THE ESTIMATION OF PARAMETERS IN POINT PROCESSES

Toshiharu HAYASHI

University of Osaka Prefecture

1989

訂 正

誤 $\Longrightarrow$ 正

p. 3 ℓ. 2 $\rho t - N(t)$ $\Longrightarrow$ $t - N(t)$

p. 13 ℓ. 8 $M_i(1) =, i - k_1 + 1$ $\Longrightarrow$ $M_i(1) = i - k_1 + 1$

p. 20 ℓ. 2 $\sum_{n=0}^{[t_{k_m}]-1}$ $\Longrightarrow$ $\sum_{n=0}^{[T_{k_m}]-1}$

p. 21 ℓ. 13, p. 22 ℓ. 3 (2カ所) $f(t)h(Y(t))$ $\Longrightarrow$ $f'(t)h(Y(t))$

p. 25 ℓ. 10

$$\begin{pmatrix} 0 & 0 & \rho\beta\psi'(Y(t,\theta))\psi(Y(t,\theta))/T^2 \\ & 0 & \rho Y(t,\theta)\psi'(Y(t,\theta))\psi(Y(t,\theta))/T^2 \\ sym. & & \{2\rho\beta\psi'(Y(t,\theta))\psi(Y(t,\theta)) \cdot t + \psi(Y(t,\theta))^2\}/T^3 \end{pmatrix}$$

$\Longrightarrow$

$$\begin{pmatrix} 0 & 0 & \rho\beta\psi'(\beta Y(t,\theta))/T^2 \\ & 0 & \rho Y(t,\theta)\psi'(\beta Y(t,\theta))/T^2 \\ sym. & & \{2\rho\beta\psi'(\beta Y(t,\theta)) \cdot t + \psi(\beta Y(t,\theta))\}/T^3 \end{pmatrix}$$

p. 27 ℓ. 7 $\dfrac{P^{(T)}_{\theta+\Phi_T(\theta)u}}{P^{(T)}_\theta}$ $\Longrightarrow$ $\dfrac{dP^{(T)}_{\theta+\Phi_T(\theta)u}}{dP^{(T)}_\theta}$

p. 36 ℓ. 8−10, p. 37 ℓ. 4 (4カ所) $\chi(X(T) \le 0)$ $\Longrightarrow$ $\chi(X(t) \le 0)$

p. 37 ℓ. 2, 4 (2カ所) $\chi(X(T) > 0)$ $\Longrightarrow$ $\chi(X(t) > 0)$

p. 65 the last line

asymptotic variance $\Longrightarrow$ reciprocal of asymptotic variance

Abstract

In the present thesis, we consider two problems, namely, the maximum likelihood estimation in self-correcting point processes and the robust estimation in the Poisson processes.

In Chapters 2–4, we treat the self-correcting point process $N(\cdot)$ with the conditional intensity $\rho\,\psi(\rho t - N(t))$, where ρ is a positive constant and $\psi(\cdot)$ is a function satisfying suitable conditions. The process $N(\cdot)$ corrects its conditional intensity through the function $\psi(\cdot)$ so that the absolute value of $\rho t - N(t)$ may tend to become small. From this self-correcting mechanism, we see that a skeleton process $\{n - N(n/\rho)\}_{n=0,1,2,\dots}$ is the ergodic (i.e. positive recurrent) Markov chain, which implies the law of large numbers for functionals of the Markov process $\rho t - N(t)$.

We consider a self-correcting point process with a parametrized conditional intensity and investigate asymptotic properties of the maximum likelihood estimator (MLE) in Chapter 3. Using the above law of large numbers, we see that a standardized information matrix converges to a positive definite matrix. Since we can not give the explicit representation of the MLE, we examine the likelihood ratio and show that the family of the measures induced by the self-correcting processes is locally asymptotically normal. This property is closely related to the likelihood ratio and plays an important role in investigating asymptotic behavior of the MLE. From these results, we obtain consistency and asymptotic normality of the MLE.

In Chapter 4, we treat a self-correcting point process whose conditional intensity has only two levels θ_1 and θ_2 ($0 < \theta_1 < 1 < \theta_2$). We explicitly give the log likelihood and the MLE of the parameter $\theta = (\theta_1, \theta_2)'$, where v' denotes the transposition of a vector v. Moreover we obtain asymptotic normality of the MLE and express its asymptotic variance

by θ_1 and θ_2.

In the last chapter, we discuss the robust estimation problem in the Poisson processes with a periodic intensity. It is well known that the MLE is a solution of the likelihood equation and is asymptotically normal and efficient under some regularity conditions. However, if the observation is contaminated by noises, the MLE is not always an appropriate estimator. So we should construct robust estimators in the sense that high efficiency is kept even if the observation is contaminated. For this purpose we treat M-estimators which are solutions of estimating equations. We investigate their asymptotic properties such as consistency in a sense and asymptotic normality. Furthermore we obtain the M-estimator which has the minimax asymptotic variance under suitable conditions.

Table of Contents

Chapter 1. Introduction

We consider a series of events which discretely occur on the time axis $[0, \infty)$, where two or more events do not occur at the same time. Let $N(t)$ be the number of events occurring in the time interval $[0, t)$ for each $t \in [0, \infty)$. The process $N(\cdot)$ is called a *point process.* Denoting occurrence time of i-th event by τ_i $(i = 1, 2, 3, \dots)$, we see that $N(t) = \sum_{i=1}^{\infty} \chi(\tau_i < t)$, where $\chi(\cdot)$ is the indicator. Clearly, the sample path of the process $N(\cdot)$ is piece-wise constant and has unit jumps. Assume that the point processes considered here have a *conditional intensity* $\lambda(t|\mathcal{F}_t)$ which is defined by

$$
\begin{aligned}
\lambda(t|\mathcal{F}_t) &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) \geq 1 | \mathcal{F}_t) \\
&= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(N(t + \Delta t) - N(t) = 1 | \mathcal{F}_t),
\end{aligned}
$$

where $\mathcal{F}_t$ is the σ-field generated by $\{N(s); 0 \leq s \leq t\}$ (see e.g. Daley and Vere-Jones (1972), Rubin (1972) and Brillinger (1978)). It is well known that the conditional intensity $\lambda(t|\mathcal{F}_t)$ determines the distribution of the point process. If the point process $N(\cdot)$ has the Markov property, its conditional intensity depends only on the count $N(t)$ up to time t but is independent of behavior of $N(\cdot)$ before time t and vice versa. Thus in order that the point process has the Markov property, it is necessary and sufficient that its conditional intensity $\lambda(t|\mathcal{F}_t)$ is equal to its count conditional intensity $\lambda(t|N(t))$ $(= E[\lambda(t|\mathcal{F}_t)|N(t)])$. The pure birth process with the instantaneous birth rate λ (> 0) is a typical example of the Markov point processes and its conditional intensity is $\lambda N(t)$. In the present thesis, we treat only Markov point processes, where we can identify the conditional intensity $\lambda(t|\mathcal{F}_t)$ with the count conditional intensity $\lambda(t|N(t))$.

We consider point processes with the conditional intensity

$$\lambda(t|N(t)) = \rho\psi(\rho t - N(t)),$$

where ρ is a positive constant and $\psi(\cdot)$ is a function satisfying the following conditions:

(C1) for any $x \in \mathbf{R}$ (the real line),

$$0 \leq \psi(x) < \infty,$$

(C2) there exists a positive constant c_0 such that

$$\psi(x) \geq c_0 \qquad \text{for all } x > 0,$$

(C3) $$\liminf_{x\to\infty} \psi(x) > 1 \quad \text{and} \quad \limsup_{x\to-\infty} \psi(x) < 1.$$

For a sufficiently large $t > 0$, if few events occur up to time t, $x = \rho t - N(t)$ is sufficiently large. Then the intensity $\rho\psi(x)$ is larger than ρ by the condition (C3). We can interpret ρ as the rate of increase of ρt and the intensity $\rho\psi(x)$ as the mean rate of increase of $N(\cdot)$ under the condition that $\rho t - N(t) = x$. Thus it is expected that the forthcoming $x = \rho(t+s) - N(t+s)$ $(s > 0)$ is smaller than the present $x = \rho t - N(t)$ if it is sufficiently large. On the other hand, if a considerable number of events occur up to time t, $x = \rho t - N(t)$ is negative and its absolute value is sufficiently large. Then it is expected that the forthcoming x is larger than the present x. As stated above, the process $N(\cdot)$ corrects its conditional intensity so that the absolute value of $\rho t - N(t)$ may tend to become small when its absolute value is sufficiently large. Isham and Westcott (1979) introduced such point processes and called them *self-correcting point processes.* They showed the following results about the mean and the r-th moment of the self-correcting point process $N(\cdot)$:

$$\limsup_{t\to\infty} \; |E[N(t)] - \rho t| \; < \infty$$

$$\limsup_{t\to\infty} E|N(t) - E[N(t)]|^r < \infty \quad (r > 0).$$

Vere-Jones and Ogata (1984) obtained a version of the law of large numbers for the process $\rho t - N(t)$ under the condition that the process $N(\cdot)$ have an exponential form intensity, that is, $\lambda(t|N(t)) = \exp\{\alpha + \beta(t - N(t))\}$, where α and β are constants. Ogata and Vere-Jones (1984) treated a self-correcting point process with a parametrized intensity $\exp\{\alpha+\beta(t-N(t))+\gamma t/T\}$ (T is the observation time) and obtained asymptotic normality of the MLE's of the parameters α, β and γ. We extend their results to self-correcting point processes with more general intensity.

In Chapter 2, we shall show a version of the law of large numbers for the process $\rho t - N(t)$ under the conditions (C1)–(C3) and

(C4) for any $K > 0$, there exists an $M > 0$ such that

$$\psi(x) \leq M \qquad \text{for } |x| \leq K.$$

The condition (C4) holds for ordinary functions (e.g. piece-wise continuous functions).

We can choose the scale of the time axis so that $\rho = 1$, which is verified as follows. Let $\widetilde{N}(t) = N(ct)$, where c is a positive constant. Then the intensity of the process $\widetilde{N}(t)$ is given by

$$\lambda(t|\widetilde{N}(t)) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P(\widetilde{N}(t+\Delta t) - \widetilde{N}(t) = 1 | \widetilde{N}(t)).$$

It is easy to check that

$$\begin{aligned}\lambda(t|\widetilde{N}(t)) &= c \lim_{\Delta t \downarrow 0} \frac{1}{c\Delta t} P(N(ct + c\Delta t) - N(ct) = 1 | N(ct)) \\ &= c\lambda(ct|N(ct)) \\ &= c\rho\psi(c\rho t - N(ct)) \\ &= c\rho\psi(c\rho t - \widetilde{N}(t)).\end{aligned}$$

For $c = 1/\rho$, the intensity of the process $\widetilde{N}(\cdot)$ is $\psi(t - \widetilde{N}(t))$. Thus, without loss of generality, we may suppose that the conditional intensity of the process $N(\cdot)$ is

$$\lambda(t|N(t)) = \psi(t - N(t)) = \psi(X(t)),$$

where

$$X(t) = t - N(t).$$

The process $X(\cdot)$ inherits the Markov property from the self-correcting point process $N(\cdot)$.

In Section 2.2, we consider the Markov chain $\{X(n)\}_{n=0,1,2,\ldots}$ and show that this chain is irreducible, aperiodic and positive recurrent (i.e. ergodic). Hence the Markov chain $\{X(n)\}$ has the invariant distribution $\{\pi_j\}$. Vere-Jones and Ogata (1984) used Theorem 4 in Tweedie (1983a) to show that the Markov chain $\{X(n)\}$ is ergodic under the condition that the process $N(\cdot)$ have an exponential form intensity. Their idea is very useful here as well. To check the conditions of Theorem 4 in Tweedie (1983a), upper bounds of a conditional expectation $E[b^{-X(n+1)}|X(n) = i]$ are required, where n and i are integers and b is a positive constant. We give these upper bounds in Lemmas 2.3–2.5. Furthermore, we obtain existence of an exponential form expectation with respect to the invariant distribution $\{\pi_j\}$ of the Markov chain $\{X(n)\}$.

In Section 2.3, we show that the weighted time average of a functional of the Markov process $X(\cdot)$ converges in probability to its mean on the sample space, that is,

$$\int_0^T w(t,T)h(X(t))\,dt \to \sum_{j\in S} \pi_j E\left[\int_0^1 h(X(t))\,dt \,\middle|\, X(0) = j\right]$$

in probability as $T \to \infty$, where $w(t,T)$ is a weight, $h(\cdot)$ is a function, S is the state space of the Markov chain $\{X(n)\}$ and $\{\pi_j\}$ is the invariant distribution of $\{X(n)\}$ (see Theorem 2.9). This is a version of the law of large numbers.

In Chapter 3, we consider a self-correcting point process with a parametrized conditional intensity

$$\lambda(t,\theta) = \rho\,\psi(\beta\{\rho t - N(t) + \alpha\}),$$

where the parameter $\theta = (\alpha, \beta, \rho)'$ (v' denotes the transposition of a vector v) belongs to the parameter space $\Theta = \mathbf{R} \times (0, M_\beta) \times (0, \infty)$ and M_β (> 1) is a constant. The parameters α, β and ρ are respectively related to the origin of the time axis, sensitivity of the self-correcting and the scale of the time axis. We shall investigate asymptotic properties of the maximum likelihood estimator (MLE) of the parameter θ (but we do not discuss the estimation problem of the function $\psi(\cdot)$ or we treat it as a known function). Unfortunately, we can not give the explicit representation of the MLE. Thus we do not examine the MLE but asymptotic behavior of the likelihood ratio which is normalized by a matrix associated with the information matrix. In Section 3.2, we show that a standardized information matrix converges to a positive definite matrix by using the law of large numbers shown in Chapter 2. In Section 3.3, we shall review definition and some basic properties of local asymptotic normality. Moreover we obtain local asymptotic normality of the family of the measures induced by the self-correcting processes. This property is closely related to the likelihood ratio and plays an important role in investigating asymptotic behavior of the MLE. In Section 3.4, we obtain that the MLE is consistent and asymptotically normal under suitable conditions.

In Chapter 4, we treat a point process with a very simple conditional intensity

$$\lambda(t|N(t)) = \rho\psi(\rho t - N(t) + \alpha)$$
$$= \begin{cases} \rho\theta_1 & \text{if } \rho t - N(t) + \alpha \le 0, \\ \rho\theta_2 & \text{if } \rho t - N(t) + \alpha > 0, \end{cases}$$

where θ_1 and θ_2 are constants satisfying that $0 < \theta_1 < 1 < \theta_2$. We easily check the conditions (C1)–(C4) for the function $\psi(x) = \theta_1$ if $x \le 0$ and $= \theta_2$ if $x > 0$. Hereafter we call this process $N(\cdot)$ a *simple self-correcting point process.*

We concentrate our interest on estimation of the intensity levels and treat α and ρ as known constants. We can choose the location and the scale of the time axis so that $\alpha = 0$ and $\rho = 1$. Hence, without loss of generality, we may assume that the conditional intensity of the process $N(\cdot)$ is

$$\lambda(t|N(t)) = \psi(X(t)) = \begin{cases} \theta_1 & \text{if } X(t) \leq 0, \\ \theta_2 & \text{if } X(t) > 0, \end{cases}$$

where $X(t) = t - N(t)$ and $\theta = (\theta_1, \theta_2)'$. In Section 4.2, we explicitly give the log likelihood and the MLE of the parameter θ. In Section 4.3, we calculate the invariant distribution of the Markov chain $\{X(n)\}$ and obtain that

$$\frac{|D_1(T)|}{T} \rightarrow \frac{\theta_2 - 1}{\theta_2 - \theta_1}$$

in probability as $T \to \infty$ from the law of the large numbers shown in Chapter 2, where $|D_1(T)|$ denotes the measure of the region $D_1(T) = \{t \in [0,T]; X(t) \leq 0\}$. From the above result, a standardized information matrix converges to a positive definite matrix which is represented by θ_1 and θ_2. In Section 4.4, we show that the MLE is asymptotically normal and explicitly give its asymptotic variance. Furthermore we obtain that the family of the measures induced by simple self-correcting point processes is locally asymptotically normal.

In Chapter 5, we consider a Poisson process $N(t)$ with a parametrized intensity $\lambda(t,\theta)$, where the parameter θ belongs to a bounded open interval Θ of $\mathbf{R}$. Note that the intensity $\lambda(t,\theta)$ is deterministic. The log likelihood function based on the observation $(N(t); 0 \leq t \leq T)$ up to time T is given by

$$\ell(T,\theta) = \int_0^T \log \lambda(t,\theta) dN(t) - \int_0^T \lambda(t,\theta) dt\,.$$

The MLE maximizes the log likelihood $\ell(T,\theta)$ and is a solution of the likelihood equation

$$\int_0^T \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)} dN(t) - \int_0^T \dot{\lambda}(t,\theta)dt = 0$$

under some regularity conditions, where $\dot{\lambda}(t,\theta)$ is the derivative of $\lambda(t,\theta)$ with respect to θ. Moreover it is well known that the MLE is consistent, asymptotically normal and efficient (see e.g. Kutoyants (1984)).

If the artificial model does not sufficiently reflect the generation mechanism of the data or if the data are contaminated by noises, the true intensity $\mu(t)$ of the process $N(\cdot)$ may not belong to the parametric model $\{\lambda(t,\theta); \theta \in \Theta\}$. In such circumstances, the MLE is not always an appropriate estimator of the parameter θ.

For example, we shall consider the following point process $N(\cdot)$. Let the true intensity $\mu(t)$ be $(1-\varepsilon)f(t)+\varepsilon c(t)$ and $\lambda(t,\theta) = f(t-\theta)$, where ε denotes the rate of contamination, $f(\cdot)$ and $c(\cdot)$ are periodic even functions with the period 1 and the phase parameter θ belongs to $\Theta = (-\frac{1}{2}, \frac{1}{2})$. If the data are not contaminated (i.e. $\varepsilon = 0$), the MLE is a very good estimator. However its asymptotic efficiency diminishes for $\varepsilon > 0$. So our purpose is to construct robust estimators in the sense that high efficiency is kept even if the data are contaminated.

