MAXIMUM LIKELIHOOD ESTIMATION FOR STATIONARY POINT PROCESSES. (U)
AUG 80  P TUAN, M L PURI
AFOSR-TR-80-0688
Summary. In this paper we derive the log likelihood function for point processes in terms of their stochastic intensities, using the martingale approach. For practical purposes we work with an approximate log likelihood function which is shown to possess the usual asymptotic properties of a log likelihood function. The resulting estimates are strongly consistent and asymptotically normal (under some regularity conditions). As a by-product, a strong law of large numbers and a central limit theorem for continuous martingale are derived.

---

Key words and phrases. Maximum likelihood, compensator, stochastic intensity, martingale, natural increasing process, point process, predictable process, likelihood function, stochastic integral.
Maximum Likelihood Estimation for Stationary Point Processes

Madan L. Puri and Pham-dinh Tuan

Indiana University
Department of Mathematics
Bloomington, Indiana 47405

Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332

August, 1980

27

Maximum likelihood; compensator, stochastic intensity, martingale, natural increasing process, point process, predictable process, likelihood function, stochastic integral.

In this paper we derive the log likelihood function for point processes in terms of their stochastic intensities, using the martingale approach. For practical purposes we work with an approximate log likelihood function which is shown to possess the usual asymptotic properties of a log likelihood function. The resulting estimates are strongly consistent and asymptotically normal (under some regularity conditions). As a by-product, a
strong law of large numbers and a central limit theorem for continuous martingale are derived.
1. **Introduction.** The maximum likelihood estimates for point process model have been occasionally used especially for Poisson model (see e.g. Snyder (1975)). The log likelihood function has also been used in some cases in a somewhat heuristic manner (see e.g. Snyder (1975), Vere-Jone(1975)). Although it is expected that the maximum likelihood estimates possess the usual asymptotic properties, to the best of our knowledge, there has not been a rigorous proof for it. In this paper, we use some of the ideas of Jacod (1975) and Lipster and Shirayev (1978) on martingale theory for point processes to derive the log likelihood function, and we prove under some regularity conditions the asymptotic properties of the log likelihood and the maximum likelihood estimators. The case of self exciting processes is of great interest since the log likelihood can be written down easily and hence is treated in greater detail.

2. **The Log Likelihood Function for Point Process.**

We shall be concerned with multitype point processes, that is, point events of \( r \) different types randomly occurring along the real line. This process can be described by a multivariate counting process \( N(t) = \{N_1(t), \ldots, N_r(t)\}, t \in \mathbb{R} \), defined on some probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \). Here \( N_j(t) - N_j(s), t > s \) denotes the number of point events of type \( j \) which occur in \( (s, t] \).

By convention \( N_j(0) = 0 \). We shall suppose that at each \( t \), at most one event regardless of its type can occur. Let \( \mathcal{A}_t, t \geq 0 \) be an increasing family of sub-\( \sigma \)-fields such that \( N(t) \) is \( \mathcal{A}_t \)-measurable, \( t > 0 \). Then (Lipster and Shirayev (1977),
p. 239) there is for each \( j = 1, \ldots, r \) a natural increasing process \( \Lambda_j(t) \), called the compensator of \( N_j(t) \), relative to \((A_t, P)\) such that \( m_j(t) = N_j(t) - \Lambda_j(t) \) \( t \geq 0 \) is a \( A_t\)-local martingale. Here we shall be interested only in the case where the measure \( d\Lambda_j(t) dP \) admits a density \( \lambda_j(t) \) with respect to \( dt dP \). The process \( \lambda_j(t) \) can be chosen to be \( A_t\)-predictable, that is \( (t, \omega) \rightarrow \lambda_j(t, \omega) \) is measurable with respect to the \( \sigma\)-field generated by all the \( A_t\)-adapted processes with left continuous sample paths, and is called the stochastic intensity of the \( N(t) \) process. Intuitively \( \lambda_j(t) \) can be interpreted as

\[
\lim_{\Delta t \to 0} P\{\Delta N_j(t) = 1 | A_t\} / \Delta t
\]

where \( \Delta N(t) = N(t + \Delta t) - N(t) \). Indeed if \( \tau_n^j \) is the time of occurrence of the \( n \)-th event of type \( j \) after the origin, then \( m_j(t \wedge \tau_n^j), t \geq 0 \) is a martingale and so for \( s > t, A \in A_t \),

\[
E[1_A \{N_j(s \wedge \tau_n^j) - N_j(t \wedge \tau_n^j)\}] = E\left[1_A \int_{t \wedge \tau_n^j}^{s \wedge \tau_n^j} d\Lambda_j(u)\right]
\]

\[
= \int_t^s E[1_A 1_{\{u \leq \tau_n^j\}} \lambda_j(u)] du .
\]

Heuristically if \( s = t + \Delta t \) and \( \Delta t \) is small, then the left hand side of (2.2) is approximately \( P(A \cap \{t \leq \tau_n^j\} \cap \{\Delta N(t) = 1\}) \), and the right hand side of (2.2) is approximately

\[
E[1_A 1_{\{t \leq \tau_n^j\}} \lambda(t)] \Delta t .
\]

Since \( \{\tau_n^j < t\} \in A_t \), we obtain the interpretation (2.1). Historically (2.1) was proposed as the definition of the stochastic intensity. This definition requires the existence of the limit in (2.1) and is equivalent to our
definition only under some regularity conditions.

If \( A_t = A_t \), the sub \( \sigma \)-field generated by \( N(s), 0 \leq s \leq t \), then the stochastic intensity, which we write now \( \lambda_j(t) \) completely defines the probability distribution of the process (Jacod (1975), Lipster and Shiryayev (1977), p. 252). Now let \( \pi \) be the probability such that the \( N_j(t) \), relative to \( \pi \), are independent Poisson processes with unit rate. Thus, relative to \( \pi \), the stochastic intensity of \( N_j(t) \) is one. Observe that the random measure \( \{j\} \times (s,t) \to \int_s^t \lambda_j(u) \, du \) on \( \{1, \ldots, r\} \times \mathbb{R}^+ \) is precisely the dual predictable projection in the sense of Jacod (1975) of the random measure \( \{j\} \times (s,t) \to N_j(t) - N_j(s) \), by the result of this paper, \( P \) is absolutely continuous with respect to \( \pi \) on any \( A_T, T > 0 \), with density

\[
\frac{dP}{d\pi} = \exp \left[ \int_0^T \sum_{j=1}^r \{ \log \lambda_j(t) \, dN(t) - \lambda_j(t) \, dt + dt \} \right].
\]

Let now \( \{P_\theta, \theta \in \Theta\} \) be a family of probability distributions on \((\Omega, \mathcal{A})\) and let \( \~\lambda_{\theta,j}(t) \) be the corresponding stochastic intensity of the \( N_j(t) \) process. Then the above result shows that the log likelihood function corresponding to an interval of observation \([0, T]\) is, up to an additive constant:

\[
(2.3) \quad L_T(\theta) = \int_0^T \sum_{j=1}^r \{ \log \~\lambda_{\theta,j}(t) \, dN_j(t) - \~\lambda_{\theta,j}(t) \, dt \}.
\]

Remark

The multitype process is a special case of the marked point process when the space of the marks is just \( \{1, \ldots, r\} \). There is no difficulty to write down the log likelihood function for marked point process. Also, it is not necessary to suppose the existence of the stochastic intensity to write down the log likelihood func-
tion; it suffices that the measure \( d\Lambda_{\theta,j}(t) dP_\theta \) be absolutely continuous with respect to some measure \( dv(t) dP_\theta \), \( \Lambda_{\theta,j}(t) \) being the natural increasing process of \( N_j(t) \) relative to \( \Lambda_t \) and to \( P_\theta \). The special case we considered is convenient for further developments concerning the asymptotic properties of the log likelihood function.

