_3(i-1)\}.$$

The gradiant and Hessian are given similarly. It is worthwhile to note that the simulation is much quicker if we make use of the recursive structure.

Bivariate Wold Process

The intensity functions are given by

$$\begin{pmatrix}\lambda_1(t)\\ \lambda_2(t)\end{pmatrix} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} + \begin{pmatrix}\alpha_1 & \beta_1\\ \gamma_1 & \delta_1\end{pmatrix}\begin{pmatrix}t-t_{(1)}^{(1)}\\ t-t_{(1)}^{(2)}\end{pmatrix} + \sum_{k=2}^{p}\begin{pmatrix}\alpha_k & \beta_k\\ \gamma_k & \delta_k\end{pmatrix}\begin{pmatrix}t_{(k-1)}^{(1)}-t_{(k)}^{(1)}\\ t_{(k-1)}^{(2)}-t_{(k)}^{(2)}\end{pmatrix}.$$

The coefficients must be non-negative for the asymptotic stationarity under condition th μ_1 and μ_2 are strictly positive. If $\mu_1 = 0$ or $\mu_2 = 0$ then there are explosive cases even if the other coefficients are non-negative. For example consider the simplest case p=1. If $\mu_1 > 0$, $\mu_2 = 0$, $\alpha_1 = 0$, $\beta_1 > 0$, $\gamma_1 > 0$ and $\delta_1 = 0$, then this provides an explosive process.

Consider an asymptotically stationary case with p = 1. Let $\{t_i^{(1)}\}$, $i=1,2,\cdots,n$ and $\{t_j^{(2)}\}$, $j=1,2,\cdots,m$ be the data from the model. Then the log-likelihood function is given by

$$L_T(\mu_1,\mu_2,\alpha,\beta,\gamma,\delta) = L_T^{(1)}(\mu_1,\alpha,\beta) + L_T^{(2)}(\mu_2,\gamma,\delta)$$

and

$$L_T^{(1)} = \sum_{i=1}^{n} \log \{\mu_1+\alpha(t_i^{(1)}-t_{i-1}^{(1)})+\beta(t_i^{(1)}-t_{(i)}^{(2)})\}$$

$$- \mu_1 T - \alpha\sum_{i=1}^{n}(t_i^{(1)}-t_{i-1}^{(1)})^2/2 - \beta\sum_{j=1}^{m}(t_j^{(2)}-t_{j-1}^{(2)})^2/2$$

$$L_T^{(2)} = \sum_{j=1}^{m} \log \{\mu_2+\gamma(t_j^{(2)}-t_{(j)}^{(1)})+\delta(t_j^{(2)}-t_{j-1}^{(2)})\}$$

$$- \mu_2 T - \delta\sum_{j=1}^{m}(t_j^{(2)}-t_{j-1}^{(2)})^2/2 - \gamma\sum_{i=1}^{n}(t_i^{(1)}-t_{i-1}^{(1)})$$

where $t_{(i)}^{(2)}$ is the last point on the line $t^{(2)}$ before $t_i^{(1)}$, also $t_{(j)}^{(1)}$ is the last point on the line $t^{(1)}$ before $t_j^{(2)}$, and $t_0^{(1)} = t_0^{(2)} = 0$. Notice that the minimization of L_T is equivalent to the separate minimizations of $L_T^{(1)}$ and $L_T^{(2)}$, provided the parameters are independent.

The simulation algorithm is performed in accordance with Algorithm 1 and modified Algorithm 3.

Bivariate Hawkes' Mutually Exciting Process [4]

The intensity function are given by

$$\lambda_1(t) = \mu_1 + \int_0^t \nu_{11}(t-s)dN_1(s) + \int_0^t \nu_{12}(t-s)dN_2(s),$$

$$\lambda_2(t) = \mu_2 + \int_0^t \nu_{21}(t-s)dN_1(s) + \int_0^t \nu_{22}(t-s)dN_2(s),$$

where $\nu_{ij}(s) \geqq 0$. For the asymptotic stationarity it is necessary that the moduli of a eigen-values of the matrix $\{\nu_{ij}\}$ are less than 1, where $\nu_{ij} = \int_0^\infty \nu_{ij}(s)ds$. Parametrizing the functions $\nu_{ij}(s) = \alpha_{ij}e^{-\beta_i s}$, we can simulate the data in accordance with Algorithms 1 and 2. For data $\{t_i^{(1)}\}$, i=1,2,$\cdots$,n and $\{t_j^{(2)}\}$, j=1,2,$\cdots$,m, the log-likelihood of the model is