To get robust estimators, we investigate M-estimators which are solutions of generalized likelihood equations. In Section 5.2, we obtain that our M-estimators are consistent in a sense and asymptotic normal even if the true intensity $\mu(t)$ does not belong to the parametric model $\{\lambda(t,\theta); \theta \in \Theta\}$. In Section 5.3, we illustrate how to get a robust M-estimator for the above model with an unknown phase parameter. Moreover we show that our robust estimator has the minimax asymptotic variance provided that the true intensity belongs to a suitable class.

Chapter 2. Law of Large Numbers in Self-Correcting Point Processes

2.1. Introduction

We consider a probability space $(\Omega, \mathcal{B}, \mathcal{P})$. For $i = 1, 2, 3, \ldots$ and $\omega \in \Omega$, we denote occurrence time of i-th event by $\tau_i(\omega)$ (≥ 0), where the events discretely occur on the time axis $[0, \infty)$ and two or more events do not occur at the same time. Let

$$N(t, \omega) = \sum_{i=1}^{\infty} \chi(\tau_i(\omega) < t),$$

where $\chi(\cdot)$ is the indicator. Then $N(t, \omega)$ is called a point process and denotes the number of events occurring in the time interval $[0, t)$. We usually abbreviate $N(t, \omega)$ as $N(t)$.

In the present chapter, we consider the self-correcting point process $N(\cdot)$ with the conditional intensity

$$\lambda(t|\mathcal{F}_t) = \lambda(t|N(t)) = \rho\,\psi(\rho t - N(t)), \tag{1.1}$$

where ρ is a positive constant and $\psi(\cdot)$ is a function satisfying the conditions (C1)–(C4) in Chapter 1. We can choose the scale of the time axis so that $\rho = 1$. Hence, without loss of generality, we may assume that the conditional intensity of the process $N(\cdot)$ is

$$\lambda(t|N(t)) = \psi(t - N(t)) = \psi(X(t)), \tag{1.2}$$

where

$$X(t) = t - N(t). \tag{1.3}$$

The process $X(\cdot)$ inherits the Markov property from the self-correcting point process $N(\cdot)$.

Under the condition that the process $N(\cdot)$ have an exponential form intensity, that is, $\psi(x) = \exp\{x\}$, Vere-Jones and Ogata (1984) obtained the law of large numbers for the Markov process $X(\cdot)$. We shall show the law of large numbers under more general conditions, namely, (C1)–(C4). For this purpose, we shall give upper bounds of conditional expectations of the Markov chain $\{X(n)\}_{n=0,1,2,\ldots}$ and show that this chain is irreducible, aperiodic and positive recurrent (i.e. ergodic) in Section 2.2. Vere-Jones and Ogata (1984) used Theorem 4 in Tweedie (1983a) to show that the Markov chain $\{X(n)\}$ is ergodic. Their idea is very useful here as well. In Section 2.3, we show that the weighted time average of a functional of the Markov process $X(\cdot)$ converges in probability to its mean on the sample space under the conditions (C1)–(C4). This is a version of the law of large numbers.

2.2. Ergodicity of the Markov chain $\{X(n)\}$

For fixed $x \in \mathbf{R}$, we consider a point process $N_x(\cdot)$ with the conditional intensity

$$\lambda(t|N_x(t)) = \psi(t + x - N_x(t)).$$

Then we obtain the following lemma because the distribution of a point process is completely specified by its conditional intensity (see e.g. Bremaud (1981)).

Lemma 2.1. *For any $\tau \geq 0$, any $s \geq 0$ and any $n = 0, 1, 2, \ldots,$*

$$P(N(\tau + s) - N(\tau) = n|X(\tau) = x) = P(N_x(s) = n)$$

provided that $P(X(\tau) = x) > 0$, where $N(\cdot)$ is the self-correcting point process with the conditional intensity (1.2) and $X(\cdot) = t - N(t)$.

From the above lemma, the transition probability $P(X(n+1) = j|X(n) = i)$ ($= P(N_i(1) = i + 1 - j)$) is independent of n, which implies that the Markov chain $\{X(n)\}$

has the stationary transition probability p_{ij} $(= P(X(n+1) = j|X(n) = i))$. Since $X(n) - X(n+1) = N(n+1) - N(n) - 1 \geq -1$, we have that $p_{ij} = 0$ for $j \geq i+2$. Therefore the transition matrix of the Markov chain $\{X(n)\}$ is written as

$$
(2.1) \qquad \begin{pmatrix}
\cdots & \ddots & & & & & \\
\cdots & p_{-2,-2} & p_{-2,-1} & & & & 0 \\
\cdots & p_{-1,-2} & p_{-1,-1} & p_{-1,0} & & & \\
\cdots & p_{0,-2} & p_{0,-1} & p_{0,0} & p_{0,1} & & \\
\cdots & p_{1,-1} & p_{1,-1} & p_{1,0} & p_{1,1} & p_{1,2} & \\
\cdots & p_{2,-2} & p_{2,-1} & p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} \\
\cdots & & \cdots & & \cdots & & \ddots
\end{pmatrix}.
$$

If for some $x_1 \in \mathbf{R}$, the integral $\int_0^1 \psi(t+x_1)dt$ of the conditional intensity is equal to 0 (from the condition (C2), such an x_1 is not greater than -1), we have that $P(N_{x_1}(1) = 0) = \exp\{-\int_0^1 \psi(t+x_1)dt\} = 1$, which implies $p_{ij} = 0$ for $i > x_1 \geq j$. Let

$$k_0 = -\max\left\{k \in \mathbf{Z}\,;\ \text{there exists an } x_1 \geq k \text{ such that } \int_0^1 \psi(t+x_1)\,dt = 0\right\},$$

where $\mathbf{Z}$ is the set of all integers. If for any $x \in \mathbf{R}$, $\int_0^1 \psi(t+x)\,dt > 0$, put $k_0 = \infty$. It is clear that k_0 is a positive integer (or infinity). We easily check that

$$p_{i,i-1} > 0 \quad \text{for } i > -k_0 + 1$$

and

$$p_{ij} = 0 \quad \text{for } i > -k_0 \geq j.$$

From the conditions (C2) and (C4), we have that $p_{i,i+1} = \exp\left\{-\int_0^1 \psi(t+i)dt\right\} > 0$ for all i and that $p_{0,0} = \int_0^1 \psi(s)\exp\left\{-\int_0^s \psi(t)dt - \int_s^1 \psi(t-1)dt\right\}ds > 0$. Consequently we obtain the following lemma.

Lemma 2.2. *The discrete-time process $\{X(n)\}_{n=0,1,2,\dots}$ is an irreducible and aperiodic Markov chain defined on the state space $S = \{i \in \mathbf{Z}\,;\, i > -k_0\}$, where $\mathbf{Z}$ is the set of all integers and k_0 is given above.*

To show that the Markov chain $\{X(n)\}$ is ergodic, the following lemma is required.

Lemma 2.3. *Let S be the state space of the Markov chain $\{X(n)\}$ and p_{ij}'s are its transition probabilities. Then we have that for any $i \in S$ and any $b \geq 1$,*

$$\begin{aligned} E[b^{-X(n+1)}|X(n) = i] &= \sum_{j \in S} p_{ij} b^{-j} \\ &\leq b^{-i-1} \exp\{\lambda_i (b-1)\}, \end{aligned}$$

where $\lambda_i = \sup\{\psi(x)\,;\, x \leq i+1\}$ ($< \infty$ by the conditions (C3) and (C4)).

Proof. From Lemma 2.1, we have that

$$\begin{aligned} E[b^{-X(n+1)}|X(n) = i] &= b^{-i} E[b^{-X(n+1)+X(n)}|X(n) = i] \\ &= b^{-i-1} E[b^{N(n+1)-N(n)}|X(n) = i] \\ &= b^{-i-1} E[b^{N_i(1)}], \end{aligned} \tag{2.2}$$

where $N_i(\cdot)$ is a point process with the conditional intensity $\psi(t + i - N_i(t))$. We easily see that $\psi(t + i - N_i(t)) \leq \lambda_i$ for $0 \leq t \leq 1$. The point process $M_i(\cdot)$ with the constant conditional intensity λ_i is the homogeneous Poisson process with the parameter λ_i. Since the conditional intensity of the process $N_i(\cdot)$ is dominated by one of the process $M_i(\cdot)$, we have that

$$\begin{aligned} E[b^{N_i(1)}] &\leq E[b^{M_i(1)}] \\ &= \exp\{\lambda_i(b-1)\} \end{aligned}$$

(see Deng (1985) for the relevant comparison theorem). From (2.2) and the above inequality, we obtain the conclusion of the present lemma.

We consider the following equation

$$\exp\{c(x-1)\} = x, \tag{2.3}$$

where c is a positive constant. It is clear that $x = 1$ is a solution of the above equation for every $c(>0)$. For $c \neq 1$, the above equation has exactly two solutions. Denoting the solution except 1 by x_0, it is easy to check that for $c > 1$, $0 < x_0 < 1$ and that for $0 < c < 1$, $x_0 > 1$. Let b_0 be the solution (except 1) of the equation (2.3) for $c = \liminf_{x\to\infty} \psi(x)$, b_1 be the solution for $c = \limsup_{x\to-\infty} \psi(x)$ and

$$a_0 = \min\left\{\frac{1}{b_0}, b_1\right\}. \tag{2.4}$$

If $\liminf_{x\to\infty} \psi(x) = \infty$, then put $b_0 = 0$ and if $\limsup_{x\to-\infty} \psi(x) = 0$, then $b_1 = \infty$. Then we can easily check that $0 < b_0 < 1$ and $a_0 > 1$ by the condition (C3). We will use a_0 and b_0 in Lemma 2.5 and the following lemma, respectively.

Lemma 2.4. *For any $b \in (b_0, 1)$, there exist an $\eta > 0$ and an $i_0 \geq 1$ such that for any $i > i_0$,*

$$\begin{aligned} E[\, b^{-X(n+1)} \,|\, X(n) = i\,] &= \sum_{j\in S} p_{ij} b^{-j} \\ &< b^{-i+\eta} \end{aligned}$$

Proof. We easily see that for any $b \in (b_0, 1)$, $\exp\{c(b-1)\} < b$, where $c = \liminf_{x\to\infty} \psi(x)$ (> 1 by the condition (C3)). Hence we can find a $\nu \in (1, c)$ such that $\exp\{\nu(b-1)\} < b$. Since $\nu < c (= \liminf_{x\to\infty} \psi(x)\,)$, there exists a positive integer k_1 such that $\psi(x) \geq \nu$ for all

$x \geq k_1$. Let $M_i(\cdot)$ be a point process with the conditional intensity

$$\nu \cdot \chi(M_i(t) \leq i - k_1) = \begin{cases} \nu & \text{if } M_i(t) \leq i - k_1\,, \\ 0 & \text{if } M_i(t) > i - k_1\,, \end{cases}$$

where $\chi(\cdot)$ is the indicator. It is easy to check that $\psi(t + i - m) \geq \nu \cdot \chi(m \leq i - k_1)$ for $m = 0, 1, 2, \ldots$ and $0 \leq t \leq 1$. We then obtain that

$$E[b^{N_i(1)}] \leq E[b^{M_i(1)}] \tag{2.5}$$

because b^x is decreasing in x (see Deng (1985)).

We have that

$$E[b^{M_i(1)}] = \sum_{m=0}^{i-k_1} b^m \frac{\nu^m}{m!} e^{-\nu} + b^{i-k_1+1} P\{M_i(1) =, i - k_1 + 1\}$$
$$\to \exp\{\nu(b-1)\} \quad \text{as} \quad i \to \infty\,.$$

Since $\exp\{\nu(b-1)\} < b$, there exists an $\eta > 0$ such that $\exp\{\nu(b-1)\} < b^{1+\eta}$ $(< b)$. Moreover we can find an $i_0 > 0$ such that

$$E[b^{M_i(1)}] < b^{1+\eta} \qquad \text{for all } i > i_0\,.$$

From (2.2), (2.5) and the above inequality, we obtain the required inequality.

The following lemma is also required to show that the Markov chain $\{X(n)\}$ is ergodic.

Lemma 2.5. *For any $b \geq 1$ and any $i \geq 0$,*

$$\sum_{j \in S_1} p_{ij} b^{-j} \leq \exp\{\lambda_0(b-1)\}\,,$$

where $S_1 = \{j \in S; j \leq 0\}$ and $\lambda_0 = \sup\{\psi(x)\,; x \leq 1\}$.

Proof. We have that

$$\sum_{j \in S_1} p_{ij} b^{-j} = E[\chi(X(n+1) \leq 0) b^{-X(n+1)} | X(n) = i]$$
$$= E[\chi(N_i(1) \geq i + 1) b^{N_i(1)-i-1}],$$

where $N_i(\cdot)$ is a point process with the conditional intensity $\psi(t+i-N_i(t))$. Let $M_i(\cdot)$ be a point process with the conditional intensity

$$\lambda(t|M_i(t)) = \begin{cases} \lambda_i & \text{if } M_i(t) \le i, \\ \lambda_0 & \text{if } M_i(t) \ge i+1, \end{cases}$$

where $\lambda_i = \sup\{\psi(x); x \le i+1\}$. Since the conditional intensity of the process $N_i(\cdot)$ is dominated by one of the process $M_i(\cdot)$, we see that

$$E[\chi(N_i(1) \ge i+1)b^{N_i(1)-i-1}] \le E[\chi(M_i(1) \ge i+1)b^{M_i(1)-i-1}].$$

Furthermore, we have that

$$\begin{aligned}
&E[\chi(M_i(1) \ge i+1)b^{M_i(1)-i-1}] \\
&= \sum_{k=0}^{\infty} b^k P\{M_i(1) = i+1+k\} \\
&= \sum_{k=0}^{\infty} b^k \int_0^1 \frac{(\lambda_i t)^i}{i!} \exp\{-\lambda_i t\} \cdot \lambda_i \cdot \frac{\{\lambda_0(1-t)\}^k}{k!} \exp\{-\lambda_0(1-t)\} dt \\
&= \frac{\lambda_i^{\,i+1}}{i!} \exp\{\lambda_0(b-1)\} \int_0^1 t^i \exp\{-[\lambda_i + \lambda_0(b-1)]t\} dt \\
&\le \frac{\lambda_i^{\,i+1}}{i!} \exp\{\lambda_0(b-1)\} \frac{1}{[\lambda_i + \lambda_0(b-1)]^{i+1}} \Gamma(i+1) \\
&\le \exp\{\lambda_0(b-1)\},
\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function. Hence we obtain the required inequality.

Using Lemmas 2.3, 2.4 and 2.5, we shall show that the Markov chain $\{X(n)\}$ is ergodic. From Theorem 4 (i) in Tweedie (1983a), it is sufficient that there exist a finite subset A of the state space S, a non-negative sequence $\{g_j\}$ and an $\varepsilon > 0$ such that

(a) $$\sup_{i \in A} \left\{ \sum_{j \in S} p_{ij} g_j \right\} < \infty$$

and for any $i \in (S - A)$,

(b)
$$\sum_{j \in S} p_{ij} g_j < g_i - \varepsilon.$$

Let $g_j = a^{|j|}$ for some $a \in (1, a_0)$, where a_0 is given by (2.4). We shall seek for a finite subset A satisfying (a) and (b).

For $i \geq 0$, we see that

$$\begin{aligned} \sum_{j \in S} p_{ij} g_j &= \sum_{j \in S} p_{ij} a^{|j|} \\ &\leq \sum_{j \in S_1} p_{ij} a^{-j} + \sum_{j \in S} p_{ij} a^{j} . \end{aligned}$$

From Lemmas 2.4 and 2.5, there exists an $i_1 > i_0$ such that for any $i > i_1$ and sufficiently small $\varepsilon > 0$,

$$\begin{aligned} \sum_{j \in S} p_{ij} g_j &\leq \exp\{\lambda_0 (a-1)\} + a^{i-\eta} \\ &< g_i - \varepsilon, \end{aligned}$$

where i_0 (≥ 1) and η (> 0)are given in Lemma 2.4.

For $i \leq -1$, we see that

$$\sum_{j \in S} p_{ij} g_j = \sum_{j \in S} p_{ij} a^{-j}$$

because $p_{ij} = 0$ for $j \geq 1$ ($\geq i + 2$). By Lemma 2.3 we have that for any $i \in S$,

$$\sum_{j \in S} p_{ij} a^{-j} \leq a^{-i-1} \exp\{\lambda_i (a-1)\} ,$$

where $\lambda_i = \sup\{\psi(x) ; x \leq i + 1\}$. Since $1 < a < a_0$, we easily see that $\exp\{c(a-1)\} < a$ for $c = \limsup_{x \to -\infty} \psi(x)$. Hence we can find an $\varepsilon > 0$ and a $\nu_0 \in (c, 1)$ such that for any $\nu < \nu_0$, $\exp\{\nu(a-1)\} < a - \varepsilon$, which implies that there exists an $i_2 > 0$ such that for any $i < -i_2$,

$$\exp\{\lambda_i (a-1)\} < a - \varepsilon.$$

Thus we obtain that for any $i < -i_2$,

$$\sum_{j \in S} p_{ij} g_j < a^{-i} - a^{-i-1}\varepsilon$$
$$\leq g_i - \varepsilon.$$

Putting $A = \{i \in S\,; -i_2 \leq i \leq i_1\}$, the condition (b) holds as we have seen. We easily see the condition (a) by finiteness of A. Hence we have the following theorem.

Theorem 2.6. *The Markov chain $\{X(n)\}$ is ergodic (i.e. positive recurrent).*

Vere-Jones and Ogata (1984) thought out the idea of using Tweedie's theorem. They obtained the same result for the self-correcting point processes with the exponential form conditional intensity.

From the above theorem, the Markov chain $\{X(n)\}$ has the invariant (stationary) distribution $\{\pi_j\}$. Moreover we can show the following theorem about the moment of the Markov chain $\{X(n)\}$.

Theorem 2.7. *For any $a \in (1, a_0)$,*

$$\sum_{j \in S} \pi_j a^{|j|} < \infty,$$

where a_0 is given by (2.4).

Proof. We easily see that

$$\sum_{j \in S} \pi_j a^{|j|} \leq \sum_{j \in S} \pi_j a^j + \sum_{j \in S} \pi_j a^{-j}.$$

By Lemmas 2.3, 2.4 and Theorem 1 in Tweedie (1983b), we obtain that the first and the second term of the right-hand side are finite.