3. The Approximate Log Likelihood Function

The considered process is defined on the whole line although only an observation on \([0,T]\) is available. Denote by \( \lambda_{\theta,j}(t) \) the stochastic intensity of the \( N_j(t) \) process relative to \( P_\theta \) and the sub \( \sigma \)-fields \( A_t \), \( t \in \mathbb{R} \) generated by \( N(s), s \leq t \). Models of point processes are usually described in terms of \( \lambda_\theta(t) = (\lambda_{\theta,1}(t), ..., \lambda_{\theta,r}(t)) \). For example, the self exciting process introduced by Hawkes (1972) can be defined by

\[
(3.1) \quad \lambda_\theta(t) = \alpha_\theta + \int_{-\infty}^{t} g_\theta(t-s) dN(s)
\]

where \( \alpha_\theta \) is a constant vector and \( g_\theta(.) \) is some appropriate matrix function. Thus, it is desirable to obtain the log likelihood function in terms of \( \lambda_\theta(t) \). We are led to the problem of computing \( \tilde{\lambda}_\theta(t) \) in terms of \( \lambda_{\theta,j}(t) \). Now, from the interpretation (2.1) of \( \lambda_{\theta,j}(t) \) one can expect that

\[
\tilde{\lambda}_{\theta,j}(t) = E_\theta(\lambda_{\theta,j}(t)|A_t)
\]

in case when \( \lambda_{\theta,j}(t) \) is integrable.

The rigorous result is

**Theorem 1.** Let \( \tau_n^j \) be the time of occurrence of the \( n \)-th event of type \( j \) after the origin. Then for almost all \( t \), the random
variable \( l_{\{ t \leq \tau_n^j \}} \lambda_{\theta,j}(t) \) is \( P_\theta \)-integrable and

\[
l_{\{ t \leq \tau_n^j \}} \lambda_{\theta,j}(t) = E_\theta[l_{\{ t \leq \tau_n^j \}} \lambda_{\theta,j}(t) | \tilde{A}_t]
\]

almost surely.

Proof.

From (2.2), we have for all \( s > 0 \)

\[
n \geq E_\theta[N_j(s \wedge \tau_n^j)] = \int_0^s E_\theta[l_{\{ t \leq \tau_n^j \}} \lambda_{\theta,j}(t)] dt
\]

and hence the function

\[
(t, \omega) \mapsto X_n(t, \omega) = l_{\{ t \leq \tau_n^j \}} \lambda_{\theta,j}(\omega)
\]

belongs to \( L^1(R^+ \times \Omega, B_{R^+} \otimes A, dt \, dP_\theta) \). We shall show that there exists a \( \tilde{A}_t \)-predictable process \( \tilde{X}_n(t) \) with \( \tilde{X}_n(t) = E[X_n(t) | \tilde{A}_t] \) for almost all \( t \), almost surely. Indeed, there exists a sequence of simple functions of the form

\[
X_{n,k}(t, \omega) = \sum_{m=1}^{m_k} z_m^{(k)}(\omega) l_{(t_{m-1}, t_m]}(t)
\]

which converges in \( L^1 \) to \( X_n \). For \( t_{m-1}^{(k)} < t \leq t_m^{(k)} \), set \( \tilde{X}_{n,k}(t) \) to be a version of \( E_\theta[z_m^{(k)} | \tilde{A}_t] \), such that the process \( \tilde{X}_{n,k}(t) \) has left continuous sample path and is \( \tilde{A}_t \)-adapted, which is possible because of the martingale property of the \( E_\theta[z_m^{(k)} | \tilde{A}_t] \), \( t_{m-1}^{(k)} < t \leq t_m^{(k)} \). Hence \( \tilde{X}_{n,k}(t) \) is \( \tilde{A}_t \)-predictable.

From

\[
|\tilde{X}_{n,k}(t) - E_\theta[X_n(t) | \tilde{A}_t]| = |E_\theta[\{X_{n,k}(t) - X_n(t)\} | \tilde{A}_t]|
\]

\[
\leq E_\theta[|X_{n,k}(t) - X_n(t)| | \tilde{A}_t]
\]
and the fact that $X_n,k + X_n$ in $L^1$ as $k \to \infty$, we get that the sequence $X_n,k$, $k \geq 1$ is a Cauchy sequence in $L^1$ and hence converges to some $\tilde{\lambda}_t$-predictable process $\tilde{X}_n$ which equals $E_\theta\{X_n(t) | \tilde{\lambda}_t\}$ almost surely for almost all $t$.

Now, from the definition of $X_n(t)$ and the fact that $\{t \leq \tau^n_j\} \in \tilde{\lambda}_t$, we have, almost surely, \[ 1_{\{t \leq \tau^n_j\}} E_\theta\{X_m(t) | \tilde{\lambda}_t\} = E_\theta\{X_n(t) | \tilde{\lambda}_t\} \quad \text{for } m \geq n \]
and hence there is a $\tilde{\lambda}_t$-predictable process $\tilde{X}(t)$ such that \[ 1_{\{\tau_j^n > t\}} \tilde{X}(t) = \tilde{X}_n(t) = E_\theta[1_{\{\tau_j^n > t\}} \lambda_{\theta,j}(t) | \tilde{\lambda}_t] \]
for almost all $t$, almost surely.

We now show that $\tilde{X}(t)$ is $\tilde{\lambda}_{\theta,j}(t)$. For this let $A \in \tilde{\lambda}_t$, then the right hand side of (2.2) is equal to
\[
E_\theta[\int_t^s 1_A X_n(u) \, du] = \int_t^s E_\theta[1_A E_\theta\{X_n(u) | \tilde{\lambda}_u\}] \, du \\
= \int_t^s E_\theta[1_A \tilde{X}_n(u)] \, du \\
= E_\theta[1_A \int_t^s \tilde{X}(u) \, du].
\]
Hence, by (2.2) the process $N(t) = \int_0^t \tilde{X}(u) \, du$, $t \geq 0$ is a local martingale, which gives the desired result.