$$L_T(\mu_1,\mu_2,\beta_1,\beta_2,\alpha_{11},\alpha_{12},\alpha_{21},\alpha_{22})$$
$$= L_T^{(1)}(\mu_1,\beta_1,\alpha_{11},\alpha_{12}) + L_T^{(2)}(\mu_2,\beta_2,\alpha_{21},\alpha_{22})$$

and

$$L_T^{(1)}(\mu_1,\beta_1,\alpha_{11},\alpha_{12}) = \sum_{i=2}^{n} \log\{\mu_1 + \alpha_{11}R_{11}(i) + \alpha_{12}R_{12}(i)\}$$
$$- \mu_1 T - \frac{\alpha_{11}}{\beta_1}\sum_{i=1}^{n}\{1-e^{-\beta_1(T-t_i^{(1)})}\} - \frac{\alpha_{12}}{\beta_1}\sum_{j=1}^{m}\{1-e^{-\beta_1(T-t_i^{(2)})}\}$$

$$L_T^{(2)}(\mu_2,\beta_2,\alpha_{21},\alpha_{22}) = \sum_{j=2}^{m} \log\{\mu_2 + \alpha_{21}R_{21}(j) + \alpha_{22}R_{22}(j)\}$$
$$- \mu_2 T - \frac{\alpha_{21}}{\beta_2}\sum_{j=1}^{m}\{1-e^{-\beta_2(T-t_i^{(1)})}\} - \frac{\alpha_{22}}{\beta_2}\sum_{j=1}^{m}\{1-e^{-\beta_2(T-t_j^{(2)})}\},$$

where R_{ij}'s are given recursively as follows;

$$R_{11}(1) = R_{12}(1) = R_{21}(1) = R_{22}(1) = 0,$$

$$R_{11}(i) = e^{-\beta_1(t_i^{(1)}-t_{i-1}^{(1)})}\cdot\{1.0 + R_{11}(i-1)\},$$

$$R_{12}(1) = e^{-\beta_1(t_i^{(1)}-t_{i-1}^{(1)})}\cdot R_{12}(i-1) + \sum_{\{j:\ t_{i-1}^{(1)}\leqq t_j^{(2)}<t_i^{(1)}\}} e^{-\beta_1(t_i^{(1)}-t_j^{(2)})},$$

$$R_{21}(j) = e^{-\beta_2(t_j^{(2)}-t_{j-1}^{(2)})}\cdot R_{21}(j-1) + \sum_{\{i:\ t_{j-1}^{(2)}\leq t_i^{(1)}<t_j^{(2)}\}} e^{-\beta_2(t_j^{(2)}-t_i^{(1)})}$$

and

$$R_{22}(j) = e^{-\beta_2(t_j^{(2)}-t_{j-1}^{(2)})}\cdot\{1.0 + R_{22}(j-1)\}.$$

The gradient vector and Hessian matrix are given similarly in a recursive formula. Also the simulation algorithm should be performed recursively for the efficiency.

IV. CONCLUDING REMARKS

In this section we discuss whether the simulation data are statistically accurate enough in each case. In [10] a collection of the regularity conditions is given to prove the following:

1. The maximum likelihood estimator is consistent, i.e. $\hat{\theta}_T \to \theta_0$ a.s. as $T\to\infty$.
2. $\sqrt{T}(\hat{\theta}_T-\theta_0)$ is asymptotically normal according to $N(\underline{0}, I(\theta_0)^{-1})$ as $T\to\infty$, where each component of the Fisher's mean rate information matrix is given $I_{ij}(\theta_0) = E\{(1/\lambda_\theta)(d\lambda/d\theta_i)(d\lambda/d\theta_j)\}_{\theta=\theta_0}$.
3. $2\{L_T(\hat{\theta})-L_T(\theta_0)\}$ is asymptotically χ_k^2-distributed as $T\to\infty$, where k is the dimension of the parameter θ.