2.3. Law of large numbers

The following lemma is an extension of the L_2-ergodic theorem and is the same as Lemma 3 in Vere-Jones and Ogata (1984) with a little modification of the condition (i).

Lemma 2.8. *Let $\{U(n)\}$ be a stationary and ergodic process with finite second moments and $\{w_{n,k}\}$ $(n = 0, 1, \ldots, m_k\,; k = 1, 2, \ldots)$ be a sequence of weights satisfying*

(i). $w_{n,k} \geq 0$ and $\sum_{n=0}^{m_k} w_{n,k} \to 1$ as $k \to \infty$,

(ii). $w_{n,k} \leq w_{n+1,k}$ *for* $n = 0, 1, \ldots, m_k - 1$,

(iii). $w_{m_k,k} \to 0$ *as* $k \to \infty$.

Then

$$P\text{-}\lim_{k\to\infty} \sum_{n=0}^{m_k} w_{n,k} U(n) = E[U(0)]\,,$$

*where P-*lim *denotes the convergence in probability.*

This lemma is obtained by the similar way of showing the L_2-ergodic theorem (e.g. Theorem 2.1 in Billingsley (1965)). So we omit the proof.

As we have seen in Section 2.2, the Markov chain $\{X(n)\}$ is ergodic (i.e. positive recurrent). When its initial distribution, namely, the distribution of $X(0)$ is the invariant distribution $\{\pi_j\}$, the Markov chain $\{X(n)\}$ is stationary. Let $h(\cdot)$ be a function satisfying that for some $a \in (1, a_0)$,

$$|h(x)| \leq a^{|x|/2} \qquad (x \in \mathbf{R}), \tag{3.1}$$

where a_0 is defined by (2.4). Then, from Theorem 2.7, we see that for $n = 0, 1, 2, \ldots$,

$$\begin{aligned} E[h(X(n))^2] &= \sum \pi_j h(j)^2 \\ &\leq \sum \pi_j a^{|j|} \\ &< \infty. \end{aligned}$$

Thus we can use Lemma 2.8 for the process $\{h(X(n))\}_{n=0,1,2,\ldots}$.

For any $s \in [0,1]$, we consider processes $\{X(n+s)\}_{n=0,1,2,\ldots}$, where $X(n+s) = (n+s) - N(n+s)$. These processes $\{X(n+s)\}_{n=0,1,2,\ldots}$ are also ergodic and stationary (when $X(0)$ is distributed as the invariant distribution $\{\pi_j\}$). We can easily check that for any $n = 0,1,2,\ldots$ and any $s \in [0,1]$,

$$\begin{aligned} X(n+s) &\leq n+1-N(n) \\ &= X(n)+1 \\ &\leq |X(n)|+1 \end{aligned}$$

and that

$$\begin{aligned} -X(n+s) &\leq -(n-N(n+1)) \\ &= -X(n+1)+1 \\ &\leq |X(n+1)|+1. \end{aligned}$$

From the condition (3.2) and the above inequalities, we have that for any $n = 0,1,2,\ldots$ and any $s \in [0,1]$,

$$\begin{aligned} |h(X(n+s))| &\leq a^{|X(n+s)|/2} \\ &\leq a^{X(n+s)/2} + a^{-X(n+s)/2} \\ &\leq a^{1/2}(a^{|X(n)|/2} + a^{|X(n+1)|/2}). \end{aligned} \tag{3.2}$$

Similarly we see that

$$h(X(n+s))^2 \leq a(a^{|X(n)|} + a^{|X(n+1)|}),$$

which implies that

$$\begin{aligned} E[h(X(n+s))^2] &\leq a(E[a^{|X(n)|}] + E[a^{|X(n+1)|}]) \\ &< \infty. \end{aligned}$$

By using Lemma 2.8 for the processes $\{h(X(n+s))\}_{n=0,1,2,\ldots}$ ($s \in [0,1]$), we can show the following theorem.

Theorem 2.9. *Let $h(\cdot)$ be a function satisfying (3.1) and $\{w(t,T)\}$ $(0 \le t \le T, 0 < T < \infty)$ be a family of weights satisfying that*

(i)′. $w(t,T) \ge 0$ and $\int_0^T w(t,T)\,dt \to 1$ as $T \to \infty$,

(ii)′. $w(t,T)$ is monotone increasing in t,

(iii)′. $w(T,T) \to 0$ as $T \to \infty$.

Then

$$P\text{-}\lim_{T\to\infty} \int_0^T w(t,T)h(X(t))\,dt = \sum_{j\in S} \pi_j E\left[\int_0^1 h(X(t))\,dt \,\middle|\, X(0)=j\right], \tag{3.3}$$

where S is the state space of the Markov chain $\{X(n)\}$.

Proof. First we assume that $X(0)$ is distributed as the invariant distribution $\{\pi_j\}$. From the conditions (i)′–(iii)′, for any $s \in [0,1]$, the weight $\{w(n+s,T)\}$ $(n = 0,1,\ldots,[T]-1, 0 < T < \infty)$ satisfies the conditions (i)–(iii) in Lemma 2.8, where $[T]$ denotes the integral part of T. Therefore, for any $s \in [0,1]$,

$$P\text{-}\lim_{T\to\infty} \sum_{n=0}^{[T]-1} w(n+s,T)\,h(X(n+s)) = E[h(X(s))]. \tag{3.4}$$

In particular, for $s = 0,1$,

$$P\text{-}\lim_{T\to\infty} \sum_{n=0}^{[T]-1} w(n+s,T)a^{|X(n+s)|/2} = E[a^{|X(s)|/2}] < \infty.$$

Hence, for any sequence $\{T_k\}$, there exists a subsequence $\{T_{k_m}\}$ such that

$$P\left\{\lim_{m\to\infty} \sum_{n=0}^{[T_{k_m}]-1} w(n+s,T_{k_m})a^{|X(n+s)|/2} = E[a^{|X(s)|/2}] \text{ for } s = 0,1\right\} = 1. \tag{3.5}$$

From (3.2), (3.4) and (3.5), we have that

$$P\left\{\lim_{m\to\infty}\sum_{n=0}^{[t_{k_m}]-1} w(n+s, T_{k_m})h(X(n+s)) = E[h(X(s))] \text{ for every } s\in[0,1)\right\} = 1.$$

Thus we obtain the following equation with probability 1

$$\begin{aligned}
&\lim_{m\to\infty}\int_0^{[T_{k_m}]} w(t, T_{k_m})h(X(t))\,dt \\
&= \lim_{m\to\infty}\int_0^1 \sum_{n=0}^{[T_{k_m}]-1} w(n+s, T_{k_m})h(X(n+s))\,ds \\
&= \int_0^1 E[h(X(s))]ds\,,
\end{aligned}$$

which implies that

$$\begin{aligned}
&P\text{-}\lim_{T\to\infty}\int_0^{[T]} w(t,T)h(X(t))\,dt \\
&= E\left[\int_0^1 h(X(t))\,dt\right] \\
&\left(= \sum_{j\in S}\pi_j E\left[\int_0^1 h(X(t))\,dt\,\middle|\, X(0)=j\right]\right).
\end{aligned}$$

It follows from (4.2) that $\int_{[T]}^{T} w(t,T)h(X(t))\,dt$ converges to 0 in mean square as T tends to infinity. Hence we obtain (3.3).

See Vere-Jones and Ogata (1984) for proof of this theorem without the assumption for the initial distribution of the process $X(\cdot)$.

Let $f(\cdot)$ be a differentiable and strictly monotone increasing function satisfying that $f(0)=0$ and $\lim_{x\to\infty} f(x) = \infty$. We consider a point process $M(\cdot)$ with the conditional intensity

$$\lambda(t|M(t)) = f'(t)\psi(f(t) - M(t))\,,$$

where the function $\psi(\cdot)$ satisfies the conditions (C1)–(C4). The conditional intensity function of the process $\widetilde{M}(t) = M(f^{-1}(t))$ is

$$\lambda(t|\widetilde{M}(t)) = \psi(t - \widetilde{M}(t)).$$

Hence for the process $\widetilde{Y}(t) = t - \widetilde{M}(t)$, the above-mentioned results hold, for instance, the Markov chain $\{\widetilde{Y}(n)\}_{n=0,1,2,\dots}$ is ergodic and has the invariant distribution $\{\widetilde{\pi}_j\}$. By using Theorem 2.9 for the process $\widetilde{M}(\cdot)$, we have the following corollary.

Corollary 2.10. *Let $h(\cdot)$ be a function satisfying (3.1) and $\{w(t,T)\}$ $(0 \le t \le T, 0 < T < \infty)$ be a family of weights satisfying that*

(i)″. $w(t,T) \ge 0$ *and* $\int_0^T w(t,T)\,dt \to 1$ *as* $T \to \infty$,

(ii)″. $\dfrac{w(t,T)}{f'(t)}$ *is monotone increasing in* t,

(iii)″. $\dfrac{w(T,T)}{f'(T)} \to 0$ *as* $T \to \infty$.

Then

$$P\text{-}\lim_{T\to\infty} \int_0^T w(t,T)h(Y(t))\,dt = \sum \widetilde{\pi}_j E\left[\int_0^{f^{-1}(1)} f(t)h(Y(t))\,dt \middle| Y(0) = j\right],$$

where $Y(t) = f(t) - N(t)$.

Proof. Let $\widetilde{w}(s,S) = \dfrac{w(f^{-1}(s), f^{-1}(S))}{f'(f^{-1}(s))}$. Then we can easily verify the conditions (i)′–(iii)′ in the previous theorem. Using the previous theorem for the process $\widetilde{M}(\cdot)$ and the weights $\widetilde{w}(s,S)$, we obtain that

$$P\text{-}\lim_{T\to\infty} \int_0^T w(t,T)h(Y(t))\,dt$$

$$= P\text{-}\lim_{T\to\infty} \int_0^{f(T)} w(f^{-1}(s),T)h(Y(f^{-1}(s)))\frac{1}{f'(f^{-1}(s))}\,ds$$

$$= P\text{-}\lim_{T\to\infty} \int_0^{f(T)} \widetilde{w}(s, f(T))h(\widetilde{Y}(s))\,ds$$

$$= \sum \widetilde{\pi}_j E\left[\int_0^1 h(\widetilde{Y}(t))\,ds \,\middle|\, \widetilde{Y}(0) = j\right]$$

$$= \sum \widetilde{\pi}_j E\left[\int_0^{f^{-1}(1)} f(t)h(Y(t))\,dt \,\middle|\, Y(0) = j\right]$$

Chapter 3. Maximum Likelihood Estimation in Self-Correcting Point Processes

3.1. Introduction

In the previous chapter, we have shown a version of the law of the large number for functionals of the process $t - N(t)$, where $N(\cdot)$ is the self-correcting point process, that is, its conditional intensity is $\psi(t - N(t))$, where the function $\psi(\cdot)$ satisfies the conditions (C1)–(C4). In the present chapter, we consider a self-correcting point process with a parametrized conditional intensity

$$\lambda(t,\theta) = \rho\,\psi(\beta\{\rho t - N(t) + \alpha\})\,, \tag{1.1}$$

where the function $\psi(\cdot)$ satisfies the conditions (C1)–(C4). The parameter $\theta = (\alpha, \beta, \rho)'$ belongs to the parameter space $\Theta = \mathbf{R} \times (0, M_\beta) \times (0, \infty)$, where $\mathbf{R}$ is the real line, M_β (> 1) is a constant and v' denotes the transposition of a vector v. The parameters α, β and ρ are respectively related to the origin of the time axis, sensitivity of the self-correcting and the scale of the time axis.

We shall investigate asymptotic properties of the maximum likelihood estimator (MLE) of the parameter θ (but we do not consider the estimation problem of the function $\psi(\cdot)$ or we treat it as a known function). Unfortunately, we can not give the explicit expression of the MLE. Thus we do not examine the MLE but asymptotic behavior of the likelihood ratio normalized by a matrix which is associated with the information matrix. In Section 3.2, we investigate asymptotic behavior of the information matrix. The process

$$Y(t,\theta) = \rho t - N(t) + \alpha \tag{1.2}$$

inherits the Markov property from the process $N(\cdot)$. Using the law of large numbers for the process $Y(t,\theta)$, which is ensured by Corollary 2.10, we see that a standardized information matrix converges to a positive definite matrix. In Section 3.3, we shall review definition and some basic properties of local asymptotic normality (LAN) and show that the family of the measures induced by the self-correcting processes is LAN. In Section 3.4, we obtain that the MLE is consistent and asymptotically normal.

3.2. Asymptotic behavior of the information matrix

We assume the following conditions to obtain Lemma 3.1 below about convergence of the information matrix $I_T(\theta)$:

(C1)′. $0 < \psi(x) < \infty$ for all $x \in \mathbf{R}$,

(A1). the function $\psi(\cdot)$ is continuously differentiable in $\mathbf{R}$.

(A2). there exist an $M > 0$ and an $a \in (1, a_0)$ such that

$$\frac{|\psi'(x)|^2}{\psi(x)} \le a^{|x|/(2M_\beta)} + M$$

for all $x \in \mathbf{R}$, where M_β is the upper bound of the parameter β and a_0 is defined by (2.4) in Chapter 2.

We easily see that the condition (A1) implies the condition (C4).

The log likelihood $\ell(T,\theta)$ based on the observation $(N(t);\ 0 \le t \le T)$ up to time T is written as

$$\ell(T,\theta) = \int_0^T \log\lambda(t,\theta)dN(t) - \int_0^T \lambda(t,\theta)dt \tag{2.1}$$

(see e.g. Liptser and Shiryayev (1977) and Rubin (1972)). The derivative of $\lambda(t,\theta)$ with respect to the parameter θ is

$$\begin{aligned} &\overset{\bullet}{\lambda}(t,\theta) \\ &= (\rho\beta\psi'(\beta Y(t,\theta)),\ \rho Y(t,\theta)\psi'(\beta Y(t,\theta)),\ \psi(\beta Y(t,\theta)) + \rho\beta\psi'(\beta Y(t,\theta))\cdot t)', \end{aligned} \tag{2.2}$$

where $Y(t,\theta)$ is defined by (1.2). We see that

(2.3)
$$
\begin{aligned}
&I_T(\theta)\\
&= E_\theta\left[\left(\frac{\partial}{\partial\theta}\ell(T,\theta)\right)\left(\frac{\partial}{\partial\theta}\ell(T,\theta)\right)'\right]\\
&= E_\theta\left[\left(\int_0^T \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}(dN(t)-\lambda(t,\theta)dt)\right)\left(\int_0^T \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}(dN(t)-\lambda(t,\theta)dt)\right)'\right]\\
&= E_\theta\left[\int_0^T \frac{\dot{\lambda}(t,\theta)\dot{\lambda}(t,\theta)'}{\lambda(t,\theta)}dt\right].
\end{aligned}
$$

See Kutoyants (1984) for the last equation.

Let $D_T = \mathrm{diag}(T,T,T^3)$. From (1.1) and (2.2), we have the following equation:

$$
\begin{aligned}
&D_T{}^{-1/2}\int_0^T \frac{\dot{\lambda}(t,\theta)\dot{\lambda}(t,\theta)'}{\lambda(t,\theta)}dt\, D_T{}^{-1/2}\\
&= \int_0^T \frac{\rho^2\psi'(\beta Y(t,\theta))^2}{\rho\psi(\beta Y(t,\theta))}\begin{pmatrix} \beta/T & \beta Y(t,\theta)/T & \beta^2\cdot t/T^2\\ & Y(t,\theta)^2/T & \beta Y(t,\theta)\cdot t/T^2\\ sym. & & \beta^2\cdot t^2/T^3\end{pmatrix}dt\\
&+\int_0^T \frac{1}{\rho}\begin{pmatrix} 0 & 0 & \rho\beta\psi'(Y(t,\theta))\psi(Y(t,\theta))/T^2\\ & 0 & \rho Y(t,\theta)\psi'(Y(t,\theta))\psi(Y(t,\theta))/T^2\\ sym. & & \{2\rho\beta\psi'(Y(t,\theta))\psi(Y(t,\theta))\cdot t+\psi(Y(t,\theta))^2\}/T^3\end{pmatrix}dt.
\end{aligned}
$$

Let the operator $E_Y[\,\cdot\,]$ denote the expectation with respect to the equilibrium distribution of the process $Y(t,\theta)$, that is,

$$
E_Y[f(Y(t,\theta))] = \sum \tilde{\pi}_j E_\theta[f(Y(t,\theta))\,|Y(0,\theta)=j] \tag{2.4}
$$

for every function $f(\cdot)$ for which there exists the expectation in the right-hand side, where $\{\tilde{\pi}_j\}$ is the invariant distribution of the skeleton Markov chain $\{Y(n/\rho,\theta)\}_{n=0,1,2,\dots}$. Using

Corollary 2.10 for the process $Y(t,\theta)$, we obtain that

$$D_T^{-1/2}\int_0^T \frac{\dot{\lambda}(t,\theta)\dot{\lambda}(t,\theta)'}{\lambda(t,\theta)}dt\, D_T^{-1/2} \to I(\theta) \tag{2.5}$$

in probability as $T\to\infty$ and this convergence is uniform in every compact subset $K\subset\Theta$, where

$$I(\theta) = E_Y\left[\rho\int_0^{1/\rho}\frac{\rho\psi'(\beta Y(t,\theta))^2}{\psi(Y(t,\theta))}\begin{pmatrix}\beta^2 & \beta Y(t,\theta) & \beta^2/2\\ & Y(t,\theta)^2 & \beta Y(t,\theta)/2\\ sym. & & \beta^2/3\end{pmatrix}dt\right]. \tag{2.6}$$

Moreover we see that the matrix $I(\theta)$ is positive definite. Indeed, for any vector $v=(v_1,v_2,v_3)'$,

$$v'I(\theta)v = E_Y\left[\rho\int_0^{1/\rho}\frac{\rho\psi'(\beta Y(t,\theta))^2}{\psi(Y(t,\theta))}\{(\beta v_1+Y(t,\theta)v_2+\frac{1}{2}\beta v_3)^2+\frac{1}{12}\beta^2 {v_3}^2\}\,dt\right]$$

and $\psi'(x)$ is positive on an interval from the conditions (C3) and (A1). Thus we obtain the following lemma.