**Corollary.** Suppose that $\lambda_{\theta,j}(t)$ is integrable for almost all $t$. Then $\tilde{\lambda}_{\theta,j}(t) = E_\theta\{\lambda_{\theta,j}(t) | \tilde{\lambda}_t\}$ almost surely, for almost all $t$.

Although the above result provides a means of computing $\tilde{\lambda}_\theta(t)$
in terms of $\lambda_\theta(t)$, the actual computation is not easy. So we are led to approximate $\lambda_\theta(t)$ by some $\hat{\lambda}_\theta(t)$ which depend only on $N(s), 0 \leq s \leq t$. The approximate log likelihood function is then

$$L_T(\theta) = \int_0^T \left\{ \sum_{j=1}^r \log \hat{\lambda}_{\theta,j}(t) \, dN_j(t) - \hat{\lambda}_{\theta,j}(t) \, dt \right\}$$

Since $T^{-1} \tilde{L}_T(\theta)$ depends essentially on the values of $\hat{\lambda}_\theta(t)$ for large $t$ if $T$ is large, we would expect that $L_T$ is a good approximation to $\tilde{L}_T$ for large $T$ if $\hat{\lambda}_\theta(t)$ is a good approximation to $\tilde{\lambda}_\theta(t)$ for large $t$. But by the corollary of Theorem 1 and the stationarity of $N(t)$,

$$E \| \tilde{\lambda}_\theta(t) - \lambda_\theta(t) \| \to 0 \text{ as } t \to \infty.$$ Therefore one could expect that $L_T$ is a good approximation to $\tilde{L}_T$ for large $T$ if $\hat{\lambda}_\theta(t)$ is a good approximation to $\lambda_\theta(t)$ for large $t$. We will make our assumptions on $\lambda_\theta(t)$ precise later on.

As example, consider the self exciting process when $\lambda_\theta(t)$ is given by (3.1). In order that $\lambda_{\theta,j}(t) \geq 0$ for all $t$, we shall assume that $\alpha_{\theta,j} \geq 0, g_{\theta,j}(t) \geq 0$ for all $t$. If the $g_{\theta,j}$ are integrable and $N(t)$ is of stationary increments with $EN_j(1) = \nu_{\theta,j} < \infty$, then the integral in (3.1) exists almost surely. Indeed

$$E_{\theta} \left\{ \int_t^\infty g_{\theta,j}(t-s) \, dN_j(s) \right\} = \int_{-\infty}^t g_{\theta,j}(t-s) \nu_{\theta,j} \, ds$$

$$\leq \nu_{\theta,j} \int_{-\infty}^\infty g_{\theta,j}(t) \, dt < +\infty$$

and hence the integral in the above left hand side is finite almost surely. As an approximation to $\lambda_\theta(t)$, one might consider

$$\hat{\lambda}_\theta(t) = \alpha_\theta + \int_0^t g_\theta(t-s) \, dN(s)$$
which is evidently \( \bar{A}_t \)-measurable and would be a good approximation to \( \lambda_\theta(t) \) for large \( t \) if \( g_\theta(t) \to 0 \) sufficiently fast.

4. **Asymptotic Properties of the Approximate Log Likelihood**

From now on, we shall suppose that the \( N(t) \) process is of stationary increments and metrically transitive in the sense of Doob (1953, p. 510). Let \( \tilde{\Omega} \) be the space of \( \tilde{\omega} = (\tilde{\omega}_1, ..., \tilde{\omega}_r) \) where the \( \tilde{\omega}_i \) are non-decreasing integral valued functions on \((-\infty, \infty)\), \( \tilde{\mathcal{A}} \) the \( \sigma \)-field generated by the projections \( \pi_{st} : \tilde{\omega} \mapsto \tilde{\omega}(t) - \tilde{\omega}(s), s < t \), \( \mathcal{P}_\theta \) the restriction to \( \tilde{\mathcal{A}} \) of the image of \( \mathcal{P}_\theta \) by the application \( N : \omega \mapsto N(.) \), and \( T_h \) the shift operator \( (T_h \tilde{\omega})(t) = \tilde{\omega}(t + h) \). Then stationarity means that \( T_h \) conserves the probability \( \mathcal{P}_\theta \), that is \( \mathcal{P}_\theta(T_h \mathcal{A}) = \mathcal{P}(\mathcal{A}) \), \( \forall \mathcal{A} \in \tilde{\mathcal{A}} \), and metric transitivity means that the invariant sets, that is, those sets \( \mathcal{A} \in \tilde{\mathcal{A}} \) for which \( T_h^{-1} \mathcal{A} = \mathcal{A} \), a.s., have probabilities zero or one. Now, from the fact that \( N(t) - N(s) = \pi_{st} \circ N \) and that \( T_h \) conserves the probability \( \mathcal{P}_\theta \), one can show that \( \lambda_\theta(t) = \overline{\lambda}_\theta(t) \circ N \) with \( \overline{\lambda}_\theta(t + h) = \overline{\lambda}_\theta(t) \circ T_h \). Hence if \( \lambda_\theta(t) \) is integrable

\[
\lim_{T \to \infty} \frac{1}{T} \int_0^T \lambda_\theta(t) \, dt = \mathbb{E}\{\lambda_\theta(0)\},
\]

almost surely. (see Doob, 1953, p. 515). Here the expectation is computed with respect to the true probability.

We suppose in the sequel that \( 0 \in \mathbb{R}^k \) and \( \lambda_\theta(t), \overline{\lambda}_\theta(t) \)
are twice continuously differentiable with respect to $\theta$, almost surely. We shall use the notation $x^{(1)}_\theta$, $x^{(2)}_\theta$ to denote the vector and the matrix of first and second derivatives of $x_\theta$ with respect to $\theta$. Let $\psi(t)$ be a function of the $\lambda^i_\theta,j(t)$, $\lambda^{(i)}_\theta,j(t)$, $i = 1,2$ such that $\psi(t)$ is integrable. Then by the same argument as above

\begin{equation}
\frac{1}{T} \int_0^T \psi(t) dt \rightarrow E\psi(0), \text{ a.s. } (T \to \infty)
\end{equation}

