The Annals of Statistics
1987, Vol. 15, No. 2, 473–490

MAXIMUM LIKELIHOOD ESTIMATION IN THE MULTIPLICATIVE INTENSITY MODEL VIA SIEVES[1]

BY ALAN F. KARR

The Johns Hopkins University and University of North Carolina

For point processes comprising i.i.d. copies of a multiplicative intensity process, it is shown that even though log-likelihood functions are unbounded, consistent maximum likelihood estimators of the unknown function in the stochastic intensity can be constructed using the method of sieves. Conditions are given for existence and strong and weak consistency, in the L^1-norm, of suitably defined maximum likelihood estimators. A theorem on local asymptotic normality of log-likelihood functions is established, and applied to show that sieve estimators satisfy the same central limit theorem as do associated martingale estimators. Examples are presented. Martingale limit theorems are a principal tool throughout.

1. Introduction. The multiplicative intensity model introduced by Aalen (1978) has engendered numerous papers on "martingale methods" of estimation for point processes; Jacobsen (1982) and Karr (1986) contain expository treatments. The martingale method enjoys a number of important advantages: It is particularly suited (in some ways) for estimation of functions arising in point process inference problems, martingale estimators and their first and second moments are equally amenable to calculation and, in the case of the latter, estimation, and there exist powerful and widely applicable martingale inequalities and central limit theorems [especially those of Liptser and Shiryayev (1980) and Rebolledo (1980)] that can be used to establish consistency and asymptotic normality of martingale estimators. However, while the martingale method can be interpreted as a conditional form of the method of moments, it is not based on an optimality principle, nor does it estimate the unknown function itself, but rather indefinite integrals of it. Our purpose in this paper is to pursue some aspects of maximum likelihood estimation for the multiplicative intensity model. After describing the model in more detail, along with difficulties arising from frontal approaches to maximum likelihood estimation, we shall indicate briefly the nature of the results in the remaining sections.

For a general discussion of point processes on $\mathbb{R}_+$ with stochastic intensities, we refer to Brémaud (1981), Gill (1980), Jacobsen (1982), Karr (1986) and Liptser and Shiryayev (1978). Let $(\Omega, \mathscr{F}, P)$ be a probability space over which are defined a history $\mathscr{H} = (\mathscr{H}_t)_{t \geq 0}$ satisfying the "conditions habituelles" and an

Received December 1983; revised September 1986.

[1]Supported by the Air Force Office of Scientific Research, AFSC, USAF, under grant number 82-0029 and contract number F49620-82-C-0009. Reproduction in whole or in part is permitted for any purpose of the United States Government.

AMS 1980 *subject classifications.* Primary 62M09, 60G55; secondary 62F12, 62G05.

Key words and phrases. Point process, counting process, stochastic intensity, multiplicative intensity model, martingale, martingale estimator, maximum likelihood estimator, method of sieves, consistency, c_n-consistency, asymptotic normality, local asymptotic normality.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

adapted point process (counting process) $N = (N_t)_{t \geq 0}$. An $\mathscr{H}$-predictable, nonnegative process $\lambda = (\lambda_t)$ is the $(P, \mathscr{H})$*-stochastic intensity* of N if the process

$$M_t = N_t - \int_0^t \lambda_s \, ds \tag{1.1}$$

is a (local) $\mathscr{H}$-martingale, which we call the *innovation martingale* for N. (Heuristically, $\lambda_t \, dt = E[dN_t | \mathscr{H}_{t-}]$; for ways of making this precise, see the references cited above.) If N is integrable in the sense that $E[N_t] < \infty$ for each t, then M is a martingale and is square integrable over each bounded time interval. The predictable variation of M (cf. Section 2) satisfies

$$\langle M \rangle_t = \int_0^t \lambda_s \, ds. \tag{1.2}$$

(Interpretation: N is, locally and conditionally, a Poisson process in the sense that the conditional mean $\lambda_t \, dt = E[dN_t | \mathscr{H}_{t-}]$ equals the conditional variance $d\langle M \rangle_t = E[(dM_t)^2 | \mathscr{H}_{t-}] = E[(dN_t - E[dN_t | \mathscr{H}_{t-}])^2 | \mathscr{H}_{t-}]$.)

The *multiplicative intensity model* is a statistical model for point processes on $[0,1]$ in which the stochastic intensity contains an unknown deterministic function as a factor; it is formulated as follows. Let $(\Omega, \mathscr{F}, P)$, $\mathscr{H}$, N and λ be as in the preceding paragraph, but defined over $[0,1]$ rather than $\mathbb{R}_+$. The model is indexed by the set I of functions $\alpha \in L^1_+[0,1]$ that are left-continuous with right-hand limits, or by a subset. The family of candidate probability laws is $\mathscr{P} = \{P_\alpha \colon \alpha \in I\}$, where for each α the point process N has $(P_\alpha, \mathscr{H})$-stochastic intensity

$$\lambda_t(\alpha) = \alpha_t \lambda_t. \tag{1.3}$$

Note well that α is deterministic. One observes a single realization of both the point process N and the baseline stochastic intensity λ over $[0,1]$, and the objective is to draw inferences concerning the unknown function α.

Put succinctly, the martingale method of estimation is based on the dictum "martingale = noise;" in other words, one can estimate a given unobservable process by an observable process differing from it by a martingale. For the multiplicative intensity model, direct estimation of α is usually not attempted, because of the analogy between α and density or hazard functions; recently, though, Ramlau-Hansen (1983) has used kernel smoothing methods from density estimation to convert martingale estimators (see below) to estimators of α itself.

Instead of α, the martingale method estimates indefinite integrals $\int_0^t \alpha_s \, ds$, but since it is hopeless to ascertain information about α on the (random) set where λ is zero—no points of N occur there—one estimates the stochastic process

$$B_t(\alpha) = \int_0^t \alpha_s 1(\lambda_s > 0) \, ds, \tag{1.4}$$

where "1" denotes an indicator function. Under suitable assumptions, the difference

$$\int_0^t 1(\lambda_s > 0) \lambda_s^{-1} (dN_s - \alpha_s \lambda_s \, ds) = \int_0^t 1(\lambda_s > 0) \lambda_s^{-1} \, dN_s - B_t(\alpha),$$

as the integral of a predictable process with respect to the innovation martingale, is itself a martingale, which yields the *martingale estimator* (a slight misnomer: the error process, not the estimator, is a martingale)

$$\hat{B}_t = \int_0^t 1(\lambda_s > 0)\lambda_s^{-1}\, dN_s. \tag{1.5}$$

Throughout, integrals denoted by $\int_0^t$ are over $[0, t]$.