Lemma 3.1. *Under the conditions (C1)′, (C2), (C3), (A1) and (A2),*

$$D_T^{-1/2}I_T(\theta)D_T^{-1/2}\to I(\theta)$$

as $T\to\infty$ and this convergence is uniform in any compact subset $K\subset\Theta$, where $D_T = diag(T,T,T^3)$ and $I(\theta)$ is given by (2.6).

3.3. Local asymptotic normality

In the present section, we shall review definition and some basic properties of local asymptotic normality (LAN) and show that the family of the measures induced by the self-correcting point processes is LAN.

We consider a stochastic process $X(\cdot)$ which satisfies that all realizations of $X_T = \{X(t)\,;\,0 \le t \le T\}$ belong to a measurable space $(\mathcal{X}_T, \mathcal{B}_T)$. Let $P_\theta^{(T)}$ denote the measure induced by X_T on the measurable space $(\mathcal{X}_T, \mathcal{B}_T)$, where $\theta \in \Theta \subset \mathbf{R}^k$ (k is a positive integer). Assume that all the measures in the family $\{P_\theta^{(T)}\,;\theta \in \Theta\}$ are equivalent, The family $\{P_\theta^{(T)}\}$ is said to be LAN at θ as T tends to infinity if there exists a non-singular $k \times k$ matrix $\Phi_T(\theta)$ such that the likelihood ratio normalized by $\Phi_T(\theta)$ is written as

$$\frac{P^{(T)}_{\theta+\Phi_T(\theta)u}}{P_\theta^{(T)}} = \exp\{u'\delta_T(\theta, X) - \frac{1}{2}u'u + g_T(\theta, u, X)\}\,, \tag{3.1}$$

where the vector $u(\in \mathbf{R}^k)$ satisfies that $\theta + \Phi_T(\theta)u \in \Theta$,

$$\delta_T(\theta, X) \to N(0, I) \qquad \text{in law} \quad \text{as } T \to \infty \tag{3.2}$$

and

$$g_T(\theta, u, X) \to 0 \qquad \text{in probability} \quad \text{as } T \to \infty \tag{3.3}$$

($N(0, I)$ denotes the normal distribution with 0 mean vector and identity variance matrix). The family is LAN in Θ if it is LAN at θ for every $\theta \in \Theta$. We call uniformly LAN if it is LAN in Θ and each convergence of (3.2) and (3.3) is uniform in θ. The matrix $\Phi_T(\theta)$ is called the normalizing matrix and is usually equal to the inverse of the square root of the information matrix. Under suitable conditions, uniformly LAN implies that the MLE is consistent and asymptotic normal (see e.g. Ibragimov and Has'minskii (1981) and Kutoyants (1984)), which will be proved for the present model in the following section.

We shall show the following theorem by using Lemma 3.1 and Theorem 4.5.3 in Kutoyants (1984).

Theorem 3.2. *Assume the conditions (C1)', (C2), (C3), (A1), (A2) and*

(A3). there exists a $\xi_0 > 0$ *such that*

$$\sup_{x\in\mathbf{R},\,|\xi|<\xi_0}\left\{\frac{|\psi'(x+\xi)|}{\psi(x)}\right\}<\infty.$$

Let $P_\theta^{(T)}$ *denote the measure induced by* $\{N(t)\,;0\le t\le T\}$*, where the process* $N(\cdot)$ *has the intensity given in (1.1). Then for any compact set* $K\subset\Theta$*, the family* $\{P_\theta^{(T)}\,;\theta\in\Theta\}$ *is uniformly LAN in* K *with the normalizing matrix*

$$\Phi_T(\theta)=D_T{}^{-1/2}I(\theta)^{-1/2}$$

and the vector

$$\delta_T(\theta,N)=\Phi_T(\theta)'\int_0^T\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}(dN(t)-\lambda(t,\theta)dt).$$

Proof. To show the present theorem it is enough to check the conditions (G1)–(G4) of Theorem 4.5.5 in Kutoyants (1984), which implies the assumption of Theorem 4.5.3 in Kutoyants (1984). The conditions (G1) and (G2) are easily seen and the condition (G3) is shown as follows. From (2.5), we have that the integral

$$\int_0^T\Phi_T(\theta)'\frac{\dot{\lambda}(t,\theta)\dot{\lambda}(t,\theta)'}{\lambda(t,\theta)}\Phi_T(\theta)\,dt \tag{3.4}$$

uniformly converges in probability to the identity matrix as $T\to\infty$. For every $\varepsilon>0,\theta\in K$ and $y>0$,

$$E_\theta\left[\int_0^T\frac{|\Phi_T(\theta)'\dot{\lambda}(t,\theta)|^2}{\lambda(t,\theta)}\chi\left(\frac{|\Phi_T(\theta)'\dot{\lambda}(t,\theta)|}{\lambda(t,\theta)}>\varepsilon\right)dt\right]$$

$$\le E_\theta\left[\int_0^T\frac{|\Phi_T(\theta)'\dot{\lambda}(t,\theta)|^2}{\lambda(t,\theta)}\chi\left(\frac{|\Phi_T(\theta)'\dot{\lambda}(t,\theta)|}{\lambda(t,\theta)}>\varepsilon\right)\chi(|Y(t,\theta)|\le y)\,dt\right] \tag{3.5}$$

$$+E_\theta\left[\int_0^T\frac{|\Phi_T(\theta)'\dot{\lambda}(t,\theta)|^2}{\lambda(t,\theta)}\chi(|Y(t,\theta)|>y)\,dt\right].$$

Since the first term of the third element of $\overset{\bullet}{\lambda}(t,\theta)$ is asymptotically negligible, we have that for any sufficiently large T,

$$\frac{|\Phi_T(\theta)'\overset{\bullet}{\lambda}(t,\theta)|}{\lambda(t,\theta)} \leq \frac{1}{\sqrt{T}} \frac{\psi'(\beta Y(t,\theta))}{\psi(\beta Y(t,\theta))} \sqrt{\lambda_1(Y(t,\theta)^2 + 2\beta^2 + \eta)}, \tag{3.6}$$

where $\lambda_1(> 0)$ is the maximum characteristic root of $I(\theta)^{-1}$ and η is a positive constant. Thus for any sufficiently large T, the first term of the right-hand side of (3.5) is equal to 0 by the condition (A3). From the condition (A2), (3.6) and Corollary 2.10, we obtain that for any sufficiently large y,

$$\begin{aligned} &\int_0^T \frac{|\Phi_T(\theta)'\overset{\bullet}{\lambda}(t,\theta)|^2}{\lambda(t,\theta)} \chi(|Y(t,\theta)| > y)\,dt \\ &\leq \frac{1}{T}\int_0^T \rho\lambda_1 \frac{\psi'(\beta Y(t,\theta))^2}{\psi(\beta Y(t,\theta))} (Y(t,\theta) + 2\beta + \eta)\,\chi(|Y(t,\theta)| > y)\,dt \\ &\leq \frac{1}{T}\int_0^T \rho\lambda_1 a^{|Y(t,\theta)|/2}\,\chi(|Y(t,\theta)| > y)\,dt \\ &\to E_Y\left[\rho\int_0^{1/\rho} \rho\lambda_1 a^{|Y(t,\theta)|/2}\,\chi(|Y(t,\theta)| > y)\,dt\right] (< \infty) \end{aligned} \tag{3.7}$$

in probability as $T \to \infty$, where a is the constant determined by the condition (A2) and $E_Y[\cdot]$ is the operator defined by (2.4), namely, the expectation with respect to the equilibrium distribution of the process $Y(t,\theta)$. Since the above expectation converges to zero as y tends to infinity, the condition (G3) holds.

Let $\theta(u) = \theta_T(u) = \theta + \Phi_T(\theta)u$ for $u \in \mathbf{R}^k$ and $U(T,\theta) = \{u\,;\, \theta_T(u) \in \Theta\}$. To obtain the condition (G4) we shall show (3.8) below for any sufficiently large T, any non-negative bounded function $f(\cdot)$ and any vectors u, $v \in U(T,\theta)$ which satisfy $|u| + |v| < T^{1/3}$,

$$\int_0^T \left|\Phi_T(\theta)'\frac{\overset{\bullet}{\lambda}(t,\theta(u))}{\lambda(t,\theta(v))}\right|^2 \lambda(t,\theta)f(t)\,dt \leq \frac{1}{T}\int_0^T \rho\lambda_1\, a^{|Y(t,\theta)|/2} f(t)\,dt, \tag{3.8}$$

where λ_1 is the maximum characteristic root of the matrix $I(\theta)^{-1}$. We get that $Y(t, \theta_T(u))$ uniformly converges to $Y(t, \theta)$ as T tends to infinity, because

$$Y(t, \theta_T(u)) = Y(t, \theta) + w_3 t T^{-3/2} + w_1 T^{-1/2},$$

where $(w_1, w_2, w_3)' = I(\theta)^{-1/2} u$.

There exists a ξ such that $\psi(x + \nu) - \psi(x) = \nu \psi'(x + \xi)$. By the condition (A3), $\dfrac{\psi(x+\nu)}{\psi(x)}$ uniformly converges to 1 as ν tends to 0. Thus we have that

$$\frac{\psi(\beta(u) Y(t, \theta(u)))}{\psi(\beta Y(t, \theta))} \to 1 \qquad \text{as} \quad T \to \infty, \tag{3.9}$$

where $(\alpha(u), \beta(u), \rho(u))' = \theta(u)$. We obtain (3.8) from (3.6), (3.9) and (A2).

We easily see that

$$\int_0^T \left| \Phi_T(\theta)' \left(\frac{\dot{\lambda}(t, \theta(u))}{\lambda(t, \theta(v))} - \frac{\dot{\lambda}(t, \theta)}{\lambda(t, \theta)} \right) \right|^2 \lambda(t, \theta)\, dt$$

$$\leq \int_0^T \left| \Phi_T(\theta)' \left(\frac{\dot{\lambda}(t, \theta(u))}{\lambda(t, \theta(v))} - \frac{\dot{\lambda}(t, \theta)}{\lambda(t, \theta)} \right) \right|^2 \lambda(t, \theta) \chi(|Y(t, \theta)| \leq y)\, dt$$

$$+ 2 \int_0^T \left| \Phi_T(\theta)' \frac{\dot{\lambda}(t, \theta(u))}{\lambda(t, \theta(v))} \right|^2 \lambda(t, \theta) \chi(|Y(t, \theta)| > y)\, dt$$

$$+ 2 \int_0^T \left| \Phi_T(\theta)' \frac{\dot{\lambda}(t, \theta)}{\lambda(t, \theta)} \right|^2 \lambda(t, \theta) \chi(|Y(t, \theta)| > y)\, dt.$$

The first term of the right-hand side tends to zero as T tends to infinity because $\left(\dfrac{\dot{\lambda}(t, \theta(u))}{\lambda(t, \theta(v))} - \dfrac{\dot{\lambda}(t, \theta)}{\lambda(t, \theta)} \right) \chi(|Y(t, \theta)| \leq y)$ uniformly converges to 0. From (3.8), we have that

$$\int_0^T \left| \Phi_T(\theta)' \frac{\dot{\lambda}(t, \theta(u))}{\lambda(t, \theta(v))} \right|^2 \lambda(t, \theta) \chi(|Y(t, \theta)| > y)\, dt$$

$$\leq \frac{1}{T}\int_0^T \rho\lambda_1(a^{|Y(t,\theta)|+\xi}+M)\chi(|Y(t,\theta)|>y)\,dt$$

$$\to E_Y\left[\rho\int_0^{1/\rho} \rho\lambda_1(a^{|Y(t,\theta)|+\xi}+M)\chi(|Y(t,\theta)|>y)\,dt\right] \qquad \text{as} \quad T\to\infty$$

and the above expectation converges to zero as y tends to infinity. Hence we obtain that

$$\lim_{T\to\infty}\sup_{\theta\in K}\sup_{|u|+|v|<T^{1/3}}\int_0^T \left|\Phi_T(\theta)'\left(\frac{\dot{\lambda}(t,\theta(u))}{\lambda(t,\theta(v))}-\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right)\right|^2 \lambda(t,\theta)\,dt = 0,$$

that is, the condition (G4).

3.4. Asymptotic properties of the MLE

We shall show that the MLE is consistent and asymptotic normal.

Theorem 3.3. *Let $\widehat{\theta}_T = \widehat{\theta}_T(\theta, N)$ denote the MLE of the parameter θ based on the observation $(N(t)\,;\,0\leq t\leq T)$ up to time T. Under the conditions (C1)′,(C2), (C3) and (A1)-(A3),*

$$\widehat{\theta}_T \to \theta \qquad \text{in } P_\theta \quad \text{as} \quad T\to\infty \tag{4.1}$$

and

$$\Phi_T(\theta)^{-1}(\widehat{\theta}_T-\theta)\to N(0,I) \qquad \text{in law} \quad \text{as} \quad T\to\infty, \tag{4.2}$$

where $\Phi_T(\theta)$ is given in Theorem 3.2.

Proof. It is sufficient to show the conditions (G5) and (G6) of Theorem 4.5.5 in Kutoyants (1984). For any $\theta, \theta(u)\in K$, there exists a $\theta(v)\in\widetilde{K}$ such that

$$\sqrt{\lambda(t,\theta(u))}-\sqrt{\lambda(t,\theta)} = u'\Phi_T(\theta)'\frac{\dot{\lambda}(t,\theta(v))}{2\sqrt{\lambda(t,\theta(v))}},$$

where $\widetilde{K}$ is the convex hull of K. Thus we have that

$$-\frac{1}{2}\int_0^T (\sqrt{\lambda(t,\theta(u))} - \sqrt{\lambda(t,\theta)})^2 dt$$

$$= -\frac{1}{8}\int_0^T u'\Phi_T(\theta)'\frac{\dot{\lambda}(t,\theta(v))\dot{\lambda}(t,\theta(v))'}{\lambda(t,\theta(v))}\Phi_T(\theta)u\,dt.$$

From the conditions(G3), (G4) and (3.9), the right-hand side uniformly converges to $-\frac{|u|}{8}$. Hence there exist a $c > 0$ such that for any sufficiently large T,

$$E_\theta\left[\exp\left\{-\frac{1}{2}\int_0^T (\sqrt{\lambda(t,\theta(u))} - \sqrt{\lambda(t,\theta)})^2 dt\right\}\right] \le \exp\{-c|u|^2\},$$

that is , the condition (G5) holds.

Finally, we shall show the condition

(G6). for some $p > \frac{3}{2}$,

$$\sup_{T>0;\ \theta,\theta_1\in K} J(T,\theta,\theta_1) < \infty,$$

where

$$J(T,\theta,\theta_1)$$

$$= E_\theta\left[\int_0^T \left|\Phi_T(\theta)'\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right|^{2p} \lambda(t,\theta)\,dt\right] + E_\theta\left(\int_0^T \left|\Phi_T(\theta)'\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right|^{2} \lambda(t,\theta)\,dt\right)^p$$

From the conditions (A2), (A3) and (3.6), we have that

$$\left|\Phi_T(\theta_1)'\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right|^{2p} \lambda(t,\theta)$$

$$= \left|\Phi_T(\theta_1)'\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right|^{2(p-1)} \times \left|\Phi_T(\theta_1)'\frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}\right|^{2} \lambda(t,\theta)$$

$$\le \zeta \times \left\{ \left(\frac{Y(t,\theta)}{\sqrt{T}} + \Delta\rho + \frac{\Delta\alpha}{\sqrt{T}} \right)^2 + \frac{2}{T}{\beta_1}^2 + \frac{\eta}{T} \right\}^{p-1} \times \frac{a^{|Y(t,\theta)|/2}}{T},$$

where $(\Delta\alpha, \Delta\beta, \Delta\rho) = (\alpha_1 - \alpha, \beta_1 - \beta, \rho_1 - \rho)$ and ζ is a constant which is independent of T. We obtain that

$$\int_0^T \left| \Phi_T(\theta)' \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)} \right|^{2p} \lambda(t,\theta)\, dt$$

$$\le \frac{1}{T} \int_0^T \zeta \times \left\{ \left(\frac{Y(t,\theta)}{\sqrt{T}} + \Delta\rho + \frac{\Delta\alpha}{\sqrt{T}} \right)^2 + \frac{2}{T}{\beta_1}^2 + \frac{\eta}{T} \right\}^{p-1} \times a^{|Y(t,\theta)|/2}\, dt$$

$$\to \zeta(\Delta\rho)^{2(p-1)} E_Y \left[\rho \int_0^{1/\rho} a^{|Y(t,\theta)|/2}\, dt \right],$$

where $E_Y[\,\cdot\,]$ denotes the expectation with respect to the equilibrium distribution of the process $Y(t,\theta)$. Thus we have that

$$\sup_{T>0;\ \theta,\theta_1 \in K} \text{ the first term of } J(T,\theta,\theta_1) < \infty.$$

It is similarly shown that the second term of $J(T,\theta,\theta_1)$ is bounded. Hence we obtain the condition (G6).

Chapter 4. Estimation of Intensity Levels in Simple Self-Correcting Point Processes

4.1. Introduction

In the previous chapter, we have investigated asymptotic properties of the maximum likelihood estimator (MLE) in a self-correcting point process $N(\cdot)$ with the intensity $\rho\,\psi(\beta\{\rho t - N(t) + \alpha\})$, where α, β and ρ are parameters. In the present chapter, we treat a simple self-correcting point process $N(\cdot)$ whose intensity has only two levels. More precisely, the intensity of the process $N(\cdot)$ is given by

$$\rho\,\psi(\beta\{\rho t - N(t) + \alpha\}) = \begin{cases} \rho\theta_1 & \text{if } \rho t - N(t) + \alpha \le 0, \\ \rho\theta_2 & \text{if } \rho t - N(t) + \alpha > 0, \end{cases}$$

where $0 < \theta_1 < 1 < \theta_2$ and

$$\psi(x) = \begin{cases} \theta_1 & \text{if } x \le 0, \\ \theta_2 & \text{if } x < 0. \end{cases}$$

Since $\psi(x)$ depends only on the sign of x, the parameter β does not make any sense. Here, we treat α and ρ as known constants and concentrate our interest on estimation of the intensity levels. We can choose the location and the scale of the time axis so that $\alpha = 0$ and $\rho = 1$. Hence, without loss of generality, we may assume that the conditional intensity of the process $N(\cdot)$ is

$$\lambda(t, \theta) = \psi(X(t)) = \begin{cases} \theta_1 & \text{if } X(t) \le 0, \\ \theta_2 & \text{if } X(t) > 0, \end{cases} \tag{1.1}$$

where $\theta = (\theta_1, \theta_2)'$ and $X(t) = t - N(t)$.