In this section, we are interested in the limiting behavior of $L_T(\theta)$, $L^{(i)}_T(\theta^*)$, $i = 1,2$ as $T \to \infty$, $\theta^*$ being the true value of $\theta$. We have, omitting the subscript $\theta$ when $\theta = \theta^*$ and putting $\hat{\phi}_{\theta,j}(t) = \log \hat{\lambda}_{\theta,j}(t)$,

\begin{equation}
dm^j = dN^j(t) - \lambda^j(t) dt:
\end{equation}

\begin{equation}
L_T(\theta) = \sum_{j=1}^{r} \{ \int_0^T \hat{\phi}_{\theta,j}(t) dN^j(t) - \int_0^T \hat{\lambda}_{\theta,j}(t) dt \}
\end{equation}

\begin{equation}
= \sum_{j=1}^{r} \left[ \int_0^T \hat{\phi}_{\theta,j}(t) dm^j(t) + \int_0^T \{ \hat{\phi}_{\theta,j}(t) \lambda^j(t) - \hat{\lambda}^j(t) \} dt \right]
\end{equation}

\begin{equation}
L^{(1)}_T(\theta) = \sum_{j=1}^{r} \{ \int_0^T \hat{\phi}^{(1)}_{\theta,j}(t) dN^j(t) - \int_0^T \hat{\lambda}^{(1)}_{\theta,j}(t) dt \}
\end{equation}

\begin{equation}
= \sum_{j=1}^{r} \left[ \int_0^T \hat{\phi}^{(1)}_{\theta,j}(t) (\lambda^j(t) - \hat{\lambda}_{\theta,j}(t)) dt + \int_0^T \hat{\phi}^{(1)}_{\theta,j}(t) dm^j(t) \right]
\end{equation}
In order to use the result (4.1) we are led to replace
\( \hat{\lambda}_\theta, j(t), \hat{\phi}_\theta, j(t), \ldots \) by \( \lambda_\theta, j(t), \phi_\theta, j(t) = \log \lambda_\theta, j(t), \ldots \).
The following result, which can be easily proved, is quite useful.

**Lemma 1.** Let \( \varepsilon_t, x_t \) be such that \( \varepsilon_t/x_t \to 0 \) as \( t \to \infty \). Also let \( \int_0^T |\varepsilon_t| d\mu(t) < \infty \) and \( \mu(T)^{-1} \int_0^T |x_t| d\mu(t) \) be bounded where \( \mu \) is a non-decreasing function with \( \mu(\infty) = \infty \). Then \( \mu(T)^{-1} \int_0^T \varepsilon_t d\mu(t) \to 0 \) as \( T \to \infty \).

To obtain convergence results for the stochastic integrals with respect to \( dm_j(t) \) in (4.2) - (4.4), we will need the following result which is of independent interest.

**Lemma 2.** Let \( M_t, t \geq 0 \) be a locally square integrable martingale with continuous natural increasing process \( \langle M \rangle_t \). Let \( g_t, t > 0 \) be a non-decreasing left continuous function with \( g_\infty = \infty \), such that \( \langle M \rangle_t = O(g_t^{2-C}) \), \( c > 0 \), almost surely as \( t \to \infty \). Then \( M_t/g_t \to 0 \) almost surely as \( T \to \infty \).
Proof

Let $a > 0$ be such that $g_a > 0$. Let $c$ be an arbitrary positive constant. Define the stopping time

$$
\tau = \inf \{ t : \int_a^t g_s^{-2} \, d\langle M \rangle_s \geq c \}.
$$

Set $M_t = M_{\tau \wedge T}$, and

$$
Z_t = \int_a^t g_s^{-1} \, dM_s = \int_a^\tau g_s \, dM_s.
$$

The process $Z_t$, $t \geq a$ is a martingale with natural increasing process

$$
<Z>_t = \int_a^{\tau \wedge T} g_s^{-2} \, d\langle M \rangle_s \leq c, \; t \geq a.
$$

Since $EZ_t^2 \leq c$, $t \geq a$, we know (Doob, 1953, p. 354, 361) that almost surely $Z_{a+}$, $Z_{t-}$, $Z_{t+}$, $t \geq a$, $Z_\infty$ exist and $Z_t$ is bounded on any finite interval. Set $t^{(k)}_i = a + (T - a) i/k$ and write

$$
g_T Z_T = \sum_{i=1}^k \left[ \left( g_{t^{(k)}} - g_{t^{(k+1)}} \right) \int_{t^{(k+1)}}^{t^{(k)}} Z_s \, ds + g_{t^{(k)}} \left( \int_{t^{(k-1)}}^{t^{(k)}} Z_s \, ds - \int_{t^{(k-1)}}^{t^{(k+1)}} Z_s \, ds \right) \right]
$$

$$
= \int_0^T z_t \, dg_t + \int_a^T g_t^{(k)} \, dZ_t
$$

where $z_t = z^{(k)}_{t_i}$, $g_t^{(k)} = g_{t^{(k)}}$ for $t^{(k)}_{i-1} \leq t < t^{(k)}_i$.

By the Lebesgue dominated convergence Theorem and the property of stochastic integral, almost surely as $k \to \infty$

$$
\int_a^T z_t^{(k)} \, dg_t + \int_a^T z_t^{+} \, dg_t
$$
and
\[ \int_a^T g_t^{(k)} \, dt + \int_a^T g_t \, dt = \int_a^T \tilde{M}_t = \tilde{M}_T - \tilde{M}_a \]

Therefore
\[ g_T^{-1}(\tilde{M}_T - \tilde{M}_a) = Z_T - g_T^{-1} \int_a^T z_t^+ \, dt \]
\[ = Z_T - (1 - g_T g_T^{-1}) Z_{\infty} - g_T^{-1} \int_a^T (z_t^+ - z_{\infty}) \, dt \]

Since \( g_T \to \infty \) as \( T \to \infty \), from Lemma 1, we get \( g_T^{-1}(\tilde{M}_T - \tilde{M}_a) \to 0 \) and hence \( g_T^{-1} \tilde{M}_T \to 0 \), almost surely as \( T \to \infty \).

Now, we have \( \tilde{M}_t = M_t \) for all \( t \) on the set
\[ \{ \tau = \infty \} = \{ \int_d^\infty g_s^2 \, d\langle M \rangle_s \leq \sigma \} \]

Since \( c \) is arbitrary, we obtain \( g_T^{-1} \tilde{M}_T \to 0 \) almost surely as \( T \to \infty \) on the set
\[ \{ \int_d^\infty g_s^{-2} \, d\langle M \rangle_s < \infty \} \]

By assumption \( \langle M \rangle_s \leq \text{const.} \, g_s \), for all \( s \geq b \), for some \( b \geq a \), almost surely. So
\[ \int_b^\infty g_s^{-2} \, d\langle M \rangle_s \leq \text{const.} \int_b^\infty \frac{2^{-\varepsilon}}{2(\varepsilon-2)} \, \langle M \rangle_s \, dx < \infty \]

since the image of the measure \( d\langle M \rangle_s \) by the application \( s \to \langle M \rangle_s \) is just the Lebesgue measure. Hence, almost surely
\[ \int_a^\infty g_s^{-2} \, d\langle M \rangle_s < \infty \]

since \( g_s^{-2} \) is bounded on \( [a, b] \). The proof is completed.
Theorem 2. Suppose that

(i) $P(\lambda_{i,j}(t) \geq c) = 0$ for some $c > 0$ ($c$ is independent of $t$ by stationarity) and

$$\hat{\lambda}_{i,j}(t) - \lambda_{i,j}(t), \quad i = 1, 2$$

are differentiable with $\lambda_{i,j}(t)$ tending to zero almost surely as $t \to \infty$.