The following is one set of "suitable assumptions."

LEMMA 1.6. *Assume that there exists $\delta > 0$ such that $\lambda_s > \delta$ whenever $\lambda_s > 0$. Then for each α the error process $\hat{B} - B(\alpha)$ is a P_α-square integrable martingale with predictable variation*

$$\langle \hat{B} - B(\alpha)\rangle_t = \int_0^t 1(\lambda_s > 0)\lambda_s^{-1}\alpha_s\, ds. \tag{1.7}$$

The proof is given, e.g., in Aalen (1978) or Gill (1980).

In this paper, rather than estimate processes of the form (1.4), we estimate the unknown function α directly. More specifically, we are concerned with the following situation: There is a sequence of triples $(N^{(i)}, \lambda^{(i)}, \mathscr{H}^{(i)})$ on $[0,1]$ that, under each probability P_α, are independent and identically distributed (i.i.d.) with $N^{(i)}$ having $(P_\alpha, \mathscr{H}^{(i)})$-stochastic intensity $\lambda_t^{(i)}(\alpha) = \alpha_t\lambda_t^{(i)}$. Note that α does not vary with i. Our main results deal with the asymptotic properties of appropriately defined maximum likelihood estimators of α and of likelihood ratios.

Relevant properties of likelihood functions are collected in the following result. For the proof see Jacod (1975).

PROPOSITION 1.8. *Let P be a probability measure with respect to which the pairs $(N^{(i)}, \lambda^{(i)})$ are i.i.d. and $N^{(i)}$ has $(P, \mathscr{H}^{(i)})$-stochastic intensity $\lambda^{(i)}$. (One can take P to correspond to $\alpha \equiv 1$.) For each n let $N^n = \sum_{i=1}^n N^{(i)}$, $\lambda^n = \sum_{i=1}^n \lambda^{(i)}$, $\mathscr{H}_t^n = \bigvee_{i=1}^n \mathscr{H}_t^{(i)}$. Then*

(a) *N^n has $(P_\alpha, \mathscr{H}^n)$-stochastic intensity $\alpha\lambda^n$;*

(b) *each probability P_α is absolutely continuous with respect to P on $\mathscr{H}_1^n$, with Radon–Nikodym derivative*

$$dP_\alpha/dP = \exp\left[\int_0^1 \lambda_s^n(1 - \alpha_s)\, ds + \int_0^1 (\log \alpha_s)\, dN_s^n\right]; \tag{1.9}$$

(c) *(N^n, λ^n) is a sufficient statistic for α given the data $(N^{(1)}, \lambda^{(1)}), \dots, (N^{(n)}, \lambda^{(n)})$.*

We now define the log-likelihood functions

$$L_n(\alpha) = \int_0^1 \lambda_s^n(1 - \alpha_s)\, ds + \int_0^1 (\log \alpha_s)\, dN_s^n, \tag{1.10}$$

to whose study most of the paper is devoted. It is apparent from (1.10) that

$L_n(\alpha)$ is unbounded above as a function of α: Its value can be made arbitrarily large by choosing α to be large at each point of N^n and zero nearly everywhere else. In Section 3 we show that it is possible, nevertheless, to construct strongly consistent maximum likelihood estimators $\hat{\alpha} = \hat{\alpha}_n$ of α by employing the method of sieves due to Grenander and collaborators [for an application to density estimation see Geman and Hwang (1982)]. Convergence of $\hat{\alpha}$ to α takes place in the L^1-norm. In Section 4 we consider local asymptotic normality of differences of log-likelihood functions viewed as stochastic processes on $[0, 1]$, which leads in turn to asymptotic normality of the estimators $\hat{\alpha}$. Section 5 presents two examples.

Even for analysis of maximum likelihood estimators and likelihood ratios, martingale concepts and results remain central to our proofs. Thus, while maximum likelihood estimators may sometimes supplant martingale estimators, their respective theories overlap. Therefore in Section 2 we summarize martingale results needed in later sections. Section 3 also sheds light on the relationship between martingale and maximum likelihood estimation: The martingale estimator (1.5) is shown to arise by transformation of an estimator $\hat{\alpha}$ obtained by using a particular sieve. An alternative approach appears in Johansen (1983), where an extended model is formulated in which the estimators (1.5) have a direct maximum likelihood interpretation. Borgan (1984) treats the finite-dimensional case in detail, while Jacobsen (1984) is a general survey. McKeague (1986) discusses sieve methods for least-squares estimation in a semimartingale model incorporating a particular linear structure.

2. Martingale results. First, we summarize relevant results from the theory of martingales; for more detailed exposition we refer to Jacod (1979), where the setting is general, and to Gill (1980), Karr (1986) and Rebolledo (1980), which are oriented specifically to point processes. We consider only point processes on $[0, 1]$, defined over a probability space $(\Omega, \mathscr{F}, P)$ equipped with a history $\mathscr{H} = (\mathscr{H}_t)$; martingales are always presumed right-continuous with left-hand limits. (In later sections this is true by construction.)

The first result is not a martingale result per se, but we apply it to the analysis of martingales.

THEOREM 2.1 [Lenglart (1977)]. *Let X, Y be adapted, right-continuous, nonnegative processes such that Y is predictable and nondecreasing with $Y_0 = 0$. Suppose that Y dominates X in the sense that for every finite stopping time T,*

$$E[X_T] \le E[Y_T]. \tag{2.2}$$

Then for each $\varepsilon, \eta > 0$, and every finite stopping time T,

$$P\Big\{\sup_{t \le T} X_t \ge \varepsilon\Big\} \le \eta/\varepsilon + P\{Y_T \ge \eta\}. \tag{2.3}$$

We will apply Theorem 2.1 to square-integrable martingales: A martingale M is *square integrable* if $\sup_t E[M_t^2] < \infty$. The *predictable variation* process $\langle M \rangle$, the unique increasing, predictable process such that $M^2 - \langle M \rangle$ is a mean zero martingale, dominates M^2 in the sense of (2.2). In particular, if N is a point process with stochastic intensity λ, then the square of the innovation martingale $M_t = N_t - \int_0^t \lambda$ is dominated by $\langle M \rangle$, which is given by (1.2): $\langle M \rangle_t = \int_0^t \lambda$.