In Section 4.2, we explicitly give the log likelihood, the MLE and the information matrix. In Section 4.3, we calculate the invariant distribution of the Markov chain $\{X(n)\}$

and show Lemma 4.1 about convergence of a standardized information matrix by using the law of large numbers for the process $X(\cdot)$ which is ensured by Theorem 2.9. In Section 4.4, we show that the MLE is asymptotically normal and explicitly give its asymptotic variance. Moreover we obtain local asymptotic normality of the family of the measures induced by the simple self-correcting point process.

4.2. Likelihood and information

In the present section, we explicitly give the log likelihood, the MLE and the information matrix and state the lemma about convergence of a standardized information matrix. We show this lemma in the following section.

The log likelihood based on the observation $(N(t)\,;\,0 \le t \le T)$ up time T is given by

$$\begin{aligned}\ell(T,\theta) &= \int_0^T \log\lambda(t,\theta)dN(t) - \int_0^T \lambda(t,\theta)dt \\ &= \int_0^T \log\theta_1\,\chi(X(t)\le 0)\,dN(t) + \int_0^T \log\theta_2\,\chi(X(t)>0)\,dN(t) \\ &\quad - \left\{\int_0^T \theta_1\,\chi(X(t)\le 0)\,dt + \int_0^T \theta_2\,\chi(X(t)>0)\,dt\right\},\end{aligned}$$

where $\chi(\cdot)$ is the indicator. Let

$$\begin{aligned}D_1(T) &= \{t\in[0,T]\,;X(t)\le 0\}\,, & D_2(T) &= \{t\in[0,T]\,;X(t)>0\}\,,\\ N(D_1(T)) &= \int_0^T \chi(X(t)\le 0)\,dN(t)\,, & N(D_2(T)) &= \int_0^T \chi(X(t)>0)\,dN(t)\,,\\ |D_1(T)| &= \int_0^T \chi(X(t)\le 0)\,dt\,, & |D_2(T)| &= \int_0^T \chi(X(t)>0)\,dt\,.\end{aligned} \tag{2.1}$$

Then the log likelihood is written as

$$\begin{aligned}&\ell(T,\theta)\\ &= N(D_1(T))\,\log\theta_1 + N(D_2(T))\,\log\theta_2 - \{\,\theta_1|D_1(T)| + \theta_2|D_2(T)|\,\}\,.\end{aligned} \tag{2.2}$$

Hence the likelihood equation is as follows:

$$\frac{\partial}{\partial\theta}\ell(T,\theta) = \begin{pmatrix} \dfrac{N(D_1(T))}{\theta_1} - |D_1(T)| \\ \dfrac{N(D_2(T))}{\theta_2} - |D_2(T)| \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{1}{\theta_1} \int_0^T \chi(X(t) \leq 0)\,(dN(t) - \lambda(t,\theta)\,dt) \\ \dfrac{1}{\theta_2} \int_0^T \chi(X(t) > 0)\,(dN(t) - \lambda(t,\theta)\,dt) \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which implies that the MLE of the parameter θ is given by

$$\widehat{\theta}_T = \begin{pmatrix} \dfrac{N(D_1(T))}{|D_1(T)|} \\ \dfrac{N(D_2(T))}{|D_2(T)|} \end{pmatrix}. \tag{2.3}$$

We obtain that

$$E\left\{\int_0^T \chi(X(T) \leq 0)\,(dN(t) - \lambda(t,\theta)\,dt)\right\}^2$$

$$= E\left[\int_0^T \chi(X(T) \leq 0)^2\,\lambda(t,\theta)\,dt\right]$$

$$= \theta_1 E\left[\int_0^T \chi(X(T) \leq 0)\,dt\right]$$

$$= \theta_1 E|D_1(T)|$$

and that

$$E\left\{\int_0^T \chi(X(T)>0)\,(dN(t)-\lambda(t,\theta)\,dt)\right\}^2 = \theta_2 E|D_2(T)|.$$

Moreover we see that

$$E\left[\int_0^T \chi(X(T)\le 0)\,(dN(t)-\lambda(t,\theta)\,dt)\int_0^T \chi(X(T)>0)\,(dN(t)-\lambda(t,\theta)\,dt)\right]$$
$$=0.$$

Therefore the information matrix is given by

$$I_T(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ell(T,\theta)\right)\left(\frac{\partial}{\partial\theta}\ell(T,\theta)\right)'\right]$$
$$=\begin{pmatrix} \dfrac{E|D_1(T)|}{\theta_1} & 0 \\ 0 & \dfrac{E|D_2(T)|}{\theta_2} \end{pmatrix}.$$

The following lemma is shown in the following section.

Lemma 4.1. *Let*

$$I(\theta) = \begin{pmatrix} \dfrac{\theta_2-1}{\theta_1(\theta_2-\theta_1)} & 0 \\ 0 & \dfrac{1-\theta_1}{\theta_2(\theta_2-\theta_1)} \end{pmatrix}. \tag{2.4}$$

Then the matrix $I(\theta)$ is positive definite and

$$\frac{1}{T}I_T(\theta) \to I(\theta) \quad \text{as } T\to\infty.$$

4.3. Proof of Lemma 4.1

Since $\dfrac{|D_i(T)|}{T} \le 1$ $(i=1,2)$, to obtain Lemma 4.1, it is enough to show that

$$\frac{|D_1(T)|}{T} \to \frac{\theta_2-1}{\theta_2-\theta_1} \quad \text{in probability as } T\to\infty \tag{3.1}$$

and

$$\frac{|D_2(T)|}{T} \to \frac{1-\theta_1}{\theta_2-\theta_1} \quad \text{in probability as } T \to \infty \tag{3.2}$$

On the other hand, it is easy to verify the conditions (C1)–(C4) in Chapter 1 for the function $\psi(x) = \theta_1$ if $x \leq 0$ and $= \theta_2$ if $x > 0$. From Lemma 2.2 and Theorem 2.6 the Markov chain $\{X(n)\}_{n=0,1,2,\ldots}$ is irreducible, aperiodic and positive recurrent (i.e. ergodic). Furthermore its state space S is the class of all integers. We can easily check that the weight $w(t,T) = \frac{1}{T}$ satisfies the conditions (i)$'$–(iii)$'$ in Theorem 2.9 and that the function $h(x) = \chi(x \leq 0)$ satisfies (3.1) in Chapter 2. Hence we have that

$$\frac{|D_1(T)|}{T} = \frac{1}{T}\int_0^T \chi(X(t) \leq 0)dt$$
$$\to \sum_{j=-\infty}^{\infty} \pi_j E\left[\int_0^1 \chi(X(t) \leq 0)\,dt \,\middle|\, X(0)=j\right] \quad (= R_1, \text{ say}) \tag{3.3}$$

in probability as $T \to \infty$, where $\{\pi_j\}_{j=-\infty}^{\infty}$ denotes the invariant distribution of the Markov chain $\{X(n)\}$. Similarly,

$$\frac{|D_2(T)|}{T} \to R_2 \tag{3.4}$$

in probability as $T \to \infty$, where

$$R_2 = \sum_{j=-\infty}^{\infty} \pi_j E\left[\int_0^1 \chi(X(t) > 0)\,dt \,\middle|\, X(0)=j\right]. \tag{3.5}$$

Therefore, our purpose is to show that

$$R_1 = \frac{\theta_2 - 1}{\theta_2-\theta_1} \quad \text{and} \quad R_2 = \frac{1-\theta_1}{\theta_2-\theta_1}.$$

First we shall investigate the transition probability p_{ij} of the Markov chain $\{X(n)\}$ to obtain its invariant distribution $\{\pi_j\}$. We easily see that

$$\begin{aligned} p_{ij} &= P\{X(n+1)=j|X(n)=i\} \\ &= P\{X(n)-X(n+1)=i-j|X(n)=i\} \\ &= P\{N(n+1)-N(n)=i-j+1|X(n)=i\} \\ &= P\{N_i(1)=i-j+1\}, \end{aligned}$$

where $N_i(\cdot)$ is a point process with the conditional intensity $\psi(t+i-N_i(t))$ and the last equation follows from Lemma 2.1. It is clear that for $i \le j-2$, $p_{ij}=0$. Since $\psi(x)=\theta_1$ for $x \le 0$ and $=\theta_2$ for $x>0$, we have that for $i \le -1$ and $j \le i+1$,

$$\begin{aligned} p_{ij} &= P\{N_i(1)=i-j+1\} \\ &= \frac{{\theta_1}^{i-j+1}}{(i-j+1)!}\exp\{-\theta_1\}, \end{aligned} \tag{3.6}$$

for $i \ge 0$ and $1 \le j \le i+1$,

$$p_{ij} = \frac{{\theta_2}^{i-j+1}}{(i-j+1)!}\exp\{-\theta_2\} \tag{3.7}$$

and that for $i \ge 0$ and $j\ (=-k)\ \le 0$,

$$p_{ij} = \int_0^1 \frac{(\theta_2 t)^i}{i!}\exp\{-\theta_2 t\}\cdot\theta_2\cdot\frac{\{\theta_1(1-t)\}^k}{k!}\exp\{-\theta_1(1-t)\}\,dt\,. \tag{3.8}$$

The invariant distribution $\{\pi_j\}$ satisfies the equation

$$\pi_j = \sum_{i=-\infty}^{\infty} \pi_i p_{ij}\,. \tag{3.9}$$

For $j \ge 1$, the above equation is written as

$$\begin{aligned} \pi_j &= \sum_{i=j-1}^{\infty} \pi_i \frac{{\theta_2}^{i-j+1}}{(i-j+1)!}\exp\{-\theta_2\} \\ &= \sum_{h=0}^{\infty} \pi_{h+j-1}\frac{{\theta_2}^h}{h!}\exp\{-\theta_2\} \end{aligned} \tag{3.10}$$

because $p_{ij} = 0$ for $i \leq j-2$ and p_{ij} is given in (3.7) for $j \geq 1$ and $i \geq j-1\,(\geq 0)$. Putting

$$\pi_j = \pi_0 q^j \quad \text{for } j \geq 1, \tag{3.11}$$

we have that

$$\begin{aligned}\pi_0 q^j &= \sum_{h=0}^{\infty} \pi_0 q^{h+j-1} \frac{{\theta_2}^h}{h!} \exp\{-\theta_2\} \\ &= \pi_0 q^{j-1} \exp\{\theta_2(q-1)\},\end{aligned}$$

which implies that

$$q = \exp\{\theta_2(q-1)\}. \tag{3.12}$$

Since for $c > 1$, the equation (2.3) in Chapter 2 has exactly two solutions 1 and $x_0\,(0 < x_0 < 1)$, we can find a $q \in (0,1)$ satisfying (3.12). The value of π_0 will be determined later.

For $j\,(= -k) \leq 0$, the equation (3.9) is written as

$$\pi_{-k} = \sum_{i=-k-1}^{\infty} \pi_i p_{i,-k}\,.$$

From (3.6), (3.8) and (3.11), we have that

$$\begin{aligned}\pi_{-k} &= \int_0^1 \sum_{i=0}^{\infty} \pi_0 q^i \cdot \frac{(\theta_2 t)^i}{i!} \exp\{-\theta_2 t\} \cdot \theta_2 \cdot \frac{\{\theta_1(1-t)^k\}}{k!} \exp\{-\theta_1(1-t)\}\,dt \\ &\quad + \sum_{i=-k-1}^{-1} \pi_i p_{i,-k} \\ &= \pi_0 \int_0^1 \exp\{\theta_2 t(q-1)\} \cdot \theta_2 \cdot \frac{\{\theta_1(1-t)^k\}}{k!} \exp\{-\theta_1(1-t)\}\,dt \\ &\quad + \sum_{h=0}^{k} \pi_{-(k+1-h)} \frac{{\theta_1}^h}{h!} \exp\{-\theta_1\}.\end{aligned}$$

For $s \in [0,1]$, let $P(s)$ be the power series with the coefficients $\{\pi_{-k}\}_{k=0,1,2,\dots}$. From the above equation and (3.12), we have that

$$\begin{aligned}
P(s) &= \sum_{k=0}^{\infty} \pi_{-k} s^k \\
&= \pi_0 \int_0^1 \exp\{\theta_2 t(q-1)\} \cdot \theta_2 \cdot \exp\{\theta_1 s(1-t)\} \exp\{-\theta_1(1-t)\}\, dt \\
&\quad + \sum_{h=0}^{\infty} \sum_{k=h}^{\infty} \pi_{-(k+1-h)}\, s^{k-h} \frac{(\theta_1 s)^h}{h!} \exp\{-\theta_1\} \\
&= \pi_0 \theta_2 \exp\{\theta_1(s-1)\} \frac{\exp\{\theta_2(q-1) - \theta_1(s-1)\} - 1}{\theta_2(q-1) - \theta_1(s-1)} \\
&\quad + \exp\{\theta_1(s-1)\} \frac{P(s) - \pi_0}{s} \\
&= \pi_0 \theta_2 \frac{q - \exp\{-\theta_1(1-s)\}}{\theta_1(1-s) - \theta_2(1-q)} + \exp\{-\theta_1(1-s)\} \frac{P(s) - \pi_0}{s}.
\end{aligned}$$

Hence we obtain that for $s \in [0,1)$,

$$P(s) = \pi_0 \frac{s\theta_2[q - \exp\{-\theta_1(1-s)\}] - \{\theta_1(1-s) - \theta_2(1-q)\} \exp\{-\theta_1(1-s)\}}{\{\theta_1(1-s) - \theta_2(1-q)\}[s - \exp\{-\theta_1(1-s)\}]}.$$

We easily see that $0 \le P(1) = \sum_{k=0}^{\infty} \pi_{-k} \le \sum_{j=-\infty}^{\infty} \pi_j = 1$. From Abel's continuity theorem and L'Hospital's theorem, we obtain that

$$\begin{aligned}
P(1) &= \lim_{s \uparrow 1} P(s) \\
&= \pi_0 \frac{\theta_2(1-q) + \theta_1(\theta_2 q - 1)}{\theta_2(1-q)(1-\theta_1)}.
\end{aligned}$$

From (3.11), we have that

$$\begin{aligned}
P(1) &= 1 - \sum_{j=1}^{\infty} \pi_j \\
&= 1 - \frac{\pi_0 q}{1-q}.
\end{aligned}$$

Thus we obtain that

$$\pi_0 = \frac{\theta_2(1-q)(1-\theta_1)}{\theta_2 - \theta_1}. \tag{3.13}$$

Consequently we can conclude the calculation of the invariant distribution $\{\pi_j\}$ of the Markov chain $\{X(n)\}$.

We consider (3.4) again. It asserts that the ratio $\frac{|D_2(T)|}{T}$ converges in probability to $R_2 = \sum_{j=-\infty}^{\infty} \pi_j \, E[\int_0^1 \chi(X(t) > 0)\, dt \,|X(0) = j]$. Under $X(0) = j$, $X(t) = X(t) - X(0) + j = t + j - (N(t) - N(0))$. By Lemma 2.1, the conditional distribution of $N(t) - N(0)$ given $X(0) = j$ and the distribution of $N_j(t)$ are the same, where $N_j(\cdot)$ is a point process with the conditional intensity $\psi(t + j - N_j(t))$. Hence we have that

$$E\left[\int_0^1 \chi(X(t) > 0)\, dt \,\middle|\, X(0) = j\right] = \int_0^1 E[\chi(N_j(t) < t + j)]\, dt.$$

We can easily check that for any $j \geq 0$ and $0 < t \leq 1$,

$$\begin{aligned} E[\chi(N_j(t) < t + j)] &= E[\chi(N_j(t) \leq j)] \\ &= P(N_j(t) \leq j) \\ &= \sum_{k=0}^{j} \frac{(\theta_2 t)^k}{k!} \exp\{-\theta_2 t\} \\ &= \int_{\theta_2 t}^{\infty} \frac{x^j}{j!} e^{-x}\, dx \end{aligned}$$

and that $\chi(N_j(t) < t + j) = 0$ for $j \leq -1$ and $0 \leq t \leq 1$. From (3.11)–(3.13) and the above equations, we conclude that

$$\begin{aligned} R_2 &= \sum_{j=0}^{\infty} \pi_0 q^j \int_0^1 E[\chi(N_j(t) \leq j)]\, dt \\ &= \pi_0 \int_0^1 \int_{\theta_2 t}^{\infty} \sum_{j=0}^{\infty} \frac{(qx)^j}{j!} e^{-x}\, dx\, dt \end{aligned}$$

$$= \pi_0 \int_0^1 \int_{\theta_2 t}^{\infty} \exp\{(q-1)x\}\, dx\, dt$$

$$= \frac{\pi_0(1 - \exp\{-\theta_2(1-q)\})}{\theta_2(1-q)^2}$$

$$= \frac{1-\theta_1}{\theta_2 - \theta_1}.$$

Since $|D_1(T)| + |D_2(T)| = T$, we obtain that

$$\begin{aligned} R_1 &= 1 - R_2 \\ &= \frac{\theta_2 - 1}{\theta_2 - \theta_1}. \end{aligned}$$

4.4. Asymptotic normality of the MLE

It is easy to show that

$$\begin{aligned} &\sqrt{T}(\widehat{\theta}_T - \theta) \\ &= \sqrt{T} \begin{pmatrix} \dfrac{N(D_1(T))}{|D_1(T)|} - \theta_1 \\ \dfrac{N(D_2(T))}{|D_2(T)|} - \theta_2 \end{pmatrix} \\ &= \sqrt{T} \begin{pmatrix} \dfrac{1}{|D_1(T)|} \int_0^T \chi(X(t) \le 0)\, (dN(t) - \lambda(t,\theta)\, dt) \\ \dfrac{1}{|D_2(T)|} \int_0^T \chi(X(t) > 0)\, (dN(t) - \lambda(t,\theta)\, dt) \end{pmatrix} \\ &= T \begin{pmatrix} \dfrac{\theta_1}{|D_1(T)|} & 0 \\ 0 & \dfrac{\theta_2}{|D_2(T)|} \end{pmatrix} \Delta_T(\theta), \end{aligned}$$

where the MLE $\widehat{\theta}_T$ is given by (2.3) and

$$\Delta_T(\theta) = \begin{pmatrix} \dfrac{1}{\sqrt{T}\,\theta_1} \int_0^T \chi(X(t) \le 0)\,(dN(t) - \lambda(t,\theta)\,dt) \\ \dfrac{1}{\sqrt{T}\,\theta_2} \int_0^T \chi(X(t) > 0)\,(dN(t) - \lambda(t,\theta)\,dt) \end{pmatrix} \tag{4.1}$$

$$\left(= \frac{1}{\sqrt{T}} \frac{\partial}{\partial \theta} \ell(T,\theta) \right).$$

From the central limit theorem for martingales, we obtain that $\Delta_T(\theta)$ is asymptotically normal, that is,

$$\Delta_T(\theta) \to N(0, I(\theta)) \quad \text{in law} \quad \text{as } T \to \infty \tag{4.2}$$

By (3.1) and (3.2), we have that

$$T \begin{pmatrix} \dfrac{\theta_1}{|D_1(T)|} & 0 \\ 0 & \dfrac{\theta_2}{|D_2(T)|} \end{pmatrix} \to I(\theta)^{-1}$$

in probability as $T \to \infty$. Hence we obtain the following theorem.