(ii) $\lambda_{i,j}(t)$ is integrable, and $\phi_{\theta,j}(t)$ and the elements of $\phi_{\theta,j}(t)$ are square integrable with respect to the measure $\lambda_{j}(t) \, d\P$.

Then almost surely as $T \to \infty$

$$T^{-1} L_T(\theta) \to E[ \sum_{j=1}^{\infty} \phi_{\theta,j}(0) \lambda_{j}(0) - \lambda_{\theta,j}(0)] = \Lambda(\theta)$$

$$T^{-1} L_T^{(1)}(\theta^*) \to 0$$

$$T^{-1} L_T^{(2)}(\theta^*) \to -E[ \sum_{j=1}^{\infty} \phi_{j}(1)(0) \phi_{j}(1)(0) \phi_{j}(0) = -J$$

Proof

Consider the right hand side of (4.2). By (i)

$$\hat{\phi}_{\theta,j}(t)\lambda_{j}(t) - \hat{\lambda}_{\theta,j}(t) = \phi_{\theta,j}(t)\lambda_{\theta,j}(t) - \lambda_{\theta,j}(t) + o(t)$$

where in this proof $o(t)$ denotes a quantity tending to zero almost surely as $T \to \infty$. By lemma 1 and the fact that

$\phi_{\theta,j}(t)\lambda_{j}(t)$ is integrable since $\phi_{\theta,j}(t)(\lambda_{j}(t))^{1/2}$ is square integrable, we obtain from (4.1) that $T^{-1}$ times the last term of the extreme right hand side of (4.2) tends to $\Lambda(\theta)$ almost surely as $T \to \infty$. On the other hand, $m_{j}(t)$ is a martingale with natural increasing process given by

$$d\langle m_{j} \rangle_t = \lambda_j(t) \, dt$$

Therefore
15.

\[ M_{j,t} = \int_0^t \hat{\phi}_{\hat{\theta},j}(t) \, dm_j(t) \]

is also a martingale with natural increasing process

\[ \langle M_j \rangle_t = \int_0^t \hat{\phi}_{\hat{\theta},j}(t) \lambda_j(t) \, dt \]

If \( \langle M_j \rangle_t = 0(t) \) almost surely as \( t \to \infty \), then by lemma 2, \( M_{j,T/T} \to 0 \) and hence \( T^{-1} L_T(\theta) + \Lambda(\theta) \) almost surely as \( T \to \infty \). Since from (i)

\[ \hat{\phi}_{\hat{\theta},j}(t) \lambda_j(t) = \phi_{\hat{\theta},j}(t) \lambda_j(t) + \phi_{\hat{\theta},j}(t) \lambda_j(t) \, o(t) + \lambda_j(t) \, o(t) \]

and therefore by (4.1) and Lemma 1, \( M_{j,T/T} \to E[\phi_{\hat{\theta},j}(0) \lambda_j(0)] \)

almost surely as \( T \to \infty \), we obtain the result

The proof for the convergence of \( T^{-1} L_T^{(1)}(\theta^*) \) and \( T^{-1} L_T^{(2)}(\theta^*) \) uses the same idea. We have from (i)

\[ \phi_j^{(1)}(t)(\lambda_j(t) - \hat{\lambda}_j(t)) = [\phi_j^{(1)}(t) + \phi_j^{(2)}(t) \, o(t) + o(t)] \, o(t) \]

\[ = \phi_j^{(1)}(t) \, o(t) + o(t) \]

\[ \phi_j^{(2)}(t)(\lambda_j(t) - \hat{\lambda}_j(t)) = [\lambda_j^{(2)}(t) / \hat{\lambda}_j(t) - \hat{\lambda}_j^{(1)}(t) / \phi_j^{(1)}(t)] \, o(t) \]

\[ = [\phi_j^{(2)}(t) + \lambda_j^{(2)}(t) \, o(t) + \phi_j^{(1)}(t) \, o(t) \]

\[ + \phi_j^{(1)}(t) \phi_j^{(1)}(t) \, o(t) + o(t)] \, o(t) \]

\[ = \phi_j^{(2)}(t) \, o(t) + \lambda_j^{(2)}(t) \, o(t) + \phi_j^{(1)}(t) \, o(t) + o(t) \]

Since \( \lambda_j(t) \) is bounded below, from (ii), \( \phi_j^{(1)}(t) \), \( \lambda_j^{(i)}(t) \), \( i = 1,2 \) are integrable. By lemma 1 and (4.1), the first term of the extreme right hand side of (4.3), (4.4), divided by \( T \)
tend to zero as $T \to \infty$.

Finally by a similar argument as above, the second terms in the extreme right hand side of (4.3) and (4.4) are $T_0(T)$ and the last term in the right hand side of (4.4) divided by $T$ converge to $-J$ almost surely as $T \to \infty$. The proof is completed.

**Remark**

Condition (i) of the theorem is introduced for convenience. The result might hold under weaker conditions. In fact, it suffices that

$$\frac{1}{T} \int_0^T \{ \hat{\lambda}_j(t) - \lambda_j(t) \} dt, \quad \frac{1}{T} \int_0^T \{ \hat{\phi}_{\theta,j}(t) - \phi_{\theta,j}(t) \} \lambda_j(t) dt,$$

$$\frac{1}{T} \int_0^T \hat{\phi}_j^{(i)}(t) \{ \hat{\lambda}_j(t) - \lambda_j(t) \} dt, \quad i = 1, 2,$$

$$\frac{1}{T} \int_0^T \{ \phi_j^{(1)}(t) \phi_j^{(1)}(t) \lambda_j(t) - \phi_j^{(1)}(t) \phi_j^{(1)}(t) \lambda_j(t) \} dt$$

tend to zero almost surely as $T \to \infty$ and

$$\frac{1}{T} \int_0^T \hat{\phi}_{\theta,j}^2(t) \lambda_j(t) dt, \quad \frac{1}{T} \int_0^T \| \hat{\phi}_j^{(i)}(t) \|^2 \lambda_j(t) dt, \quad i = 1, 2$$

are bounded almost surely, and condition (ii) to obtain the result.