The examples in the preceding section satisfy the conditions basically, although the multivariate case is not treated there. So the adoption of the minimum AIC procedure [1] is justified here. That is to say, we consider two competing models H_0 and H_1, where H_0 is the model supposed to have the true parameter θ_0, and H_1 is any other model with the fixed dimension k of the parameter θ. The value of the AIC for each model is

$$AIC_0 = (-2)(\text{value of log-likelihood at } \theta_0),$$

since the number of unknown parameters in H_0 is zero, and

$$AIC_1 = (-2)(\text{maximized value of log-likelihood}) + 2k.$$

The AIC is an estimator of the expected negative entropy which is a natural measure of discrimination between the true distribution for the data and the estimated probability distribution. Therefore if the simulated data are correctly distributed according to H_0, then we can expect $AIC_0 < AIC_1$. Another useful method is to adopt the likelihood ratio test of H_0 against H_1. Under regular situations such as nested sequence of models, the relationship between the AIC's and the likelihood ratio statistic $\Delta(\theta_0, \hat{\theta}_T)$ is given by

$$\Delta(\theta_0, \hat{\theta}_T) = AIC_0 - AIC_1 + 2k,$$

which is asymptotically χ_k^2-distributed. See [13] for an extensive discussion of the relation between minimum AIC procedure and likelihood ratio test procedure.

Using physically generated random numbers, we performed the simulation experiments five times for each case of the examples, and then the maximum likelihood estimates and minus of the log-likelihood are listed at the tables in the APPENDIX. We made use of the Davidon-Flecher-Powell method for the non-linear optimization. From the tables we can see that the maximum likelihood estimates get more accurate as the sample size (number of points) or the length of observed interval increases. Also it is seen that for each sample size or length of the interval, AICs and log-likelihood ratio tests work well for the justification of the accuracy of the simulations.

ACKNOWLEDGEMENT

The idea of the thinning method was communicated by Prof. P. A. W. Lewis, who kindly corrected the earlier versions of the paper. For these the author is very grateful. The author also would like to thank G. Kitagawa for providing a program in [2] and valuable suggestions such as recursive calculation technique, and to K. Katsura for his generous help in testing the algorithms.

REFERENCES

[1] Akaike, H. (1976). "On entropy maximization principle", Application of Statistics, Krishnaiah, P.R. ed., North-Holland, Amsterdam, 27-41.

[2] Akaike, H., Kitagawa, G., Arahata, E. and Tada, F. (1979). "TIMSAC-78", Computer Science Monographs, The Institute of Statistical Mathematics, Tokyo.

[3] Athreya, K. B., Tweedie, R. L. and Vere-Jones, D. (1978). "Asymptotic behaviour of point processes with Markov-dependent intervals", preprint.

[4] Hawkes, A. G. (1971). "Spectra of some self-exciting and mutually exciting point processes", Biometrika 58, 83-95

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

[6] Isham, V. and Westcott, M. (1979). "A self-correcting point process", Stoc. Proc. Appl. 8, 335-348.

[7] Jacod, J. (1975). "Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales", Z. f. Wahr. Verw. Geb. 31, 235-253.

[8] Kabanov, Yu. M., Lipcer, R. S. and Shiryaev, A. N. (1976). "Criteria of absolute continuity of measures corresponding to multivariate point processes", Proc. 3rd U.S.S.R. and Japan Symposium on Probability Theory, Springer, Berlin.

[9] Lewis, P. A. W. and Shedler, G. S. (1979). "Simulation of non-homogeneous Poisson processes by thinning", Naval Research Logistics Quarterly, 26, No. 3, 403-413.

[10] Ogata, Y. (1978). "The asymptotic behavior of maximum likelihood estimators for stationary point processes", Ann. Inst. Stat. Math. 30, No. 2, A, 243-261.

[11] Ozaki, T. (1978). "Maximum likelihood estimation of Hawkes' self-exciting point processes", Ann. Inst. Stat. Math. 30, No. 3, B, 145-155.

[12] Ozaki, T. and Akaike, H. (1979). "On a generalized parametric model for Hawkes' self-exciting point processes", Research Memo. No. 149.

[13] Sakamoto, Y. and Akaike, H. (1978). "Analysis of cross classified data by AIC", Ann. Inst. Stat. Math. 30, No. 1, B, 185-197.

[14] Vere-Jones, D. (1978). "Earthquake prediction - A statistician's view", J. Phys. Earth, 26, 129-146.