Returning to a general square-integrable martingale M, let $[M]$ be the quadratic variation process

$$[M]_t = \langle M^c \rangle_t + \sum_{s \le t} \Delta M_s^2,$$

where M^c is the continuous martingale part of M and $\Delta M_t = M_t - M_{t-}$ [see Jacod (1979)]. The predictable variation $\langle M \rangle$ is the predictable compensator (dual predictable projection) of the quadratic variation; in particular, the two differ by a mean zero martingale. The innovation martingale M engendered by a point process N is a "compensated sum of jumps," so that

$$[M]_t = N_t \tag{2.4}$$

for each t.

The following inequality will be used as well.

THEOREM 2.5 [Burkholder (1973)]. *For each $p \in (1, \infty)$ there exist constants c_p, C_p such that for every martingale M,*

$$c_p E\left[[M]_1^{p/2}\right] \le E\left[\sup_{t \le 1} |M_t|^p\right] \le C_p E\left[[M]_1^{p/2}\right]. \tag{2.6}$$

In particular, for $p = 2$,

$$E\left[\sup_t M_t^2\right] \le C_2 E\left[[M]_1\right] = C_2 E\left[\langle M \rangle_1\right]. \tag{2.6$'$}$$

Finally, we require a central limit theorem for sequences of martingales; the version below, albeit not the most general extant, suffices for this paper. Let (M^n) be a sequence of square-integrable mean zero martingales, with respect to histories $\mathscr{H}^n$, possibly but not necessarily defined over the same probability space.

THEOREM 2.7 [Rebolledo (1977), (1980)]. *Let V be a continuous, nondecreasing function on $[0,1]$ such that $V_0 = 0$, and suppose that*

(a) *for each t, $\langle M^n \rangle_t \to V_t$ in probability;*
(b) *there exist constants $b_n \downarrow 0$ such that*

$$P\left\{\sup_{t \le 1} |\Delta M_t^n| \le b_n\right\} \to 1.$$

Then there is a continuous, mean zero Gaussian martingale M with $\langle M \rangle_t = V_t$, such that $M^n \to_d M$ on $D[0,1]$, where $\to_d$ denotes convergence in distribution.

The martingale M is a centered Gaussian process with independent increments and $E[M_t M_s] = V_{t \wedge s}$.

3. Consistent maximum likelihood estimation via sieves. The setting remains as above: With respect to P_α the $(N^{(i)}, \lambda^{(i)})$ are i.i.d. copies of a pair (N, λ) such that the point process N has stochastic intensity $\alpha_t \lambda_t$. The index set I consists of all (left-continuous, right-hand limited) nonnegative functions $\alpha \in L^1[0,1]$. The log-likelihood functions $L_n(\alpha)$ are given by (1.10), where $N^n = \sum_{i=1}^n N^{(i)}$ and $\lambda^n = \sum_{i=1}^n \lambda^{(i)}$. We utilize an indirect approach to maximum likelihood estimation modeled on Grenander (1981, Section 8.1); the basic ideas are:

(1) For each n devise a subset I_n of I within which there does exist a maximum likelihood estimator $\hat{\alpha} = \hat{\alpha}_n$.
(2) Choose the I_n to increase as n does, and such that $\cup_n I_n$ is a dense subset of I.
(3) Given the appropriate additional assumptions, establish consistency of the sequence $(\hat{\alpha}_n)$ in the sense that $\|\hat{\alpha}_n - \alpha\|_1 \to 0$.

The family (I_n) is termed a *sieve*.

Note the resemblance to kernel methods of density estimation [cf. Ramlau-Hansen (1983)] and, in particular, the two simultaneous limit processes, as the sample size increases and the sieve mesh decreases. Not only consistency but also computational tractability is an issue in regard to choice of the sieve; we examine this issue following Theorem 3.13.

We present two consistency theorems, a strong consistency theorem giving conditions under which $\|\hat{\alpha} - \alpha\|_1 \to 0$ almost surely and a theorem providing less stringent conditions for c_n-consistency in the sense of convergence in probability. It is important to keep in mind that we estimate α itself rather than indefinite integrals; note also that the L^1-convergence is suited exactly to the role of α.

First, however, we must introduce a sieve. For each $a > 0$, let $I(a)$ denote the family of absolutely continuous functions $\alpha \in I$ satisfying

$$a \le \alpha \le a^{-1} \tag{3.1}$$

and

$$|\alpha'|/\alpha \le a^{-1}, \tag{3.2}$$

everywhere on $[0,1]$. As the *sieve mesh* a decreases to 0, $I(a)$ increases to a dense subset of I. We show that for each n and a, there exists a maximum likelihood estimator $\hat{\alpha}(n, a)$ relative to $I(a)$, then we choose $a = a_n$ converging to zero sufficiently slowly that $\hat{\alpha}(n, a_n) \to \alpha$ in $L^1[0,1]$.

Here is the main result.

THEOREM 3.3. *Assume that*:

(i) *The function* $m_s(\alpha) = E_\alpha[\lambda_s]$ *is bounded and bounded away from zero on* $[0,1]$.

(ii) *The "entropy"*

$$H(\alpha) = -\int_0^1 [1 - \alpha_s + \alpha_s \log(\alpha_s)]\, m_s(\alpha)\, ds$$

is finite.

(iii) *With* $\mathrm{Var}_\alpha(\lambda_s)$ *denoting the variance of* λ_s *under* P_α,

$$\int_0^1 \mathrm{Var}_\alpha(\lambda_s)\, ds < \infty. \tag{3.4}$$

(iv) $E_\alpha[N_1^2] < \infty$.

Then:

(a) *For each* n *and each* $a > 0$, *there exists a* (*not necessarily unique*) *maximum likelihood estimator* $\hat{\alpha}(n, a) \in I(a)$ *satisfying*

$$L_n(\hat{\alpha}(n, a)) \geq L_n(\alpha), \tag{3.5}$$

for all $\alpha \in I(a)$.