Condition (i) is also not very restrictive. In case of the self-exciting process (3.1) with $g_{\theta,jk} \geq 0$, then $\lambda_{\theta,j}(t)$ is bounded below by a $\alpha_{\theta,j}(t)$, which we assume to be strictly positive. If we also assume that

$$\overline{g}_\theta(t) = \sup_{h>0} || g_\theta(t + h) ||$$

is integrable, then
\[ |\hat{\lambda}_\theta(t) - \lambda_\theta(t)| \leq \text{const.} \int_0^\infty \| g_\theta(t - s) \| \sum_{j=1}^r dN(s) \]

\[ \leq \text{const.} \int_0^\infty \overline{g}_\theta(t - s) \sum_{j=1}^r dN(s) \]

where the last integral is almost surely finite since it has a finite expectation and converges to zero as \( t \to \infty \) by the monotonous convergence theorem. In the same way, if \( g_\theta(t) \)

is twice differentiable with respect to \( \theta \) with derivatives satisfying the same condition as \( g_\theta(t) \) as above, then

\[ ||\hat{\lambda}_{\theta,j}^{(1)}(t) - \lambda_{\theta,j}^{(1)}(t)||, \quad i = 1, 2, \quad \text{tend to zero almost surely as} \quad t \to \infty. \]

Consider now the asymptotic distribution of \( T^{-\frac{1}{2}} L_T(\theta^*) \).

From (4.3), we expect that the above is asymptotically distributed like

\[ Z_T = T^{-\frac{1}{2}} \int_0^T \sum_{j=1}^r \phi_j^{(1)}(t) d\mu_j(t) \]

The asymptotic distribution of \( Z_T \) can be obtained from the following result, which is of independent interest.

\[ \text{Lemma 3. Let} \quad M_t, A_t, t \geq 0 \quad \text{be a square integrable martingale with natural increasing process} \quad \langle M_t \rangle \quad \text{satisfying} \quad d\langle M \rangle_t = X_t dt. \]

\[ \text{Suppose that there is a semi group of shift operator} \quad T_s, s \geq 0, \quad \text{conserving the probability and metrically transitive, such that} \]

\[ T_{-1}A_t = A_{t+h}, \quad (M_t - M_s) \circ T_h = M_{t+h} - M_s + h. \quad \text{Then as} \quad T \to \infty, \]

\[ T^{-\frac{1}{2}}M_T \quad \text{is asymptotically normal with zero mean and variance} \quad EX_0. \]

\[ \text{Proof.} \]

\[ \text{Let} \quad n_T \quad \text{be integers such that} \quad \Delta_T = T/n_T \to 0 \quad \text{as} \quad T \to \infty. \]
Set
\[ Y_{T,j} = T^{-\frac{1}{2}} (M_{j} \Delta T - M_{(j-1)} \Delta T) \]

Then from the result of Durett and Resnick (1978), as \( T \to \infty \),
\[ T^{-\frac{1}{2}} M_{T} = T^{-\frac{1}{2}} M_{0} + \sum_{j=1}^{n_{T}} Y_{T,j} \]
is asymptotically normal with zero mean and variance \( \text{EX}_{0} \), provided

(i) \[ \sum_{j=1}^{n_{T}} \text{E}(Y_{T,j}^{2} | A_{(j-1) \Delta T}) \to \text{EX}_{0} \quad \text{in probability} \]

(ii) \[ \sum_{j=1}^{n_{T}} \text{E}(Y_{T,j}^{2} 1\{Y_{T,j}^{2} > \varepsilon\}) \to 0, \quad \varepsilon > 0. \]

To verify (i), observe that the sum in (i) is equal to
\[ \frac{1}{T} \int_{0}^{T} X_{t}(T) \, dt = \langle M \rangle_{T} / T + \frac{1}{T} \int_{0}^{T} (X_{t}(T) - X_{t}) \, dt \]
where \( X_{t}(T) = E(X_{t} | A_{j \Delta T}) \) for \( j \Delta T \leq t < (j + 1) \Delta T \). Clearly
\[ X_{t+h} = X_{t} \circ T_{h} \quad \text{and} \quad E(X_{t+h} | A_{s+h}) = E(X_{t} | A_{s}) \circ T_{h}. \]
Therefore, as \( \Delta T \to 0 \),
\[ \sup_{0 \leq t \leq T} E|X_{t}(T) - X_{t}| \leq \sup_{0 \leq u \leq \Delta T} E|E(X_{1} | A_{1-u}) - X_{1}| + 0 \]
Note that \( X_{t} \) is \( A_{t} \)-measurable because this process is predictable. Thus, the sum in (i) differs from \( \langle M \rangle_{T} / T \) by a term
tending to 0 in the mean and since \( \langle M \rangle_{T} / T + \text{EX}_{0} \) by the metric transitivity of \( T_{h} \), we obtain the result.

To verify (ii), observe that the \( Y_{T,j}, \quad j = 1, \ldots, n_{T} \) have
the same distribution. Put \( \xi(t) = M_{t}^{2} - M_{0} \), we are led to verify that:
n_T[E(Y_T^2,1_{Y_T^2,1 > \epsilon})] = \Delta_T^{-1} E[\xi(\Delta_T) 1_{\{\xi(\Delta_T) > \epsilon T\}}]

tend to zero as \( T \to \infty \). This is true if \( \Delta_T \to 0 \) sufficiently slow. Indeed, \( \xi(\Delta), 0 \leq \Delta \leq 1 \) being a positive sub-martingale is uniformly integrable (Doob, 1953, p. 359), that is:

\[
\alpha(T) = \sup_{0 \leq \Delta \leq 1} E[\xi(\Delta) 1_{\{\xi(\Delta) > \epsilon T\}}] \to 0.
\]
as \( T \to \infty \). So all we have to do is to choose \( n_T \) such that \( \alpha(T)/\Delta_T \to 0 \) as \( T \to \infty \).

We now show that the difference between \( T^h L_T^{(0*)} \) and \( Z_T \), defined by (4.3), (4.5) tend to zero in probability as \( T \to \infty \).

**Lemma 4.** Let \( M_t, t \geq 0 \) be a locally integrable martingale with continuous natural increasing process \( <M>_t \). Let \( h_t \geq 0 \), if \( <M>_t/h_t^2 \to 0 \) in probability as \( T \to \infty \), then so is \( M_T/h_T \).

**Proof**

Let \( \epsilon > 0 \). Define the stopping time \( \sigma_T \) to be the value of \( s \) such that \( <M>_s = \epsilon h_T^2 \). Then as \( T \to \infty \),

\[
P\{M_T \sigma_T \neq M_T\} \to 0
\]
since \( P\{\sigma_T \leq T\} = P\{<M>_T \geq \epsilon h_T^2\} \to 0 \). On the other hand

\[
E(M_T^2 \sigma_T) = E<M>_T^2 \leq E<M>_{\sigma_T} = \epsilon h_T^2
\]

and hence by Tchebycheff inequality

\[
P\{|M_T/h_T| > \delta\} \leq P\{M_T \sigma_T \neq M_T\} + \epsilon/\delta^2
\]
from which the result follows.

Theorem 3. Under the condition of Theorem 2 and suppose that

(i) \[ T^{-\frac{1}{2}} \int_0^T \phi_j^{(1)}(t) \{ \lambda_j(t) - \lambda_j(t) \} dt \to 0 \]

in probability as \( T \to \infty \). Then \( T^{-\frac{1}{2}} L_T(\theta^*) \) is asymptotically normal with zero-mean and covariance matrix \( J \) given in Theorem 2.