APPENDIX

Table 1. Hawkes' self-exciting process

$$\lambda(t) = \mu + \int_0^t \{(\alpha_0 + \alpha_1(t-s) + \alpha_2(t-s)^2\} e^{-\beta(t-s)} dN(s),$$

where

$$\begin{pmatrix} \mu, \beta \\ \alpha_0, \alpha_1, \alpha_2 \end{pmatrix} = \begin{pmatrix} 0.700 & 1.100 & \\ 0.045 & -0.300 & 0.500 \end{pmatrix}$$

numbers of data	-log(likelihood) at the true parameter	-log(maximum of likelihood)	maximum likelihood estimates
n = 500	218.895	216.715	0.816 0.970 -0.136 -0.161 0.382
	214.112	213.181	0.760 1.125 0.113 -0.427 0.570
	308.358	307.341	0.794 1.234 0.089 -0.288 0.531
	368.521	365.038	0.923 1.213 0.030 -0.412 0.487
	292.764	289.974	0.540 1.095 0.257 -0.434 0.494
n = 50000	26376.292	26374.593	0.721 1.115 0.032 -0.305 0.520
	26104.488	26099.494	0.716 1.099 0.045 -0.378 0.539
	25925.125	25922.638	0.707 1.134 0.040 -0.344 0.567
	25835.985	25832.676	0.724 1.108 0.056 -0.361 0.532
	25574.279	25570.103	0.715 1.107 0.034 -0.244 0.483

Table 2. Linear Wold process

$$\lambda(t) = \mu + \alpha(t-t_{(1)}) + \beta(t_{(1)}-t_{(2)}) + \gamma(t_{(2)}-t_{(3)}),$$

where

$$(\mu, \alpha, \beta, \gamma) = (2.000, 1.400, 3.900, 2.700).$$

number of data	-log(likelihood at the true parameter)	-log(maximum of likelihood)	maximum likelihood estimates			
n = 500	-158.717	-160.946	1.637	1.505	3.459	3.273
	-178.251	-179.852	2.540	1.067	2.624	1.814
	-165.349	-167.672	2.192	0.529	2.874	3.075
	-199.151	-203.298	1.418	3.049	3.490	4.649
	-190.100	-190.534	1.750	1.861	4.632	2.726
n = 50000	-18794.831	-18796.546	2.008	1.386	3.792	2.802
	-18862.763	-18865.039	1.944	1.474	3.932	2.863
	-18655.569	-18656.813	2.002	1.458	3.797	2.680
	-18664.021	-18664.801	2.005	1.426	3.806	2.695
	-18766.239	-18767.236	1.978	1.358	3.953	2.785

Table 3 Stress-release process

$$\lambda(t) = e^{\alpha+\beta t-\gamma N[0,t)},$$

where

$(\alpha, \beta, \gamma) = (3.000 \ \ 2.000 \ \ 1.000)$

number of data	-log(likelihood) at the true parameter	-log(maximum of likelihood)	maximum likelihood estimates		
n = 500	4.318	4.012	2.957	1.844	0.924
	4.419	2.621	2.662	2.002	0.995
	5.513	4.530	2.855	1.808	0.906
	-3.353	-4.842	3.374	2.529	1.262
	7.023	5.656	2.678	1.641	0.820
n = 50000	621.435	619.599	2.993	2.017	1.009
	623.923	623.372	2.994	2.007	1.004
	580.052	578.178	3.024	2.046	1.023
	583.369	581.362	3.014	2.041	1.021
	631.726	629.010	3.003	2.007	1.004

Table 4. A Hawkes' type non-linear process

$$\lambda(t) = \mu + \int_0^t \alpha e^{-\beta(t-s)} dN(s) + \int_0^t \int_0^s \gamma e^{-\beta(t-u)} dN(s) dN(u),$$

where

$$(\mu, \alpha, \beta, \gamma) = (0.550, 0.850, 4.750, 0.350)$$

number of data	-log(likelihood at the true parameter)	-log(maximum of likelihood)	maximum likelihood estimates			
n = 500	591.036	590.325	0.564	0.476	5.682	0.677
	458.429	452.586	0.677	0.507	6.166	0.859
	543.007	542.315	0.554	1.326	5.106	0.149
	548.165	545.468	0.543	1.225	3.355	-0.083
	561.440	557.988	0.600	1.393	6.998	0.226
n = 50000	56876.433	56870.358	0.560	0.778	4.758	0.372
	56632.821	56632.009	0.552	0.842	4.669	0.338
	56383.413	56382.695	0.554	0.824	4.791	0.372
	56524.242	56521.422	0.556	0.884	4.916	0.347
	56708.677	56707.395	0.551	0.804	4.653	0.360