(b) *For* $a_n = n^{-1/4+\eta}$, *with* $0 < \eta < 1/4$, *and* $\hat{\alpha} = \hat{\alpha}(n, a_n)$, $\|\hat{\alpha} - \alpha\|_1 \to 0$, *almost surely with respect to* P_α.

PROOF. We follow the pattern of argument in Grenander (1981, Section 8.2, Theorem 1). For notational simplicity we omit variables of integration; if no measure is given an integral is with respect to Lebesgue measure; finally, all integrals are over $[0,1]$, but this also is suppressed.

To begin, it is shown in Grenander (1981) that

(i) $I(a)$ is uniformly bounded and equicontinuous;
(ii) $\{\alpha' \colon \alpha \in I(a)\}$ is uniformly bounded.

With n (and ω) fixed and $(\alpha(j))$ a sequence in $I(a)$ such that $L_n(\alpha(j)) \to \sup\{L_n(\alpha)\colon \alpha \in I(a)\}$, it follows that there is a subsequence, which we continue to denote by $(\alpha(j))$, such that the derivatives $\alpha'(j)$ converge in $L^1[0,1]$ and such that the real sequence $(\alpha_0(j))$ converges. Hence there is $\alpha(\infty) \in I(a)$ such that $\alpha(j) \to \alpha(\infty)$ uniformly. Since $\alpha(\infty)$ is continuous and bounded away from zero, (1.10) implies that

$$L_n(\alpha(\infty)) = \lim_j L_n(\alpha(j)) = \sup\{L_n(\alpha)\colon \alpha \in I(a)\},$$

proving (a).

To establish (b), we note first that by reasoning in Grenander (1981), given $\alpha \in I$ and $\delta > 0$, for sufficiently small $a > 0$, there exists $\tilde{\alpha} \in I(a)$—depending on δ and a, which we suppress—such that $\|\alpha - \tilde{\alpha}\|_1 < \delta$ and

$$\left| \int m(\alpha)[1 - \tilde{\alpha} + \alpha(\log \tilde{\alpha})] - \int m(\alpha)[1 - \alpha + \alpha(\log \alpha)] \right| < \delta. \tag{3.6}$$

Recalling that (N, λ) is a generic copy, let

$$L(\beta) = \int \lambda(1 - \beta) + \int (\log \beta)\, dN$$

be the log-likelihood function. By Jensen's inequality, the expectation of a log-likelihood function is maximized at the "true" index value; consequently,

$$H(\alpha) = -E_\alpha[L(\alpha)] \leq -E_\alpha[L(\beta)] = -\int m(\alpha)[1 - \beta + \alpha(\log \beta)]. \tag{3.7}$$

Suppose now that $\alpha \in I$ and $\delta > 0$ are fixed and let $\hat{\alpha} = \hat{\alpha}(n, a)$ be as in (a), with n and a variable. Then for n sufficiently large and a sufficiently small,

$$\begin{aligned}
&\left| -\int m(\alpha)[1 - \hat{\alpha} + \alpha(\log \hat{\alpha})] - H(\alpha) \right| \\
&\quad = -\int m(\alpha)[1 - \hat{\alpha} + \alpha(\log \hat{\alpha})] - H(\alpha) \quad \text{(by (3.7))} \\
&\quad = -\int m(\alpha)[1 - \hat{\alpha} + \alpha(\log \hat{\alpha})] + (1/n)\left[\int \lambda^n(1 - \hat{\alpha}) + \int (\log \hat{\alpha})\, dN^n\right] \\
&\qquad -(1/n)\left[\int \lambda^n(1 - \hat{\alpha}) + \int (\log \hat{\alpha})\, dN^n\right] \\
&\qquad + (1/n)\left[\int \lambda^n(1 - \tilde{\alpha}) + \int (\log \tilde{\alpha})\, dN^n\right] \\
&\qquad - (1/n)\left[\int \lambda^n(1 - \tilde{\alpha}) + \int (\log \tilde{\alpha})\, dN^n\right] \\
&\qquad + \int m(\alpha)[1 - \tilde{\alpha} + \alpha(\log \tilde{\alpha})] \\
&\qquad - \int m(\alpha)[1 - \tilde{\alpha} + \alpha(\log \tilde{\alpha})] - H(\alpha) \\
&\quad \leq \left| \int m(\alpha)[1 - \hat{\alpha} + \alpha(\log \hat{\alpha})] - (1/n)\left[\int \lambda^n(1 - \hat{\alpha}) + \int (\log \hat{\alpha})\, dN^n\right] \right| \\
&\qquad + \left| \int m(\alpha)[1 - \tilde{\alpha} + \alpha(\log \tilde{\alpha})] \right. \\
&\qquad \left. - (1/n)\left[\int \lambda^n(1 - \tilde{\alpha}) + \int (\log \tilde{\alpha})\, dN^n\right] \right| + \delta,
\end{aligned}$$

where we have used (3.5) and (3.6).

Assuming for the moment that by proper choice of large n and small a, the first two terms converge to zero almost surely, we will have demonstrated that

$$-\int m(\alpha)[1 - \hat{\alpha} + \alpha(\log \hat{\alpha})] \to H(\alpha),$$

almost surely, which is the same as

$$\int m(\alpha)[\hat{\alpha} - \alpha(\log \hat{\alpha})] \to \int m(\alpha)[\alpha - \alpha(\log \alpha)]. \tag{3.8}$$

Following Grenander (1981) one final time, we infer from (3.8) that for $h(y) = y - \log(1 + y)$,

$$\int h(\hat{\alpha}/\alpha - 1) m(\alpha) \alpha \to 0,$$

almost surely, which shows that $\hat{\alpha} \to \alpha$ in $L^1(m_s(\alpha)\, ds)$ and, in view of (i), suffices to confirm (b).