Proof.

Let \( \alpha \) be a vector and \( Z_T \) be given by (4.5). Then
\[ M_T = T^{\frac{1}{2}} \alpha^* Z_T \], \( T \geq 0 \) is a martingale with natural increasing process
\[ <M>_t = \int_0^t \sum_{j=1}^r \{ \alpha^* \phi_j^{(1)}(s) \}^2 \lambda_j(s) ds \]

since (Lipster and Shiryayev, 1977, p. 269)
\[ M_t^2 = \int_0^t 2M_s dM_s + \sum_{s \leq t} (\Delta M_s)^2 \]
\[ = \int_0^t 2M_s dM_s + \sum_{j=1}^r (\alpha^* \phi_j^{(1)}(s))^2 dN_j(s) \]
\[ = \int_0^t [2M_s dM_s + \sum_{j=1}^r (\alpha^* \phi_j^{(1)}(s))^2 \ d\mu_j(s)] + <M>_t \]

where the first term of the last expression is a martingale and the second is a natural process. By Lemma 3, \( \alpha^* Z_T \) is asymptotically normal with zero mean and variance \( \alpha^* J \alpha \). On the other hand, by (i) and Lemma 4 with
\[ dM_t = \sum_{j=1}^r \alpha^* \phi_j^{(1)}(t) \{ \lambda_j(t) - \lambda_j(t) \} d\mu_j(t) \], \( h_t = \sqrt{t} \), we see that
\[ T^{-\frac{1}{2}} L_T^{(1)}(\theta^*) - Z_T \to 0 \] in probability as \( T \to \infty \). The result follows.
Remark

Condition (i) of Theorem 3 is satisfied if, for example, $E|\hat{\phi}^{(1)}_j(t)|^2$ is bounded and

\begin{equation}
\int_0^\infty (E|\hat{\lambda}_j(t) - \lambda_j(t)|^2)^{\frac{1}{2}} dt < \infty
\end{equation}

since by Schwartz inequality, the first absolute moment of the expression in (i) is bounded by

\begin{equation}
T^{-\frac{1}{2}} \int_0^T (E|\hat{\phi}^{(1)}_j(t)|^2 E|\hat{\lambda}_j(t) - \lambda_j(t)|^2)^{\frac{1}{2}} dt
\end{equation}

In case of the self exciting process with $\lambda_j(t)$, $\hat{\lambda}_j(t)$ given by (3.1), (3.3) we have

\begin{equation}
\hat{\lambda}_j(t) - \lambda_j(t) = \sum_k \int_{-\infty}^0 g_{jk}(t - s) dN_k(s)
= \sum_k \left\{ \int_{-\infty}^0 g_{jk}(t - s) \lambda_k(s) ds + \int_{-\infty}^0 g_{jk}(t - s) dm_k(s) \right\}
\end{equation}

Denote by $\| \cdot \|_2$ the $L^2$ norm, by the triangular inequality

\begin{equation}
\| \hat{\lambda}_j(t) - \lambda_j(t) \|_2 \leq \sum_k \int_0^\infty |g_{jk}(t - s)|^2 \| \lambda_k(s) \|_2 ds
+ \left\{ \sum_k \int_t^\infty g_{jk}(t - s) E\{\lambda_k(s)ds\}^2 \right\}^{\frac{1}{2}}
\end{equation}

Note that as in the proof of Theorem 3, the martingale $M_s$ defined by $dM_s = \sum_k g_{jk}(t - s) dm_k(s)$ has the natural increasing process given by $d\langle M \rangle_s = \sum_k g_{jk}^2(t - s) \lambda_k(s) ds$. Therefore, the process $\lambda_k(t)$ being stationary

\begin{equation}
\int_0^\infty \| \hat{\lambda}_j(t) - \lambda_j(t) \|_2 dt
\leq \text{const.} \sum_k \int_t^\infty |g_{jk}(s)| ds + \left\{ \int_t^\infty g_{jk}^2(s)ds \right\}^{\frac{1}{2}} dt
\end{equation}
Observe that

\begin{equation}
\int_0^\infty \int_t^\infty |g_{jk}(s)| \, ds \leq \int_0^\infty t |g_{jk}(t)| \, dt
\end{equation}

and, by Schwartz inequality, for any positive function \( h \) on \((0, \infty)\) which integrates to one

\[
\int_0^\infty \left\{ \int_t^\infty g_{jk}^2(s) \, ds \right\} \, dt \leq \left\{ \int_0^\infty \frac{1}{h(t)} \int_t^\infty g_{jk}^2(s) \, ds \, dt \right\}^k
\]

\[
= \int_0^\infty \left\{ tg_{jk}^2(t)/h(t) \right\} \, dt
\]

Take \( h(t) \sim t^{-1-\alpha}, \alpha > 0 \). Then (4.6) holds if the right hand size of (4.7) and \( \int_0^\infty t^{2+\alpha} g_{jk}^2(t) \, dt \) are finite. In the same way one can show that \( \| \hat{\lambda}_j^{(1)}(t) - \lambda_j^{(1)}(t) \|_2 \to 0 \) as \( t \to \infty \) if the \( g_{jk}(t) \) are integrable and square integrable. Hence \( \mathbb{E}[\| \phi_j^{(1)}(t) \|^2 \) are bounded if the above conditions hold and \( \lambda_j(t) \) are bounded below. Thus condition (i) of the Theorem is not restrictive.

5. Asymptotic Properties of the Maximum Likelihood Estimator

We are interested in the asymptotic properties of the estimate \( \theta_T \), which maximize \( L_T \) in \( \Theta \). We shall use this general result for which the proof is quite standard.

Theorem 4. 1. Let \( \Lambda_T \) be a random function on \( \Theta \subset \mathbb{R}^k \), satisfying

(i) \( \Lambda_T(\theta) \to \Lambda(\theta) \) almost surely as \( T \to \infty \), with \( \Lambda \) being continuous, admitting a unique maximum \( \theta^* \)

(ii) For any \( \theta \neq \theta^* \)
\[
\limsup_{T \to \infty} \sup_{\theta^* \in U(\theta)} \{ \Lambda_T(\theta^*) - \Lambda_T(\theta) \} \to 0 \quad \text{a.s.}
\]

as the neighborhood \( U(\theta) \) of \( \theta \) shrinks to \( \theta \).

Then any \( \theta_T \) realising the maximum of \( \Lambda_T \) on a compact \( C \) of \( \theta \) containing \( \theta^* \), converges almost surely to \( \theta^* \) as \( T \to \infty \).