Table 5. Bivariate linear Wold process

$$\lambda_1(t) = 2.300 + 10.100(t - t_{(1)}^{(1)}) + 4.500(t - t_{(1)}^{(2)})$$
$$\lambda_2(t) = 0.000 + 7.800(t - t_{(1)}^{(1)}) + 6.900(t - t_{(1)}^{(2)})$$

Length of the interval and numbers of data	-log(likelihood) at the true parameter	-log(maximum of likelihood)	maximum likelihood estimates		
T = 100	-424.386	-425.352	2.620	10.168	3.306
$n^{(1)} \doteqdot 500$			-0.064	8.788	6.878
$n^{(2)} \doteqdot 300$	-434.336	-437.251	2.889	10.262	3.377
			0.174	7.485	5.757
	-450.781	-454.897	2.469	10.851	4.813
			-0.134	6.339	8.894
	-424.451	-427.822	2.151	10.792	5.143
			-0.621	8.884	7.815
	-421.424	-422.401	2.231	10.475	5.120
			0.112	6.383	6.936
T = 2500	-10568.720	-10573.727	2.216	10.618	4.710
$n^{(1)} \doteqdot 12500$			0.032	8.083	6.917
$n^{(2)} \doteqdot 7500$	-10373.444	-10378.002	2.130	10.465	4.875
			-0.083	8.156	7.096
	-10426.732	-10429.884	2.223	10.153	4.736
			-0.042	8.191	7.080
	-10549.615	-10558.059	2.193	10.817	4.738
			0.026	7.866	6.808
	-10517.033	-10523.479	2.084	10.565	5.223
			-0.064	7.993	7.106

Table 6. Bivariate Hawkes' mutually exciting process

$$\lambda_1(t) = \mu_1 + \int_0^t a_{11}e^{-b_1(t-s)}dN_1(s) + \int_0^t a_{12}e^{-b_1(t-s)}dN_2(s)$$

$$\lambda_2(t) = \mu_2 + \int_0^t a_{21}e^{-b_2(t-s)}dN_1(s) + \int_0^t a_{22}e^{-b_2(t-s)}dN_2(s),$$

where

$$\begin{pmatrix} \mu_1, a_{11}, a_{12}, b_1 \\ \mu_2, a_{21}, a_{22}, b_2 \end{pmatrix} = \begin{pmatrix} 1.300, 0.500, 1.500, 2.800 \\ 2.500, 0.001, 1.400, 2.100 \end{pmatrix}$$

Length of the interval and numbers of data	-log(likelihood) at the true parameter	-log(maximum of likelihood)	maximum likelihood estimates			
$T = 100$	-1573.376	-1578.276	0.899	0.530	1.504	2.610
$n^{(1)} \fallingdotseq 650$			3.138	0.225	1.364	2.588
$n^{(2)} \fallingdotseq 750$	-1385.130	-1387.372	1.949	0.500	1.453	2.955
			2.702	0.167	1.251	2.238
	-2331.594	-2336.610	1.695	0.394	1.383	2.538
			2.519	0.066	1.169	1.648
	-1159.968	-1162.594	0.661	0.513	1.374	2.344
			2.888	-0.168	1.271	1.965
	-1380.468	-1381.538	1.150	0.580	1.486	2.791
			2.079	0.104	1.138	1.725
$T = 2500$	-36844.227	-36847.240	1.312	0.425	1.452	2.651
$n^{(1)} \fallingdotseq 16000$			2.483	0.029	1.481	2.247
$n^{(2)} \fallingdotseq 19000$	-35538.868	-35542.371	1.340	0.479	1.529	2.796
			2.539	0.058	1.393	2.209
	-37916.281	-37921.286	1.433	0.504	1.530	2.865
			2.644	-0.011	1.331	2.023
	-38305.112	-38311.857	1.144	0.429	1.438	2.531
			2.628	-0.082	1.441	2.083
	-36965.147	-36969.812	1.286	0.472	1.352	2.546
			2.508	0.004	1.381	2.071