Of the two remaining terms we analyze only the first, which is the more difficult; the second is handled analogously. Using integration by parts, (3.1) and (3.2), we see that

$$\left|\int m(\alpha)[1-\hat{\alpha}+\alpha(\log\hat{\alpha})] - (1/n)\left[\int \lambda^n(1-\hat{\alpha}) + \int (\log\hat{\alpha})\,dN^n\right]\right|$$

$$\begin{aligned}
&\le \left|\int [1-\hat{\alpha}+\alpha(\log\hat{\alpha})](\lambda^n/n - m(\alpha))\right| \\
&\quad + \left|(\log\hat{\alpha}_1)(1/n)\left(N_1^n - \int_0^1 \alpha\lambda^n\right)\right| \qquad (3.9)\\
&\quad + \left|(1/n)\int(\hat{\alpha}_t'/\hat{\alpha}_t)\left(N_t^n - \int_0^t \alpha\lambda^n\right)dt\right| \\
&\le a^{-1}\int|\lambda^n/n - m(\alpha)| + 2(na)^{-1}\sup_t\left|N_t^n - \int_0^t \alpha\lambda^n\right|.
\end{aligned}$$

We consider the terms in order.

For fixed s the random variables $\lambda_s^{(i)} - m_s(\alpha)$ are i.i.d. and by (3.4) have mean zero and finite second moment; consequently, by the Marcinkiewicz–Zygmund strong law of large numbers [Chow and Teicher (1978), page 122], for $a = a_n = n^{-1/4+\eta}$,

$$a^{-1}|\lambda_s^n/n - m_s(\alpha)| \to 0, \tag{3.10}$$

almost surely. From Fubini's theorem we infer that almost surely (3.10) holds almost everywhere on $[0,1]$ with respect to Lebesgue measure. This last statement, (3.4) and the dominated convergence theorem give that almost surely

$$a^{-1}\int|\lambda^n/n - m(\alpha)| \to 0.$$

By (2.6) applied with $p = 4$ to the innovation martingale $M_t^n = N_t^n - \int_0^t\alpha\lambda^n$, for each $\varepsilon > 0$,

$$\begin{aligned}
P_\alpha\left\{(na)^{-1}\sup|M_t^n| > \varepsilon\right\} &\le (na\varepsilon)^{-4}E_\alpha\left[\sup|M_t^n|^4\right]\\
&= O\left((na\varepsilon)^{-4}E_\alpha\left[[M^n]_1^2\right]\right),
\end{aligned}$$

while by (2.4) and assumption (iv),

$$E_\alpha\left[[M^n]_1^2\right] = E_\alpha\left[(N_1^n)^2\right] = n\operatorname{Var}_\alpha(N_1) + n^2E_\alpha[N_1]^2.$$

Thus,

$$P_\alpha\left\{(na)^{-1}\sup|M_t^n| > \varepsilon\right\} = O(n^{-2}a^{-4}\varepsilon^{-4}),$$

and for the prescribed choice $a_n = n^{-1/4+\eta}$ it follows by the usual Borel–Cantelli argument that

$$(na)^{-1}\sup\left|N_t^n - \int_0^t\alpha\lambda^n\right| \to 0,$$

almost surely with respect to P_α, which completes the proof. □

The relatively slow rate of convergence in Theorem 3.3 is a disadvantage in applications since, roughly speaking, the sieve mesh a determines the accuracy of $\hat{\alpha}$ as an approximation to α. At the expense of weak rather than strong consistency, the rate of convergence can be improved significantly; moreover, one hypothesis from Theorem 3.3 can be eliminated and rates of convergence can be deduced.

THEOREM 3.11. *Suppose that assumptions* (i)–(iii) *of Theorem* 3.3 *are fulfilled and let* (c_n) *be a sequence with* $c_n \to \infty$ *and* $c_n/n^{1/2} \to 0$. *Then for* $a_n = (c_n/n^{1/2})^{1/2}$ *and* $\hat{\alpha} = \hat{\alpha}(n, a_n)$ *as in Theorem* 3.3(a) *we have*

$$\lim_{d\to\infty} \limsup_{n\to\infty} \sup_{\|\alpha'-\alpha\|_1<\varepsilon} P_{\alpha'}\{c_n\|\hat{\alpha} - \alpha'\|_1 > d\} = 0, \tag{3.12}$$

for every $\alpha \in I$ *and* $\varepsilon > 0$.

That is, the estimators $\hat{\alpha}$ are c_n-consistent; see for example Millar (1983).

PROOF. By following the proof of Theorem 3.3 through (3.9), one deduces that it suffices to show an analogue of (3.12) for each term there. By (3.4), for each d,

$$\begin{aligned} P_{\alpha'}\left\{c_n a^{-1}\int_0^1 |\lambda^n/n - m(\alpha)| > d\right\} &\le (c_n/ad)^2 E_{\alpha'}\left[\left(\int_0^1 |\lambda^n/n - m(\alpha)|\right)^2\right] \\ &\le (c_n/ad)^2 E_{\alpha'}\left[\int_0^1 (\lambda^n/n - m(\alpha))^2\right] \\ &= O\left(c_n^2/na^2d^2\right) \\ &= O\left(c_n/n^{1/2}d^2\right); \end{aligned}$$

uniformity over $\{\alpha': \|\alpha' - \alpha\| < \varepsilon\}$ is straightforward.

By Theorem 2.1 applied to the processes $X = (M^n)^2$ and $Y = \langle M^n\rangle$ and with $\eta = n^{3/2}/c_n$,

$$\begin{aligned} &P_{\alpha'}\left\{c_n(na)^{-1}\sup\left|N_t^n - \int_0^t \alpha'\lambda^n\right| > d\right\} \\ &\quad = P_{\alpha'}\left\{\sup(M_t^n)^2 > (nad)^2/c_n^2\right\} \\ &\quad \le c_n^2\left(n^{3/2}/c_n\right)/(nad)^2 + P_{\alpha'}\left\{\langle M^n\rangle_1 > n^{3/2}/c_n\right\} \\ &\quad \le d^{-2} + E_{\alpha'}\left[\int_0^1 \alpha'\lambda^n\right]/\left(n^{3/2}/c_n\right) \\ &\quad = d^{-2} + \left(c_n/n^{1/2}\right)\int_0^1 \alpha' m(\alpha'), \end{aligned}$$

for the given choice of a_n, and hence (3.12) holds. □

Note that one can come arbitrarily close to $n^{1/2}$-consistency, but that because (a_n) depends on (c_n) to actually attain $n^{1/2}$-consistency would require that

$a_n \to 0$ arbitrarily slowly. There is one (important) case where this can be accomplished: When α is known to belong to the sieve set $I(a)$ for some $a > 0$, then $I(a)$ is taken as the parameter space, and the proof of Theorem 3.11 may be repeated—with a held fixed—to obtain the following result.