2. Suppose that \( \theta^* \) is an interior point of \( \Theta \) and \( \Lambda \) admits continuous first and second derivatives with respect to \( \theta \), denoted by the vector \( \Lambda_T^{(1)} \) and the matrix \( \Lambda_T^{(2)} \), satisfying

(iii) As \( T \to \infty \), \( \Lambda_T^{(2)}(\theta^*) \to -J \) in probability and

\[
\sqrt{T} \Lambda_T^{(1)}(\theta) \quad \text{is asymptotically normal with zero mean and covariance matrix } J.
\]

(iv) For every \( \varepsilon > 0 \)

\[
\liminf_{T \to \infty} P \left[ |\Lambda_T^{(2)}(\theta) - \Lambda_T^{(2)}(\theta^*)| < \varepsilon \right] > 0, \forall \theta \in U(\theta^*)
\]

increases to 1 as the neighborhood \( U(\theta^*) \) of \( \theta^* \) shrinks to \( \theta^* \).

Then \( \hat{\theta}_T \) of (1) is asymptotically normal with mean \( \theta^* \) and covariance matrix \( T^{-1} J^{-1} \). Moreover, if \( \theta_T \) is \( T^{1/2} \)-consistent, that is the distributions of \( \sqrt{T} (\theta_T - \theta^*) \) are tight, then

\[
\hat{\theta}_T - \{ \theta_T - \Lambda_T^{(2)}(\theta_T)^{-1} \Lambda_T^{(1)}(\theta_T) \} \to 0 \quad \text{in probability as } T \to \infty.
\]

We apply the above results with \( \Lambda_T = T^{-1} \Lambda_T \). By Theorem 2, condition (i) is satisfied except the continuity of \( \Lambda(\theta) \), which we shall assume. To see that \( \theta^* \) realises the maximum of \( \Lambda(\theta) \), write

\[
\Lambda(\theta) - \Lambda(\theta^*) = \sum_{j=1}^k \left[ \mathbb{E} \log \left( \lambda_{\theta, j}(0) \lambda_j^{-1}(0) \right) \lambda_j(0) + \lambda_j(0) - \lambda_{\theta, j}(0) \right]
\]
and note that \( \log x \leq x - 1 \) with equality if and only if \( x = 1 \), we get \( \Lambda(\theta) \leq \Lambda(\theta^*) \) with equality if and only if \( \lambda_{\theta,j}(0) = \lambda_j(0) \) almost surely, implying, by stationarity \( \lambda_{\theta,j}(t) = \lambda_j(t) \) almost surely. If the parametrization is such that for \( \theta \neq \theta^* \), \( \lambda_{\theta,j}(t) \) is not equal to \( \lambda_{\theta^*,j}(t) \) for all \( t \), almost surely, then \( \theta^* \) is the unique maximum of \( \Lambda \).

Theorems 2 and 3 show that condition (iii) is satisfied. So all we need is to verify conditions (ii) and (iv). This would require rather strong assumptions on \( \lambda_{\theta}(t) \) and \( \lambda_{\theta}(t) \).

A sufficient set of assumptions is

**A0:** For any compact \( C \) of \( \theta \), there is a \( c > 0 \) such that \( \lambda_{\theta,j}(t) \geq c \) almost surely for all \( t \), all \( \theta \in C \).

**A1:** For any compact \( C \) of \( \theta^* \),

\[
\lim_{t \to \infty} \sup_{\theta \in C} |\lambda_{\theta,j}(t) - \lambda_j(t)| = 0 \text{ a.s.}
\]

\[
E \left\{ \sup_{\theta \in C} \phi_{\theta,j}^2(t) \lambda_j(t) \right\} < \infty
\]

**A2:** For some compact neighborhood \( U \) of \( \theta^* \),

\[
\lim_{t \to \infty} \sup_{\theta \in U} |\lambda_{\theta,j}(t) - \lambda_j(t)| = 0 \text{ a.s., } i = 1, 2,
\]

\[
E\{ \sup_{\theta \in U} |\phi_{\theta,j}^2(t)| \lambda_j(t) \} < \infty
\]

\[
E\{ \sup_{\theta \in U} |\lambda_{\theta,j}^{(2)}(t)| \} \leq \infty
\]

**Theorem 5.** Under the assumptions **A0, A1** condition (ii) of
Theorem 4 is satisfied, and under the assumptions $A_0, A_2$, condition (iv) of Theorem 4 is satisfied.

Proof.
Let $U(\theta)$ be a compact neighborhood of $\theta$. Then from (4.2),

$$\sup_{\theta' \in U(\theta)} T^{-1} L_T(\theta') \leq T^{-1} \sum_{j=1}^{r} \int_{0}^{T} \left\{ \sup_{\theta' \in U(\theta)} \phi_{\theta', j}(t) \right\} dN_{j}(t) \right\} \right\}$$

By the same argument as in the proof of Theorem 2, the above right hand side is seen to converge almost surely as $T \to \infty$ to

$$\sum_{j=1}^{r} \left\{ \mathbb{E} \left\{ \sup_{\theta' \in U(\theta)} \phi_{\theta', j}(t) \right\} \lambda_j(t) \right\} - \mathbb{E} \left\{ \inf_{\theta' \in U(\theta)} \lambda_{\theta', j}(t) \right\}$$

By the monotonous convergence theorem, as $U(\theta)$ shrinks to $\theta$ the above expression converges to $\Lambda(\theta)$ and hence

$$\lim_{T \to \infty} \sup_{\theta' \in U(\theta)} T^{-1} L_T(\theta') \to \Lambda(\theta) \text{ a.s.},$$

which gives the result.

To verify condition (iv) of Theorem 4, from (4.4) observe that $T^{-1} |L_T^{(2)}(\theta) - L_T^{(2)}(\theta*)|$ is bounded for all $\theta \in U$ by

$$T^{-1} \sum_{j=1}^{r} \int_{0}^{T} \left\{ \sup_{\theta \in U} \| \phi_{\theta', j}(t) - \phi_{j}(t) \| \right\} dN_{j}(t)$$

Again, by a similar argument, the above expression converges almost surely as $T \to \infty$ to
\[ \sum_{j=1}^{k} \mathbb{E} \sup_{\theta \in \mathcal{U}} \| \phi_{\theta, j}^{(2)}(t) - \phi_{j}^{(2)}(t) \| \lambda_{j}(t) \]
\[ + \mathbb{E} \{ \sup_{\theta \in \mathcal{U}} \| \lambda_{\theta, j}^{(2)}(t) - \lambda_{j}^{(2)}(t) \| \} \]

Note that we have used the fact that
\[ \{ \sup_{\theta \in \mathcal{U}} \| \phi_{\theta, j}^{(2)}(t) - \phi_{j}(t) \| \sqrt{\lambda_{j}(t)}, \sup_{\theta \in \mathcal{U}} \| \lambda_{\theta, j}(t) - \lambda_{j}(t) \| \] are square integrable and integrable which follows easily from A2. Again, by the monotonous convergence Theorem, the above expectations converge to zero as \( U \) shrinks to \( \theta^* \). The proof is completed.
REFERENCES


Université de Grenoble
Mathématiques Appliquées
38041 Grenoble cédex, France

Department of Mathematics
Indiana University
Bloomington, Indiana 47401