Theorem 4.2. *The MLE $\widehat{\theta}_T$ is asymptotically normal, that is,*

$$\sqrt{T}(\widehat{\theta}_T - \theta) \to N(0, I(\theta)^{-1}) \quad \text{in law as} \quad T \to \infty,$$

where $\widehat{\theta}_T$ and $I(\theta)$ are given in (2.3) and (2.4), respectively.

Finally, we shall show that the family of the measures induced by the processes $N(\cdot)$ is locally asymptotically normal. From (2.2), the log likelihood ratio is written as

$$\begin{aligned} &\ell(T, \theta + hT^{-1/2}) - \ell(t,\theta) \\ &= N(D_1(T))\log\left(\frac{\theta_1 + h_1 T^{-1/2}}{\theta_1}\right) + N(D_2(T))\log\left(\frac{\theta_2 + h_2 T^{-1/2}}{\theta_2}\right) \\ &\quad - \{h_1|D_1(T)|T^{-1/2} + h_2|D_2(T)|T^{-1/2}\}, \end{aligned} \tag{4.3}$$

where $h = (h_1, h_2)'$. By (3.1), (3.2) and the law of large numbers for martingales, we have that

$$\frac{N(D_1(T))}{T} = \frac{1}{T}\int_0^T \chi(X(t) \le 0)\,(dN(t) - \lambda(t,\theta)\,dt) + \frac{\theta_1 |D_1(T)|}{T} \tag{4.4}$$

$$\rightarrow \frac{\theta_1(\theta_2 - 1)}{\theta_2 - \theta_1}$$

and

$$\frac{N(D_2(T))}{T} \rightarrow \frac{\theta_2(1-\theta_1)}{\theta_2 - \theta_1} \tag{4.5}$$

in probability as $T \rightarrow \infty$. From (4.1), (4.4), (4.5) and the Taylor's expansion

$$\log\left(\frac{\theta_i + h_i T^{-1/2}}{\theta_i}\right) = \frac{h_i}{\theta_i} T^{-1/2} - \frac{1}{2}\left(\frac{h_i}{\theta_i}\right)^2 T^{-1} + o(T^{-1}) \quad (i = 1, 2),$$

we obtain that

$$\ell(T, \theta + hT^{-1/2}) - \ell(t, \theta) = h'\Delta_T(\theta) - \frac{1}{2} h' G_T(\theta) h + o_p(1), \tag{4.6}$$

where

$$G_T(\theta) = \frac{1}{T}\begin{pmatrix} \dfrac{N(D_1(T))}{{\theta_1}^2} & 0 \\ 0 & \dfrac{N(D_2(T))}{{\theta_2}^2} \end{pmatrix}. \tag{4.7}$$

It follows from (4.4) and (4.5) that $G_T(\theta) \rightarrow I(\theta)$ in probability as $T \rightarrow \infty$. Hence we can easily show the following theorem.

Theorem 4.3. *Let $P_\theta^{(T)}$ denote the measure induced by $\{N(t)\,;\, 0 \le t \le T\}$, where the process $N(\cdot)$ has the intensity given in (1.1). Then the family $\{P_\theta^{(T)}\}$ is locally asymptotically normal with the normalizing matrix*

$$\Phi_T(\theta) = \frac{1}{\sqrt{T}} I(\theta)^{-1/2}$$

and the vector

$$\delta_T(\theta) = \Phi_T(\theta)\frac{\partial}{\partial\theta}\ell(T,\theta)$$

$$= I(\theta)^{-1/2}\Delta(\theta)\,.$$

Proof. For a vector u, we substitute $I(\theta)^{-1/2}u$ for the vector h in (4.6). Since $G_T(\theta) \to I(\theta)$ in probability as $T \to \infty$, we obtain the expression (3.1) in Chapter 3, that is, the conclusion of the present theorem.

Chapter 5. Robust Estimation in the Poisson Processes with a Periodic Intensity

5.1. Introduction

We consider a Poisson process $N(t)$ with a parametrized intensity $\lambda(t,\theta)$, where the parameter θ belongs to a bounded open interval Θ of $\mathbf{R}$ (the real line). The log likelihood function based on the observation $(N(t); 0 \le t \le T)$ up to time T is given by

$$\ell(T,\theta) = \int_0^T \log \lambda(t,\theta) dN(t) - \int_0^T \lambda(t,\theta) dt .$$

The maximum likelihood estimator (MLE) maximizes the log likelihood $\ell(T,\theta)$ and is a solution of the likelihood equation

$$\int_0^T \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)} dN(t) - \int_0^T \dot{\lambda}(t,\theta) dt = 0$$

under some regularity conditions, where $\dot{\lambda}$ is the derivative of λ with respect to θ. Moreover it is well known that the MLE is consistent, asymptotically normal and efficient (see e.g. Kutoyants (1984)).

If the artificial model does not sufficiently reflect the generation mechanism of the data or if the data are contaminated by noises, the true intensity $\mu(t)$ of the process $N(\cdot)$ may not belong to the parametric model $\{\lambda(t,\theta); \theta \in \Theta\}$. In such circumstances, the MLE is not always an appropriate estimator of the parameter θ. Our purpose is to construct robust estimators in the sense that high efficiency is kept even if the true intensity $\mu(t)$ does not belong to the parametric model $\{\lambda(t,\theta); \theta \in \Theta\}$.

The robust estimation problem is studied by many statisticians. Huber (1981) and Hampel et. al. (1986) sum it up in independently and identically distributed cases. In time series, it is studied by Kleiner et. al. (1979), Denby and Martin (1979), Künsch (1984), Martin and Yohai (1985, 1986), Bustos and Yohai (1986) and many other authors. Yoshida (1988) treats it in the diffusion processes. They use the M-estimation and the GM-estimation to get robust estimators. Here, we treat M-estimators which are solutions of generalized likelihood equations, more precisely,

DEFINITION. *For functions $h(t,\theta)$ and $H(t,\theta)$, a solution of the equation*

$$C(T,\theta) = \int_0^T h(t,\theta)dN(t) - \int_0^T H(t,\theta)dt = 0$$

is called the M-estimator.

The MLE corresponds to the M-estimator for

$$h(t,\theta) = \frac{\dot{\lambda}(t,\theta)}{\lambda(t,\theta)}, \qquad H(t,\theta) = \dot{\lambda}(t,\theta)$$

when the parametric model is $\{\lambda(t,\theta); \theta \in \Theta\}$.

In Section 5.2, we show that our M-estimators are consistent in a sense and asymptotically normal. In Section 5.3, we discuss estimation of the phase parameter θ of a periodic intensity $\lambda(t,\theta) = f(t-\theta)$, where $f(\cdot)$ is a periodic and even function satisfying suitable conditions. We construct the M-estimator which has the minimax variance provided that the true intensity belongs to a suitable class.

5.2. Asymptotic behavior of the M-estimators

We assume the following conditions to show that the M-estimator $\widehat{\theta}_T$ consistent in a sense and asymptotically normal.

(1). The true intensity $\mu(t)$ of the Poisson process $N(\cdot)$ is a bounded measurable function with period $\tau(>0)$.

(2). The functions $h(t,\theta)$ and $H(t,\theta)$ are periodic in t with period τ for all $\theta \in \Theta$ and are absolutely continuous with respect to θ for all $t \geq 0$. Their Radon-Nikodym derivatives $\dot{h}(t,\theta)$ and $\dot{H}(t,\theta)$ are bounded in (t,θ).

(3). There exists a $\theta_1 \in \Theta$ such that

$$\int_0^\tau \{H(t,\theta_1) - h(t,\theta_1)\mu(t)\}dt = 0$$

(4).
$$\Gamma = \frac{1}{\tau}\int_0^\tau [\dot{H}(t,\theta_1) - \dot{h}(t,\theta_1)\mu(t)]dt > 0$$

$$\Phi = \frac{1}{\tau}\int_0^\tau h(t,\theta_1)^2\mu(t)dt > 0$$

(5). There exist constants C_1 and C_2 which are independent of θ and t, such that for any sufficiently small $\delta > 0$,

$$\nu\left(\bigcup_{|\theta-\theta_1|\leq\delta}\left\{t \in [0,\tau]\,;\,|\dot{h}(t,\theta) - \dot{h}(t,\theta_1)| \geq C_1|\theta-\theta_1|\right\}\right) \leq C_2\delta,$$

where $\nu(\cdot)$ denotes the Lebesgue measure.

(6).
$$\int_0^\tau |\dot{H}(t,\theta) - \dot{H}(t,\theta_1)|dt \to 0 \quad \text{as } \theta \to \theta_1.$$

It follows from the conditions (2) and (5) that

$$\int_0^\tau |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)|dt \to 0 \quad \text{as } \theta \to \theta_1. \tag{2.1}$$

Under these conditions, the M-estimator $\widehat{\theta}_T$ is near by θ_1 with high probability for sufficiently large T. More precisely, the M-estimator $\widehat{\theta}_T$ is consistent in the following sense.

Theorem 5.1. *For any $T \geq 0$, there exist a positive number $\delta(T)$ ($\to 0$ as $T \to \infty$) and an event $A(T)$ such that $P(A(T)) \to 1$ as $T \to \infty$ and an M-estimator $\widehat{\theta}_T$ exists in $U(\delta(T))$ on the event $A(T)$, where $U(\delta) = \{\theta; |\theta-\theta_1| \leq \delta\}$.*

Before proving this theorem, we shall make auxiliary statements. Let

$$m(T,\theta) = \frac{1}{T}\int_0^T h(t,\theta)\,(dN(t) - \mu(t)dt),$$

$$G(T,\theta) = \frac{1}{T}\int_0^T \{H(t,\theta) - h(t,\theta)\mu(t)\}dt,$$

$$\dot{m}(T,\theta) = \frac{1}{T}\int_0^T \dot{h}(t,\theta)\,(dN(t) - \mu(t)dt)$$

and

$$\dot{G}(T,\theta) = \frac{1}{T}\int_0^T \{\dot{H}(t,\theta) - \dot{h}(t,\theta)\mu(t)\}dt.$$

From the law of large numbers for martingales and the conditions (2) and (3), we have that

$$m(T,\theta_1) \to 0 \qquad \text{in probability} \quad \text{as} \quad T \to \infty$$

and

$$G(T,\theta_1) \to 0 \qquad \text{as} \quad T \to \infty,$$

where θ_1 is given in the condition (3). Furthermore we obtain the following two lemmas about $\dot{m}$ and $\dot{G}$.

Lemma 5.2.

(2.2) $$\sup_{\theta\in U(\delta)} |\dot{G}(T,\theta) - \dot{G}(T,\theta_1)| \to 0 \quad \textit{as} \quad \delta \to 0 \quad \textit{uniformly in } T \geq \tau,$$

where $U(\delta) = \{\theta;\ |\theta - \theta_1| \leq \delta\}$.

Proof. We have that for any $T \geq \tau$,

$$\sup_{\theta\in U(\delta)} |\dot{G}(T,\theta) - \dot{G}(T,\theta_1)|$$

$$\leq \sup_{\theta\in U(\delta)} \left\{\frac{1}{T}\int_0^T |\dot{H}(t,\theta) - \dot{H}(t,\theta_1)|dt\right\} + \|\mu\|_\infty \sup_{\theta\in U(\delta)} \left\{\frac{1}{T}\int_0^T |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)|dt\right\}$$

$$\le \sup_{\theta\in U(\delta)} \left\{ \frac{2}{\tau}\int_0^{\tau} |\dot{H}(t,\theta)-\dot{H}(t,\theta_1)|dt \right\} + \|\mu\|_\infty \sup_{\theta\in U(\delta)} \left\{ \frac{2}{\tau}\int_0^{\tau} |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|dt \right\},$$

where $\|\mu\|_\infty = \sup|\mu(t)|$. By the conditions (1), (6) and (2.1), the right-hand side converges to 0 as $\delta \to 0$. Hence we obtain (2.2).

Hereafter, ε denotes any fixed positive number. Since $\dot{G}(T,\theta_1) \to \Gamma$ as $T \to \infty$, we have that there exists a $T_1 \ge \tau$ such that for any sufficiently small $\delta > 0$ and any $T > T_1$,

$$\text{(2.3)} \qquad \begin{aligned} \inf_{\theta\in U(\delta)} \dot{G}(T,\theta) &\ge \inf_{\theta\in U(\delta)} \{ \dot{G}(T,\theta_1) - |\dot{G}(T,\theta)-\dot{G}(T,\theta_1)| \} \\ &> \Gamma - 2\varepsilon, \end{aligned}$$

where Γ is a positive constant given in the condition (4).

Lemma 5.3. *We can find a constant C_3 such that for any sufficiently small δ, there exists a $T_2(\delta) \ge \tau$ such that for any $T > T_2(\delta)$,*

$$\text{(2.4)} \qquad P\left\{ \sup_{\theta\in U(\delta)} |\dot{m}(T,\theta)| \ge 4\varepsilon \right\} < \frac{2C_3}{\varepsilon}\delta.$$

Proof. We easily see that for any $T \ge \tau$,

$$\begin{aligned} &|\dot{m}(T,\theta) - \dot{m}(T,\theta_1)| \\ &\le \frac{1}{T}\int_0^T |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|\,dN(t) + \frac{1}{T}\int_0^T |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|\,\mu(t)\,dt \\ &\le \frac{1}{T}\int_{D_{\theta,T}} |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|\,dN(t) + \frac{1}{T}\int_{[0,T]-D_{\theta,T}} |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|\,dN(t) \\ &\quad + \frac{2\|\mu\|_\infty}{\tau}\int_0^{\tau} |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)|\,dt, \end{aligned}$$

where $D_{\theta,T} = \{t \in [0,T]\,;\, |\dot{h}(t,\theta)-\dot{h}(t,\theta_1)| \ge C_1|\theta-\theta_1|\}$ and C_1 is the constant given in the condition (5). By (2.1), there exists a $\delta_1 > 0$ such that for any $\theta \in U(\delta_1)$, the last term of the right-hand side is less than ε. Since $N(T)$ conforms to the Poisson distribution with mean $\int_0^T \mu(t)\,dt$, we have that for any $\delta > 0$ and any $T > 0$,

$$
\begin{aligned}
& P\left\{ \sup_{\theta \in U(\delta)} \frac{1}{T} \int_{[0,T]-D_{\theta,T}} |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)|\, dN(t) \geq \varepsilon \right\} \\
& \leq P\left\{ \frac{C_1 \delta}{T} N(T) \geq \varepsilon \right\} \\
& \leq \frac{C_1 \delta}{T\varepsilon} \int_0^T \mu(t)\, dt \\
& \leq \frac{C_1 \|\mu\|_\infty}{\varepsilon} \delta .
\end{aligned}
$$

For a measurable set B, let $N(B)$ denote $\int_B dN(t)$ which is the number of events occurring in B. Since $N(B)$ conforms to the Poisson distribution with mean $\int_B \mu(t)\,dt$, we see that for any $T \geq \tau$,

$$
\begin{aligned}
& P\left\{ \sup_{\theta \in U(\delta)} \frac{1}{T} \int_{D_{\theta,T}} |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)|\, dN(t) \geq \varepsilon \right\} \\
& \leq P\left\{ \frac{2\|\dot{h}\|_\infty}{T} \sup_{\theta \in U(\delta)} N(D_{\theta,T}) \geq \varepsilon \right\} \\
& \leq P\left\{ \frac{2\|\dot{h}\|_\infty}{T} N\left(\bigcup_{\theta \in U(\delta)} D_{\theta,T} \right) \geq \varepsilon \right\} \\
& \leq \frac{2\|\dot{h}\|_\infty}{T\varepsilon} E\left[N\left(\bigcup_{\theta \in U(\delta)} D_{\theta,T} \right) \right] \\
& \leq \frac{2\|\dot{h}\|_\infty \|\mu\|_\infty}{T\varepsilon} \frac{T+\tau}{\tau} \nu\left(\bigcup_{|\theta-\theta_1| \leq \delta} \{ t \in [0,\tau] \,;\, |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)| \geq C_1 |\theta - \theta_1| \} \right) ,
\end{aligned}
$$

where $\nu(\cdot)$ denotes the Lebesgue measure. From the condition (5), we get that for any sufficiently small $\delta > 0$ and any $T \geq \tau$,

$$P\left\{\sup_{\theta \in U(\delta)} \frac{1}{T} \int_{D_{\theta,T}} |\dot{h}(t,\theta) - \dot{h}(t,\theta_1)|\, dN(t) \geq \varepsilon\right\} \leq \frac{4\|\dot{h}\|_\infty \|\mu\|_\infty C_2}{\varepsilon \tau} \delta,$$

where C_2 is a constant given in the condition (5). Hence we have that for any sufficiently small δ and any $T \geq \tau$,

$$P\left\{\sup_{\theta \in U(\delta)} |\dot{m}(T,\theta) - \dot{m}(T,\theta_1)| \geq 3\varepsilon\right\} < \frac{C_3}{\varepsilon} \delta,$$

where $C_3 = (C_1 + \frac{4\|\dot{h}\|_\infty C_2}{\tau})\|\mu\|_\infty$. Since $\dot{m}(T,\theta_1)$ converges in probability to 0 as T tends to infinity, we obtain the conclusion of this lemma.