THEOREM 3.13. *If* $\alpha \in I(a)$ *and if hypotheses* (i)–(iii) *of Theorem* 3.3 *are satisfied, then the maximum likelihood estimators* $\hat{\alpha} = \hat{\alpha}(n, a)$ *are* $n^{1/2}$-*consistent: For each* $\varepsilon > 0$,

$$\lim_{d\to\infty} \limsup_{n\to\infty} \sup_{\|\alpha'-\alpha\|_1<\varepsilon,\ \alpha'\in I(a)} P_{\alpha'}\left\{n^{1/2}\|\hat{\alpha} - \alpha'\|_1 > d\right\} = 0. \tag{3.14}$$

In Section 4 we use this result to prove a central limit theorem for $\hat{\alpha}$.

Computation of the maximum likelihood estimators $\hat{\alpha}$ is somewhat difficult. One must solve the function space optimization problem

$$\text{maximize} \quad -\int_0^1 \alpha_s \lambda_s^n\, ds + \sum\left(\log \alpha_{T_i}\right), \tag{3.15}$$

where the T_i are the points of N^n, subject to the constraints (3.1) and (3.2). Similar problems arise in the calculation of nonparametric estimators of density functions using penalized likelihood functions [see Geman and Hwang (1982), concerning the relationship to sieve methods], but the presence of λ_s^n in (3.15) and the form of the constraints seem to preclude a closed form solution. Numerical solution is feasible, of course.

Alternative sieves include the *histogram sieve* defined via the sieve sets

$$I(m) = \{\alpha\colon \alpha \text{ is constant on } ((k-1)/m, k/m],\ k = 1,\dots, m\},$$

which is more convenient computationally. Indeed, in this case the maximum likelihood estimator in $I(m)$ is

$$\hat{\alpha}(n, m) = N^n(J_k)\Big/\int_{J_k} \lambda_s^n\, ds, \tag{3.16}$$

on $J_k = ((k-1)/m, k/m]$. If one integrates the estimators (3.16) in order to estimate indefinite integrals of α, there result the estimators

$$\begin{aligned}\left(\int_0^t \alpha_s 1(\lambda_s^n > 0)\, ds\right)^{\wedge} &= \int_0^t \hat{\alpha}_s 1(\lambda_s^n > 0)\, ds \\ &\cong (1/m) \sum_{k=1}^{[mt]} \left[N^n(J_k)\Big/\int_{J_k} \lambda_s^n\, ds\right] \\ &\cong \int_0^t 1(\lambda_s^n > 0)(\lambda_s^n)^{-1}\, dN_s^n,\end{aligned}$$

where $[x]$ denotes the integer part of x. Thus, we obtain a maximum likelihood interpretation of the martingale estimator $\hat{B}$ of (1.5); see also Section 4.

4. Asymptotic normality. Likelihood ratio tests are often analyzed in terms of local asymptotic normality of likelihood ratios [introduced by Le Cam

(1960)]. In this section we first present such a theorem for the multiplicative intensity model. It not only is of independent interest but also leads to a central limit theorem for the estimators $\hat{\alpha}$.

We use martingale methods to examine the behavior of log-likelihood functions as random functions of t and, in particular, describe limit properties of differences $L_n(\alpha + n^{-1/2}\alpha^*) - L_n(\alpha)$ as $n \to \infty$. The log-likelihood process, indexed by $t \in [0, 1]$, is

$$L_n(\alpha, t) = \int_0^t \lambda_s^n(1 - \alpha_s)\, ds + \int_0^t (\log \alpha_s) N_s^n.$$

By Theorem 2.7, under the probability P_α,

$$\left(n^{-1/2}\left[N_t^n - \int_0^t \alpha\lambda^n\right]\right)_{0 \le t \le 1} \to_d (M_t(\alpha))_{0 \le t \le 1}, \tag{4.1}$$

where $M(\alpha)$ is a Gaussian martingale with variance function

$$V_t(\alpha) = \lim n^{-1}\langle M^n \rangle_t = \int_0^t m_s(\alpha)\alpha_s\, ds.$$

THEOREM 4.2. *Let α and α^* be fixed elements of I, and suppose, in addition to the structure stipulated above, that*

$$\int_0^1 \left[(\alpha_s^*)^2/\alpha_s\right] m_s(\alpha)\, ds < \infty \tag{4.3a}$$

and

$$\int_0^1 \left[(\alpha_s^*)^3/(\alpha_s)^2\right] m_s(\alpha)\, ds < \infty. \tag{4.3b}$$

Then under P_α the stochastic processes

$$\left(L_n(\alpha + n^{-1/2}\alpha^*, t) - L_n(\alpha, t) + \tfrac{1}{2}\int_0^t \left[(\alpha_s^*)^2/\alpha_s\right] m_s(\alpha)\, ds\right)_{0 \le t \le 1} \tag{4.4}$$

converge in distribution to a Gaussian martingale with variance function

$$V_t(\alpha, \alpha^*) = \int_0^t \left[(\alpha_s^*)^2/\alpha_s\right] m_s(\alpha)\, ds.$$

PROOF. Tightness of the sequence (4.4) is an immediate consequence of (4.1) and (4.5) below, so we need only identify the limit, in view of whose independent increments it suffices to consider one-dimensional distributions. For each n and t,

$$\begin{aligned} L_n(\alpha &+ n^{-1/2}\alpha^*, t) - L_n(\alpha, t) \\ &= -n^{-1/2}\int_0^t \lambda_s^n \alpha_s^*\, ds + \int_0^t \log\left(1 + n^{-1/2}\alpha_s^*/\alpha_s\right) dN_s^n \\ &= -n^{-1/2}\int_0^t \lambda_s^n \alpha_s^*\, ds \\ &\quad + \int_0^t \left[n^{-1/2}(\alpha^*/\alpha) - (2n)^{-1}(\alpha^*/\alpha)^2 + O\left(n^{-3/2}(\alpha^*/\alpha)^3\right)\right] dN_s^n \\ &= n^{-1/2}\int_0^t (\alpha_s^*/\alpha_s)(N_s^n - \alpha_s\lambda_s^n\, ds) - (1/2n)\int_0^t (\alpha_s^*/\alpha_s)^2\, dN_s^n \\ &\quad + O\left(n^{-1/2}\int_0^1 (\alpha^*/\alpha)^3 (dN^n/n)\right), \end{aligned} \tag{4.5}$$

uniformly in t. The error converges in probability to zero by the law of large numbers (cf. the proof of Theorem 3.3) and Slutsky's theorem. That the error bound is uniform in t confirms the tightness asserted at the beginning of the proof. Again by the law of large numbers, the second term converges in probability to $-(1/2)\int_0^t(\alpha_s^*/\alpha_s)^2 m_s(\alpha)\alpha_s\,ds$, while by (4.1) and the continuous mapping theorem the first term converges in distribution to $\int_0^t(\alpha_s^*/\alpha_s)\,dM_s(\alpha)$. Denoted by $M(\alpha, \alpha^*)$, the limit is the integral of a predictable process with respect to a Gaussian martingale and is hence a Gaussian martingale [by (1.2)] with variance function given above. □

Albeit of interest in its own right, Theorem 4.2 leads as well to a central limit theorem for the maximum likelihood estimators $\hat{\alpha}$, provided that the assumptions of Theorem 3.13 be fulfilled. Interestingly, this theorem, which applies to indefinite integrals of the $\hat{\alpha}$ rather than to the $\hat{\alpha}$ themselves, is the same central limit theorem satisfied by the martingale estimators (1.5).

THEOREM 4.6. *Suppose that* $\alpha \in I(a)$ *for some fixed* $a > 0$ *and that the assumptions of Theorems* 3.13 *and* 4.2 *are fulfilled, and let* $\hat{\alpha}(n, a)$ *be as in Theorem* 3.13. *Then the stochastic processes* $Z_n(t) = n^{1/2}\int_0^t(\hat{\alpha}_s - \alpha_s)\,ds$ *converge in distribution to a Gaussian martingale with variance function*

$$W_t(\alpha) = \int_0^t [\alpha_s/m_s(\alpha)]\,ds. \tag{4.7}$$

PROOF. With n fixed, define a transformation T of $L^1[0,1]$ into itself by

$$T(\alpha) = \int_0^{\cdot} \lambda^n(1-\alpha) + \int_0^{\cdot}(\log\alpha)\,dN^n.$$

The first, second and third Fréchet derivatives of T are given, following lengthy but straightforward computations, by

$$T'(\alpha)[\beta] = -\int_0^{\cdot}\beta\lambda^n\,ds + \int_0^{\cdot}(\beta/\alpha)\,dN^n,$$

$$T''(\alpha)[\beta][\gamma] = -\int_0^{\cdot}(\beta\gamma/\alpha^2)\,dN^n,$$

and

$$T'''(\alpha)[\beta][\gamma][\delta] = 2\int_0^{\cdot}(\beta\gamma\delta/\alpha^3)\,dN^n,$$

respectively. See Luenberger (1969) for details.

Then, on the one hand, with $n^{-1/2}$ denoting a constant function, by Luenberger (1969, Section 7.3)

$$\begin{aligned} L_n&(\alpha + n^{-1/2}) - L_n(\alpha) \\ &= T'(\alpha)[n^{-1/2}] + \tfrac{1}{2}T''(\alpha)[n^{-1/2}][n^{-1/2}] \\ &\quad + O(T'''(\alpha)[n^{-1/2}][n^{-1/2}][n^{-1/2}]) \\ &= T'(\alpha)[n^{-1/2}] + \tfrac{1}{2}T''(\alpha)[n^{-1/2}][n^{-1/2}] + O\left(n^{-1/2}\int_0^1\alpha^{-3}[dN^n/n]\right) \\ &= T'(\alpha)[n^{-1/2}] + \tfrac{1}{2}T''(\alpha)[n^{-1/2}][n^{-1/2}] + O_P(1), \end{aligned} \tag{4.8}$$

via the law of large numbers [this is essentially the same argument used to derive (4.5)]. The sense of the convergence in (4.8) is with respect to the norm on $L^1[0,1]$. Moreover,

$$\begin{aligned}\tfrac{1}{2}T''(\alpha)[n^{-1/2}][n^{-1/2}] &= -\tfrac{1}{2}\int_0^{\cdot}\alpha^{-2}(dN^n/n)\\ &\to -\tfrac{1}{2}\int_0^{\cdot}[m(\alpha)/\alpha]\,ds,\end{aligned} \tag{4.9}$$

in the sense of convergence in probability. Theorem 4.2, together with (4.8)–(4.9), implies that $T'(\alpha)[n^{-1/2}] \to_d M$, where M is a Gaussian martingale with variance function

$$V_t(\alpha, 1) = \int_0^t [m_s(\alpha)/\alpha_s]\,ds.$$

On the other hand, by the mean value theorem for functionals [Graves (1929)], for some $r \in (0,1)$,

$$\begin{aligned}T'(\alpha)[n^{-1/2}] &= T''(\hat\alpha + r(\alpha - \hat\alpha))[\alpha - \hat\alpha][n^{-1/2}]\\ &= -\int_0^{\cdot}\left[(\alpha - \hat\alpha)n^{-1/2}/(\hat\alpha + r(\alpha - \hat\alpha))^2\right] dN^n\\ &= n^{1/2}\int_0^{\cdot}[(\hat\alpha - \alpha)/\alpha^2][dN^n/n]\\ &\quad + O_P\left(n^{-1/2}\int_0^1\left[(\hat\alpha - \alpha)^2/\alpha^3\right] dN^n\right)\\ &= n^{1/2}\int_0^{\cdot}[(\hat\alpha - \alpha)/\alpha^2][dN^n/n] + O_P(1)\\ &\qquad \text{(by Theorem 3.13 and the law of large numbers)}\\ &\cong n^{1/2}\int_0^{\cdot}(\hat\alpha_s - \alpha_s)[m_s(\alpha)/\alpha_s]\,ds,\end{aligned}$$

from which we infer that

$$n^{1/2}\int_0^{\cdot}(\hat\alpha - \alpha)[m(\alpha)/\alpha]\,ds \to_d M.$$

From this,

$$n^{1/2}\int_0^{\cdot}(\hat\alpha - \alpha)\,ds \to_d \int_0^{\cdot}[\alpha/m(\alpha)]^2\,dM,$$

with the limit a Gaussian martingale with variance function $W_t(\alpha)$. □

That (4.5) is the same central limit theorem fulfilled by the martingale estimators (1.5) can be confirmed by Gill (1980), Jacobsen (1982) or Karr (1986, Theorem 5.16).