Proof of Theorem 5.1. From the condition (2), we have that for any $\theta \in \Theta$,

$$C(T,\theta) = C(T,\theta_1) + \int_{\theta_1}^{\theta} \dot{C}(t,u)du, \tag{2.5}$$

where

$$\dot{C}(T,u) = \int_0^T \dot{h}(t,u)dN(t) - \int_0^T \dot{H}(t,u)dt.$$

Let $A_1(T,\delta)$ be an event $\{\omega\,;\ \sup_{\theta \in U(\delta)} |\dot{m}(T,\theta)| < \frac{\Gamma}{3}\}$, where Γ is a positive constant given in the condition (4). Then we have that for any sufficiently small δ and any $T > T_2(\delta)$,

$$P(A_1(T,\delta)) \geq 1 - \frac{24C_3}{\Gamma}\delta\,(= 1 - C_4\delta,\ \text{say})$$

by using Lemma 5.3 for $\varepsilon = \frac{\Gamma}{12}$. We easily see that for any sufficiently small δ and $T > T_1$,

$$\inf_{u \in U(\delta)} \left\{-\frac{1}{T}\dot{C}(T,u)\right\} = \inf_{u \in U(\delta)} \left\{-\dot{m}(T,u) + \dot{G}(T,u)\right\} \tag{2.6}$$

$$> \frac{\Gamma}{2}$$

on the event $A_1(T,\delta)$ by using (2.3) for $\varepsilon = \frac{\Gamma}{12}$. Since $\frac{1}{T}C(T,\theta_1) = m(T,\theta_1) - G(T,\theta_1) \to 0$ in probability as $T \to \infty$, for any sufficiently small δ, there exists a $T_3(\delta)$ ($\geq T_2(\delta)$) such that for any $T > T_3(\delta)$,

$$P(A_2(T,\delta)) \geq 1 - 2C_4\delta,$$

where $A_2(T,\delta)$ denotes the event $\{\omega; \frac{|C(T,\theta_1)|}{T} \leq \frac{\Gamma\delta}{4}\} \cap A_1(T,\delta)$. From (2.5) and (2.6), we easily see that for any sufficiently small δ and $T > T_1$, $C(T,\theta_1+\delta) < 0$ and $C(T,\theta_1-\delta) > 0$ on the event $A_2(T,\delta)$, which implies that there exists a $\widehat{\theta}_T \in U(\delta)$ such that $C(T,\widehat{\theta}_T) = 0$. Consequently we obtain that for any sufficiently small $\delta > 0$, there exist an event $A_2(T,\delta)$ and a $T_0(\delta)$ ($= \max\{T_1, T_3(\delta)\}$) such that $P(A_2(T,\delta)) \geq 1 - 2C_4\delta$ for any $T > T_0(\delta)$ and an M-estimator $\widehat{\theta}_T$ exists in $U(\delta)$ on the event $A_2(T,\delta)$. We can take a monotone increasing sequence $\{T_n\}$ ($T_n \to \infty$) such that for any $T \geq T_n$, $P(A_2(T,\frac{1}{n})) \geq 1 - \frac{2C_4}{n}$. Hence we obtain the conclusion of Theorem 5.1 by setting $\delta(T) = \frac{1}{n}$ for $T_n \leq T < T_{n+1}$ and $A(T) = A_2(T,\delta(T))$.

We examine $C(T,\theta)$ to obtain asymptotic normality of the M-estimator $\widehat{\theta}_T$. From the condition (3), we get

$$\frac{1}{\sqrt{T}}C(T,\theta_1) = \frac{1}{\sqrt{T}}\int_0^T h(t,\theta_1)\,(dN(t) - \mu(t)\,dt) + o(1).$$

The first term of the right-hand side converges in distribution to the normal distribution $N(0,\Phi)$ by the central limit theorem for martingales, where Φ is a positive constant given in the condition (4).

On the other hand, we get

$$\frac{1}{\sqrt{T}}C(T,\theta_1)$$

$$= \sqrt{T}\int_{\widehat{\theta}}^{\theta_1} \frac{1}{T}\dot{C}(T,u)du$$

$$= \sqrt{T} \left(\int_{\widehat{\theta}}^{\theta_1} \overset{\bullet}{m}(T,u)du - \int_{\widehat{\theta}}^{\theta_1} \overset{\bullet}{G}(T,u)du \right)$$

$$= \sqrt{T} \left(\int_{\widehat{\theta}}^{\theta_1} \overset{\bullet}{m}(T,u)du - \int_{\widehat{\theta}}^{\theta_1} \{\overset{\bullet}{G}(T,u) - \overset{\bullet}{G}(T,\theta_1)\}du + (\widehat{\theta} - \theta_1)\overset{\bullet}{G}(T,\theta_1) \right),$$

where $\widehat{\theta} = \widehat{\theta}_T$ is a zero point of $C(T,\theta)$, that is, the M-estimator. We easily see that

$$\left| \int_{\widehat{\theta}}^{\theta_1} \overset{\bullet}{m}(T,u)du \right| \leq |\widehat{\theta} - \theta_1| \ \sup\left\{ |\overset{\bullet}{m}(T,\theta)| \, ; \, \theta \in U(|\widehat{\theta} - \theta_1|) \right\}$$

and

$$\left| \int_{\widehat{\theta}}^{\theta_1} \{\overset{\bullet}{G}(T,u) - \overset{\bullet}{G}(T,\theta_1)\}du \right|$$

$$\leq |\widehat{\theta} - \theta_1| \ \sup\left\{ |\overset{\bullet}{G}(T,\theta) - \overset{\bullet}{G}(T,\theta_1)| \, ; \, \theta \in U(|\widehat{\theta} - \theta_1|) \right\}.$$

By Lemmas 5.2, 5.3 and the previous theorem, we have that

$$\frac{1}{\sqrt{T}} C(T,\theta_1) = \sqrt{T}(\widehat{\theta} - \theta_1)(\Gamma + o_p(1)).$$

Consequently we obtain the following theorem.

Theorem 5.4. *The M-estimator $\widehat{\theta}_T$ is asymptotically normal:*

$$\sqrt{T}(\widehat{\theta}_T - \theta_1) \to N(0, \Phi\Gamma^{-2}) \quad \textit{in law} \quad \textit{as} \quad T \to \infty.$$

where Φ and Γ are positive constants given in the condition (4).

5.3. Minimax robust M-estimator

We shall estimate a phase parameter θ in a periodic intensity $\lambda(t,\theta) = f(t-\theta)$, where f is a C^2-class, strictly positive and even function with period 1 and $\theta \in \Theta = (-\frac{1}{2}, \frac{1}{2})$. For simplicity we assume that $\int_0^1 f(t)dt = 1$. We suppose that the score function $S(t,\theta)$ (=

$S(t-\theta)) = \frac{\partial}{\partial\theta}\log f(t-\theta)$ is concave in $t \in [\theta, \theta+\frac{1}{2}]$ and that $\int_0^1 S(t-\theta)^2 f(t-\theta)\,dt > 0$. Note that $\frac{\partial}{\partial\theta}S(t-\theta) = -\frac{\partial}{\partial t}S(t-\theta) = -S'(t-\theta)$.

The true intensity is given by

$$\mu(t) = (1-\varepsilon)f(t-\theta_0) + \varepsilon c(t-\theta_0), \tag{3.1}$$

where $\theta_0 \in \Theta, \varepsilon \in [0, M_\varepsilon], 0 < M_\varepsilon < 1$ and c is a periodic, even, bounded and measurable function. Without loss of generality we can assume $\theta_0 = 0$. Let $h(t,\theta) = \psi(t-\theta)$ and $H(t,\theta) = \varphi(t-\theta)$, where φ is odd and both ψ and φ are periodic functions for which the conditions (1)–(6) in the previous section hold with $\theta_1 = 0\,(=\theta_0)$. Then we easily see that Γ in the condition (4) is equal to $\int_0^1 \psi'(t)\,\mu(t)\,dt$. Hence the asymptotic variance is a measure of goodness of estimation and is related to ψ only.

We shall construct an M-estimator which has the minimax variance provided that the true intensity μ belongs to a suitable class. First we shall look for an intensity μ_0 minimizing the information $I(\mu)$ defined by

$$I(\mu) = \sup_{\psi\in Q_1} \frac{\left(\int_0^1 \psi'(t)\,\mu(t)\,dt\right)^2}{\int_0^1 \psi(t)^2\,\mu(t)\,dt}, \tag{3.2}$$

where Q_1 is a class of all periodic and continuously differentiable functions ψ with $\int_0^1 \psi(t)^2\,\mu(t)\,dt > 0$.

Putting, for $0 \le \beta \le \max_t S(t)$,

$$a = \inf\{t \in [0, \tfrac{1}{2}]\,;S(t) \ge \beta\} \tag{3.3}$$

and

$$b = \sup\{t \in [0, \tfrac{1}{2}]\,;S(t) \ge \beta\}, \tag{3.4}$$

we easily see that a and b are solutions of the equation $S(t) = \beta$ and differentiable with respect to β $(0 \le \beta < \max_t S(t))$ and that $\{t \in [0, \frac{1}{2}]; S(t) \ge \beta\} = [a, b]$ by concavity of $S(t)$ on $[0, \frac{1}{2}]$. Let

$$\mu_\beta(t) = \begin{cases} (1-M_\varepsilon)\, f(t) & \text{if } |t| \le a, \\ (1-M_\varepsilon)\, f(a) \exp\{-\beta(t-a)\} & \text{if } a \le |t| \le b, \\ (1-M_\varepsilon)\, \dfrac{f(a)}{f(b)} \exp\{-\beta(b-a)\} f(t) & \text{if } b \le |t| \le \frac{1}{2}, \\ \text{periodic} & \text{otherwise.} \end{cases} \tag{3.5}$$

Then its score function is given by

$$\begin{aligned} S_\beta(t) &= -\frac{\mu'_\beta(t)}{\mu_\beta(t)} \\ &= \max\{-\beta, \min\{S(t), \beta\}\} \\ &= \begin{cases} S(t) & \text{if } |t| < a,\ b < |t| \le \frac{1}{2}, \\ \beta & \text{if } a < t < b, \\ -\beta & \text{if } -b < t < -a, \\ \text{periodic} & \text{otherwise.} \end{cases} \end{aligned}$$

For a constant ξ $(> 1$, near by 1), we determine β by the equation

$$\int_0^1 \mu_\beta(t)\, dt = \xi,$$

equivalently

$$\begin{aligned} &\int_0^a f(t)dt + f(a) \int_a^b \exp\{-\beta(t-a)\}\, dt + \frac{f(a)}{f(b)} \exp\{-\beta(b-a)\} \int_b^{1/2} f(t)dt \\ &= \frac{\xi}{2(1-M_\varepsilon)}. \end{aligned} \tag{3.6}$$

Since the derivative of the left-hand side of (3.6) with respect to β is negative, it is decreasing in β and its maximal value and minimal value are $\frac{f(0)}{2}$ and $\frac{1}{2}$, respectively. Hence, for

$M_\varepsilon \in [0, 1 - \dfrac{\xi}{f(0)})$, the equation (3.6) has a unique solution β_0. Hereafter, we abbreviate the intensity μ_{β_0} and its score function S_{β_0} as μ_0 and S_0, respectively. Let $\mathcal{M}$ be a class of all periodic, even and measurable functions $\mu(t)$ satisfying that

$$\int_0^1 \mu(t)\, dt \leq \int_0^1 \mu_0(t)\, dt \, (= \xi),$$

and for any t,

$$(1 - M_\varepsilon) f(t) \leq \mu(t) \leq (1 - M_\varepsilon) \frac{f(a_0)}{f(b_0)} \exp\{-\beta_0 (b_0 - a_0)\} f(t)$$

where a_0 and b_0 respectively denote a and b given by (3.3) and (3.4) for $\beta = \beta_0$ and ξ is explained as the upper bound of the average number of the events which occur during one period in the case that the observation is contaminated. From $\xi > 1$ we easily see that the intensity f of the model belongs to the class $\mathcal{M}$ of the contaminated intensities.

We shall show the following lemma.

Lemma 5.5. *Under the condition*

$$\beta_0^2 + 2S'(t) - S(t)^2 \leq 0 \quad \text{for all } t \in [b_0, \tfrac{1}{2}], \tag{3.7}$$

μ_0 *minimizes the information, that is,*

$$I(\mu_0) = \min_{\mu \in \mathcal{M}} I(\mu). \tag{3.8}$$

Proof. As in Chapter 4 of Huber (1981), it is sufficient to check that for any μ_1 $(\in \mathcal{M})$ which satisfies that $I(\mu_1) < \infty$,

$$\frac{d}{ds} I(\mu_s)\Big|_{s=0} = \int_0^1 (2S_0'(t) - S_0(t)^2)\,(\mu_1(t) - \mu_0(t))\, dt \geq 0, \tag{3.9}$$

where $\mu_s = (1-s)\mu_0 + s\mu_1$, $s \in [0,1]$. We easily see that

$$\begin{aligned} \frac{1}{2}\frac{d}{ds} I(\mu_s)\Big|_{s=0} &= \int_0^{1/2} (\beta_0^2 + 2S_0'(t) - S_0(t)^2)\,(\mu_1(t) - \mu_0(t))\, dt \\ &\quad - \beta_0^2 \int_0^{1/2} (\mu_1(t) - \mu_0(t))\, dt \end{aligned}$$

$$\geq \int_0^{a_0} (\beta_0^2 + 2S'(t) - S(t)^2)\,(\mu_1(t) - \mu_0(t))\,dt$$

$$+ \int_{b_0}^{1/2} (\beta_0^2 + 2S'(t) - S(t)^2)\,(\mu_1(t) - \mu_0(t))\,dt$$

$$= \mathrm{I} + \mathrm{II} \text{ (, say).}$$

We see that $\mathrm{I} \geq 0$ because for $t \in [0, a_0], \beta_0^2 - S(t)^2 \geq 0, \mu_1(t) - \mu_0(t) \geq 0$ and $S'(t) \geq 0$. Since $\mu_1(t) - \mu_0(t) \leq 0$ for $t \in [b_0, \frac{1}{2}]$, we get $\mathrm{II} \geq 0$ by (3.7). Consequently, we obtain (3.9).

We shall show that the M-estimator corresponding to $h(t,\theta) = S_0(t-\theta)$ has the minimax asymptotic variance, that is,

$$\inf_{\psi \in \mathcal{Q}_2} \sup_{\mu \in \mathcal{M}} V(\mu, \psi) = V(\mu_0, S_0), \tag{3.10}$$

where $\mathcal{Q}_2$ a class of all periodic functions ψ with the Radon-Nikodym derivative ψ', for which the conditions (2)–(5) in the previous section hold with $\theta_1 = 0 (= \theta_0)$ and

$$V(\mu, \psi) = \frac{\displaystyle\int_0^1 \psi(t)^2 \mu(t)\,dt}{\left(\displaystyle\int_0^1 \psi'(t)\mu(t)\,dt\right)^2}$$

is the asymptotic variance of the M-estimator corresponding to $h(t,\theta) = \psi(t-\theta)$. From Lemma 4.4 in Huber (1981), $\dfrac{1}{V(\mu_s, S_0)}$ is convex in s, where $\mu_s = (1-s)\mu_0 + s\mu_1, s \in [0,1]$ and $\mu_1 \in \mathcal{M}$. We see that for any $\mu_1 \in \mathcal{M}$,

$$\frac{d}{ds}\left(\frac{1}{V(\mu_s, S_0)}\right)\Bigg|_{s=0} = \int_0^1 (2S_0'(t) - S_0(t)^2)\,(\mu_1(t) - \mu_0(t))\,dt \geq 0$$

because the last inequality follows from (3.9). Hence we have that

$$\begin{aligned} V(\mu_0, S_0) &= \sup_{\mu \in \mathcal{M}} V(\mu, S_0) \\ &\geq \inf_{\psi \in \mathcal{Q}_2} \sup_{\mu \in \mathcal{M}} V(\mu, \psi). \end{aligned} \tag{3.11}$$

We get

$$V(\mu_0, S_0) = \frac{1}{I(\mu_0)}$$
$$= \inf_{\psi_1 \in \mathcal{Q}_1} V(\mu_0, \psi_1)$$

(see e.g. Huber (1981)). For any $\psi_2 \in \mathcal{Q}_2$ and any $\varepsilon > 0$, we can find a $\psi_1 \in \mathcal{Q}_1$ such that $|\psi_2(t) - \psi_1(t)| < \varepsilon$ for all $t \in [0,1]$ from Weierstrass' theorem. For any $\psi_i \in \mathcal{Q}_i$ $(i = 1, 2)$, $\int_0^1 \psi_i'(t)\mu_0(t)\,dt = -\int_0^1 \psi_i(t)\mu_0'(t)\,dt$ because both ψ_i and μ_0 are periodic. Hence, for any $\psi_2 \in \mathcal{Q}_2$, we can approximate $V(\mu_0, \psi_2)$ by $V(\mu_0, \psi_1)$ for some $\psi_1 \in \mathcal{Q}_1$ from boundedness of ψ_1, ψ_2, μ_0 and μ_0'. Furthermore, we have that

$$V(\mu_0, S_0) = \inf_{\psi_1 \in \mathcal{Q}_1} V(\mu_0, \psi_1)$$
$$\leq \inf_{\psi_2 \in \mathcal{Q}_2} V(\mu_0, \psi_2)$$
$$\leq \inf_{\psi_2 \in \mathcal{Q}_2} \sup_{\mu \in \mathcal{M}} V(\mu, \psi_2)$$

From (3.11) and the above inequality, we obtain the following theorem.

Theorem 5.6. *Under the condition (3.7), the M-estimator corresponding to $h(t, \theta) = S_0(t - \theta)$ has the minimax asymptotic variance, that is,*

$$\inf_{\psi \in \mathcal{Q}_2} \sup_{\mu \in \mathcal{M}} V(\mu, \psi) = V(\mu_0, S_0),$$

where $\mu_0 = \mu_{\beta_0}$ is given by (3.5), β_0 is a unique solution of the equation (3.6) and $S_0(t) = -\dfrac{\mu_0'(t)}{\mu_0(t)}$.

In the first half of the present section, we have constructed the M-estimator which has the minimax variance provided that the true intensity μ belongs to the class $\mathcal{M}$. In the latter half, we shall consider the minimax problem when the true intensity $\mu(t)$ is given by (3.1). The class of all functions given by (3.1) is wider than the class $\mathcal{M}$ but we impose a restriction on the function $h(t,\theta)$. More precisely, for a fixed function ψ satisfying conditions below, the function h is given by

$$h(t,\theta) = \beta\psi\left(\frac{S(t-\theta)}{\beta}\right),$$

where β is a positive constant, $S(t-\theta) = \frac{\partial}{\partial\theta}\log f(t-\theta)$ is the score function and $\psi(x)$ is a piece-wise continuously differentiable, continuous, monotone increasing and odd function which is concave on the $[0,\infty)$. Then the asymptotic variance of the M-estimator is given by

$$V(c,\varepsilon,\beta) = \frac{\Phi(c,\varepsilon,\beta)}{\Gamma(c,\varepsilon,\beta)^2},$$

where

$$\Phi(c,\varepsilon,\beta) = \int_0^1 \beta^2\psi\left(\frac{S(t)}{\beta}\right)^2 \{(1-\varepsilon)f(t)+\varepsilon c(t)\}\,dt$$

and

$$\Gamma(c,\varepsilon,\beta) = \int_0^1 S'(t)\,\psi'\left(\frac{S(t)}{\beta}\right)\{(1-\varepsilon)f(t)+\varepsilon c(t)\}\,dt.$$

Let $\mathcal{C}$ be a class of all periodic, even and measurable functions c with $0 \le c(t) \le M_c$ for every t and $\mathcal{B}$ be a class of all β satisfying that for any $c \in \mathcal{C}$ and $0 \le \varepsilon \le M_\varepsilon$, $\Gamma(c,\varepsilon,\beta) > 0$, where M_ε is the bound of ε. We suppose that the class $\mathcal{B}$ is non-empty.

Our purpose is to determine the $\beta \in \mathcal{B}$ which minimizes $\max_{\substack{c\in\mathcal{C} \\ 0\le\varepsilon\le M_\varepsilon}} V(c,\varepsilon,\beta)$. First we shall show the following lemma.

Lemma 5.7. *For* $0 \le \alpha \le \frac{1}{2}$, *let*

$$c_\alpha(t) = \begin{cases} M_c & \alpha < |t| \le \frac{1}{2}, \\ 0 & |t| \le \alpha, \\ \textit{periodic} & \textit{otherwise.} \end{cases}$$

For $c = c_\alpha$, *we abbreviate the asymptotic variance* $V(c,\varepsilon,\beta)$ *as* $V(\alpha,\varepsilon,\beta)$. *Then for any*

$\beta \in \mathcal{B}$, there exists an $\alpha^* \in [0, \frac{1}{2}]$ such that

$$\max_{c \in \mathcal{C}} V(c, \varepsilon, \beta) = V(\alpha^*, \varepsilon, \beta). \tag{3.12}$$

Moreover, for any $\alpha \in [0, \frac{1}{2}]$ and $\beta \in \mathcal{B}$, $V(\alpha, \varepsilon, \beta)$ is monotone increasing in ε.

Proof. Since $V(\alpha, \varepsilon, \beta)$ is continuous in α, there exists an $\alpha_0 \in [0, \frac{1}{2}]$ such that

$$\max_{\alpha} V(\alpha, \varepsilon, \beta) = V(\alpha_0, \varepsilon, \beta).$$

It is sufficient to show that

$$\max_{c} V(c, \varepsilon, \beta) = V(\alpha_0, \varepsilon, \beta), \tag{3.13}$$

Let

$$\Psi(t) = \beta \psi \left(\frac{S(t)}{\beta} \right).$$

Since ψ is monotone increasing and concave on $[0, \infty)$ and $S(t)$ is concave on $[0, \frac{1}{2}]$, $\Psi(t)$ is concave on $[0, \frac{1}{2}]$. Hence $\Psi'(t)$ is decreasing on $[0, \frac{1}{2}]$. Accordingly, putting

$$t_0 = \sup\{t \in [0, \frac{1}{2}]\,;\, \Psi'(t) \geq 0\},$$

we have that $\Psi'(t) \geq 0$ for $t \in [0, t_0]$ and $\Psi'(t) \leq 0$ for $t \in [t_0, \frac{1}{2}]$. It is easy to check that

$$\frac{1}{2}\Phi(c, \varepsilon, \beta) = \int_0^{1/2} \Psi(t)^2 \{ (1-\varepsilon) f(t) + \varepsilon c(t) \} \, dt > 0, \tag{3.14}$$

because $\beta \in \mathcal{B}$. Since $S(t)$ is periodic and odd, we see that $S(0) = S(\frac{1}{2}) = 0$, which implies $\Psi(0) = \Psi(\frac{1}{2}) = 0$. From (3.14) there exists a $t_1 \in (0, \frac{1}{2})$ such that $\Psi(t_1) \neq 0$. It follows from concavity of Ψ that $\Psi(t_1) > 0$ and for any $t \in (0, \frac{1}{2})$, $\Psi(t) > 0$. Moreover we have that $0 < t_0 < \frac{1}{2}$.

Let

$$p(c) = \frac{1}{2} \int_0^1 \beta^2 \psi \left(\frac{S(t)}{\beta} \right)^2 c(t) \, dt$$

and

$$g(c) = \frac{1}{2}\int_0^1 S'(t)\,\psi'\left(\frac{S(t)}{\beta}\right) c(t)\,dt.$$

Then we easily check that

$$p(c) = \int_0^{1/2} \Psi(t)^2 c(t)\,dt$$

and

$$g(c) = \int_0^{1/2} \Psi'(t) c(t)\,dt.$$

For $c = c_\alpha$, we abbreviate $p(c)$ and $g(c)$ as $p(\alpha)$ and $g(\alpha)$, respectively. Since $p(\alpha)$ is continuous in α, for any function $c \in \mathcal{C}$, there exists an α_1 such that $p(c) = p(\alpha_1)$, equivalently,

$$\Phi(c, \varepsilon, \beta) = \Phi(c_{\alpha_1}, \varepsilon, \beta)\ (\geq 0)\,. \tag{3.15}$$

Furthermore we have that $g(c) \geq g(\alpha_1)$, equivalently,

$$\Gamma(c, \varepsilon, \beta) \geq \Gamma(c_{\alpha_1}, \varepsilon, \beta)\ (> 0)\,. \tag{3.16}$$

Indeed, if $\alpha_1 \in (0, t_0]$, we get

$$\begin{aligned} g(c) - g(\alpha) &= \int_0^{\alpha_1} \Psi'(t)c(t)dt + \int_{\alpha_1}^{1/2} \Psi'(t)\left(c(t) - M_c\right)dt \\ &\geq \int_0^{\alpha_1} \Psi'(t)c(t)dt + \int_{\alpha_1}^{t_0} \Psi'(t)\left(c(t) - M_c\right)dt \\ &\geq \Psi'(\alpha_1)\left[\int_0^{\alpha_1} c(t)dt + \int_{\alpha_1}^{t_0} \left(c(t) - M_c\right)dt\right], \end{aligned}$$

because Ψ' is monotone decreasing on $[0, \frac{1}{2}]$ and $\Psi'(t) \leq 0$ for $t \in [t_0, \frac{1}{2}]$. On the other hand, we have that

$$\begin{aligned} 0 &= p(c) - p(\alpha_1) \\ &\leq \int_0^{\alpha_1} \Psi(t)^2 c(t)dt + \int_{\alpha_1}^{t_0} \Psi(t)^2\left(c(t) - M_c\right)dt \end{aligned}$$

$$\leq \Psi(\alpha_1)^2 \left[\int_0^{\alpha_1} c(t)dt + \int_{\alpha_1}^{t_0} (c(t) - M_c)\, dt \right],$$

because Ψ is non-negative and monotone increasing on $[0, t_0]$. Since $\Psi(\alpha_1) > 0$ and $\Psi'(\alpha_1) \geq 0$ by definition of t_0, we obtain $g(c) - g(\alpha_1) \geq 0$. For $\alpha_1 \in [t_0, \frac{1}{2})$, we can similarly show that $g(c) \geq g(\alpha_1)$. We easily see that $g(c) = g(\alpha_1) = 0$ for $\alpha_1 = 0$ or $\frac{1}{2}$. Hence, for any $c \in \mathcal{C}$, we can find an α_1 satisfying (3.15) and (3.16). Consequently we obtain that

$$V(c, \varepsilon, \beta) \leq V(\alpha_1, \varepsilon, \beta) \leq V(\alpha_0, \varepsilon, \beta),$$

which implies (3.13).

We easily see that for any $\alpha \in [0, \frac{1}{2}]$ and $\beta \in \mathcal{B}$,

$$\Gamma(c_\alpha, \varepsilon, \beta) = 2 \int_0^{1/2} \Psi'(t)\{\, (1-\varepsilon) f(t) + \varepsilon c_\alpha(t) \,\}\, dt$$

$$= 2(1-\varepsilon) g(f) + 2\varepsilon g(\alpha) \ (> 0)$$

and

$$\frac{\partial}{\partial \varepsilon} \Gamma = -2g(f) + 2g(\alpha) \ (\leq 0),$$

because $g(f) \geq 0$ and $g(\alpha) \leq 0$. Similarly, we see that

$$\Phi(c_\alpha, \varepsilon, \beta) = 2(1-\varepsilon) p(f) + 2\varepsilon p(\alpha)$$

and

$$\frac{\partial}{\partial \varepsilon} \Phi = -2p(f) + 2p(\alpha).$$

Therefore

$$\frac{\partial}{\partial \varepsilon} V(\alpha, \varepsilon, \beta) = \left[\left(\frac{\partial}{\partial \varepsilon} \Phi \right) \Gamma - 2\Phi \left(\frac{\partial}{\partial \varepsilon} \Gamma \right) \right] \Big/ \Gamma^3$$

$$\geq \left[\left(\frac{\partial}{\partial \varepsilon} \Phi \right) \Gamma - \Phi \left(\frac{\partial}{\partial \varepsilon} \Gamma \right) \right] \Big/ \Gamma^3$$

$$= 4[p(\alpha)g(f) - p(f)g(\alpha)]/\Gamma^3$$
$$\geq 0.$$

Consequently we obtain the conclusion of the present lemma.

Let $\alpha^*(\beta)$ denote an α maximizing $V(\alpha, M_\varepsilon, \beta)$ and β_* denote a β minimizing $V(\alpha^*(\beta), M_\varepsilon, \beta)$, where M_ε is the bound of ε. From the previous lemma, we have that

$$V(\alpha^*(\beta_*), M_\varepsilon, \beta_*) = \min_{\beta \in \mathcal{B}} \max_{\substack{c \in \mathcal{C} \\ 0 \leq \varepsilon \leq M_\varepsilon}} V(c, \varepsilon, \beta).$$

As an example, let

$$f(x) = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-k)^2}{2\sigma^2}\right\} \qquad (\sigma = 0.1)$$

and

$$\psi(x) = \begin{cases} 1 & x > 1 \\ x & |x| \leq 1 \\ -1 & x < -1. \end{cases}$$

We give the tables of asymptotic variance V and asymptotic relative efficiency (ARE) which is asymptotic variance divided by one of the MLE in $M_\varepsilon = 0$.

Table 1. Model ($M_\varepsilon = 0$)

β	12.6	14.4	15.1	15.3	16.7
V($\times 10^2$)	1.07	1.05	1.05	1.04	1.03
ARE	.941	.960	.966	.967	.976

β	17.6	17.8	19.6	21.0	ML
V($\times 10^2$)	1.03	1.03	1.02	1.02	1.01
ARE	.981	.982	.989	.992	

Table 2. $M_\varepsilon = 0.01$

M_c	1		2		3	
	$\beta_* = 21.0$	ML	$\beta_* = 19.6$	ML	$\beta_* = 17.6$	ML
α^*	.142	.147	.143	.152	.143	.157
V($\times 10^2$)	1.06	1.09	1.09	1.17	1.13	1.25
ARE	.956	.927	.924	.863	.896	.807

M_c	4		5	
	$\beta_* = 16.7$	ML	$\beta_* = 15.1$	ML
α^*	.143	.161	.142	.166
V($\times 10^2$)	1.16	1.33	1.19	1.42
ARE	.870	.757	.846	.713

Table 3. $M_\varepsilon = 0.05$

M_c	0.5		1	
	$\beta_* = 17.8$	ML	$\beta_* = 15.3$	ML
α^*	.144	.155	.143	.167
V($\times 10^2$)	1.16	1.27	1.25	1.50
ARE	.869	.795	.807	.675

M_c	1.5		2	
	$\beta_* = 14.4$	ML	$\beta_* = 12.6$	ML
α^*	.144	.179	.126	.190
V($\times 10^2$)	1.34	1.73	1.42	1.97
ARE	.757	.584	.712	.512

Acknowledgments

I would like to express my gratitude to Professor N. Inagaki of Osaka University for his guidance, instructive discussions and constant encouragement throughout the present work. I also wish to thank him for his valuable advice and many useful comments. I am grateful to Professor N. Inagaki, Professor K. Ishii, Professor S. Shirahata and Dr. Y. Kano of Osaka University for their helpful suggestions and careful reading of earlier versions. Finally, I would like to thank Mr. N. Yoshida for his many helpful suggestions and exciting discussions.

References

Aven, T. (1985). A theorem for determining the compensator of a counting process, *Scand. J. Statist.* 12, 69-72.

Billingsley, P. (1965). *Ergodic Theory and Information*, Wiley, New York.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Bremaud, P. (1981). *Point Processes and Queues*, Springer Verlag, New York.

Brillinger, D. R. (1978). Comparative aspects of the study of ordinary time series and of point processes, in *Developments in Statistics Vol. 1*, ed. by Krishnaiah, P. R., Academic Press, New York.

Brown, T. C. (1981). Compensators and Cox convergence, *Math. Proc. Camb. Phil. Soc.* 90, 305-319.

Bustos, O. and Yohai, V. J. (1986). Robust estimators for ARMA models, *J. Amer. Statist. Assoc.* 81, 155-168.

Daley, D. J. (1971). Weakly stationary point processes and random measures, *J. Roy. Statist. Soc., Ser. B* 33, 406-428.

Daley, D. J. and Vere-Jones, D. (1972). A summary of the theory of point processes, in *Stochastic Point Processes*, ed. by Lewis, P. A. W., Wiley, New York.

Denby, I. and Martin, R. D. (1979). Robust estimation of the first order autoregressive parameter, *J. Amer. Statist. Assoc.* 74, 140-146.

Deng, Y. L. (1985). On the comparison of point processes, *J. Appl. Prob.* 22, 300-313.

Feigin, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes *Adv. Appl. Prob.* 8, 712-736.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications Vol. 2*, Wiley, New York.

Hampel, F. R. Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics*, Wiley, New York.

Hayashi, T. (1986). Laws of large numbers in self-correcting point processes, *Stoch. Proc. Appl.* 23, 319-326.

Hayashi, T. (1988). Local asymptotic normality in self-correcting point processes, in *Statistical Theory and Data Analysis II*, ed. by Matusita, K., North Holland, Amsterdam, New York.

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

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation, Asymptotic Theory*, Springer Verlag, New York.

Inagaki, N. and Hayashi, T. (1989). The estimation for the parameters of the simple self-correcting point process (to appear).

Isham, V. and Westcott, M. (1979). A self-correcting point process *Stoch. Proc. Appl.* 8, 335-347.

Ivanoff, G. (1982). Central limit theorems for point processes *Stoch. Proc. Appl.* 12, 171-186.

Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes,* Springer Verlag, New York.

Kabanov, Yu. M., Liptser, R. Sh. and Shiryaev, A. N. (1980). Some limit theorems for simple point processes (a martingale approach), *Stochastics* 3, 203-216.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes 2nd. ed.*, Academic Press, New York.

Kleiner, B., Martin, R. D. and Thomson, D. J. (1979). Robust estimation of power spectra, *J. Roy. Statist. Soc., Ser. B* 41, 313-351.

Künsch, H. (1984). Infinitesimal robustness for autoregressive processes, *Ann. Statist.* 12, 843-863.

Kurtz, T. G. (1974). Point processes and completely monotone set functions, *Z. Wahrsch.* 31, 57-67.

Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes,* Translated by B. S. L. Prakasa Rao, Heldermann Verlag, Berlin.

Liptser, R. Sh. and Shiryayev, A. N. (1977). *Statistics of Random Processes, Vol. 2*, Springer Verlag, New York.

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating ARMA models, in *Handbook of Statistics, Vol. 5*, 119-155.

Martin, R. D. and Yohai, V. J. (1986). Influence functionals for time series, *Ann. Statist.* 14, 781-818.

Ogata, Y. (1978). The asymptotic behavior of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math.* 30 A, 243-261.

Ogata, Y. and Vere-Jones, D. (1984). Inference for earthquake models: a self-correcting model, *Stoch. Proc. Appl.* 17, 337-347.

Rubin, I. (1972). Regular point processes and their detection, *IEEE Trans. Informat. Theory* IT-18, 547-557.

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

Tweedie, R. L. (1983a). Criteria for rates of convergence of Markov chains, with application to queuing and storage theory, in *Probability, Statistics and Analysis*, ed. by Kingman, J. F. C. and Reuter, G. E. H., Camb. Univ. Press, London.

Tweedie, R. L. (1983b). Existence of moments for stationary Markov chains, *J. Appl. Prob.* 20, 191-196.

Vere-Jones, D. (1988). On the variance properties of stress release models, *Austral. J. Statist.* 30 A, 123-135.

Vere-Jones, D. and Ogata, Y. (1984). On the moments of a self-correcting process, *J. Appl. Prob.* 21, 335-342.

Yoshida, N. (1988). Robust M-estimator in diffusion processes, *Ann. Inst. Statist. Math.* 40, 799-820.

Yoshida, N. and Hayashi, T. (1989). On the robust estimation in the Poisson processes with the periodic intensity, (submitted).

Author's Address

Department of Mathematics

College of Integrated Arts and Sciences

University of Osaka Prefecture

Sakai, Osaka 591

JAPAN