5. Examples. In this section we illustrate the results in Sections 3–4 using two key special cases of the multiplicative intensity model, Poisson processes and survival analysis.

Poisson processes. Recall that a Poisson process N on $[0,1]$ with intensity function α is defined by the properties:

(1) For disjoint sets $A_1, \dots, A_k$ the random variables $N(A_1), \dots, N(A_k)$ are independent (N has independent increments).
(2) For each A, $N(A)$ has a Poisson distribution with mean $\int_A \alpha_s \, ds$.

More to the point of this paper, among point processes admitting stochastic intensities, Poisson processes are characterized by their having deterministic stochastic intensities: The baseline intensity λ is identically equal to one, so that $\lambda_t(\alpha) = \alpha_t$ for each t.

The method presented in Section 3 can be used to derive consistent maximum likelihood estimators relative to the sieves there, but to illustrate the breadth of the method of sieves we describe briefly the use of histogram sieves. Moreover, we need no longer be confined to $[0,1]$, and work instead in a compact metric space E.

Given a finite measure μ on E, a point process N is a Poisson process with mean measure μ if N has independent increments and if for each A, $N(A)$ has a Poisson distribution with mean $\mu(A)$. Now let μ_0 be a fixed, finite measure on E; we consider the statistical model $\mathscr{P} = \{P_\alpha \colon \alpha \in L^1(\mu_0)\}$, where under P_α, the data are i.i.d. Poisson processes $N^{(i)}$, each with mean measure $\alpha(x)\mu_0(dx)$.

For each $m \in \mathbb{N}$ let $E = \sum_{j=1}^{l_m} A_{mj}$ be a partition of E into sets A_{mj} with $\mu_0(A_{mj}) > 0$, and assume that $\max_j \mu_0(A_{mj}) \to 0$. For each m let $I(m)$ be the set of functions α that are constant over each partition set A_{mj}. Then by (3.16) the maximum likelihood estimator $\hat{\alpha}(n, m)$ relative to sample size n and the set $I(m)$ is

$$\hat{\alpha}(n, m) = \sum_{j=1}^{l_m} \left[N^n(A_{mj}) / n\mu_0(A_{mj}) \right] 1_{A_{mj}}, \tag{5.1}$$

where $N^n = \sum_{i=1}^{n} N^{(i)}$. For suitable choice of (m_n) these estimators are strongly consistent.

THEOREM 5.2. *If $m = m_n$ is chosen in such a manner that, with $\tilde{l}_n = l_{m_n}$,*

$$\sum_{n=1}^{\infty} \left(\tilde{l}_n^4 / n^2 \right) < \infty, \tag{5.3}$$

then for $\hat{\alpha} = \hat{\alpha}(n, m_n)$, $\|\hat{\alpha} - \alpha\|_1 \to 0$, P_α-almost surely.

PROOF. Let n and m both be variable; then by (5.1)

$$\begin{aligned}\|\hat{\alpha}(n,m)-\alpha\|_1 &= \sum_{j=1}^{l_m}\int_{A_{mj}}\left|N^n(A_{mj})/n\mu_0(A_{mj})-\alpha(x)\right|\mu_0(dx)\\ &\le \sum_{j=1}^{l_m}\int_{A_{mj}}\left|N^n(A_{mj})/n\mu_0(A_{mj})-\mu_0(\alpha;A_{mj})/\mu_0(A_{mj})\right|d\mu_0\\ &\quad+\sum_{j=1}^{l_m}\int_{A_{mj}}\left|\mu_0(\alpha;A_{mj})/\mu_0(A_{mj})-\alpha(x)\right|\mu_0(dx)\\ &\le \sum_{j=1}^{l_m}\left|N^n(A_{mj})/n-E_\alpha[N(A_{mj})]\right|\\ &\quad+\sum_{j=1}^{l_m}\int_{A_{mj}}\left|\mu_0(\alpha;A_{mj})/\mu_0(A_{mj})-\alpha(x)\right|\mu_0(dx),\end{aligned}$$

where $\mu_0(\alpha; A_{mj}) = \int_{A_{mj}}\alpha\, d\mu_0$. That the second term converges to zero provided only that $m \to \infty$ can be shown by a variety of techniques; see for example Grenander (1981, pages 419–420).

In regard to the first, for $\varepsilon > 0$,

$$\begin{aligned}&P_\alpha\left\{\sum_{j=1}^{l_m}\left|N^n(A_{mj})/n-E_\alpha[N(A_{mj})]\right|>\varepsilon\right\}\\ &\qquad\le \varepsilon^{-4}E_\alpha\left[\left\{\sum_{j=1}^{l_m}\left|N^n(A_{mj})/n-E_\alpha[N(A_{mj})]\right|\right\}^4\right].\end{aligned}$$

With n and m fixed the random variables $N^n(A_{mj})/n - E_\alpha[N(A_{mj})]$ are independent in j and by expansion of the fourth power of the summation followed by repeated application of the Cauchy–Schwarz inequality,

$$P_\alpha\left\{\sum_{j=1}^{l_m}\left|N^n(A_{mj})/n-E_\alpha[N(A_{mj})]\right|>\varepsilon\right\}=O(\varepsilon^{-4}n^{-2}l_m^4);$$

the dominant term comes from the $\binom{l_m}{4} \sim l_m^4$ summands with all indices distinct. Thus (5.3) suffices to give

$$\sum_n P_\alpha\left\{\sum_{j=1}^{\tilde{l}_n}\left|N^n(A_{mj})/n-E_\alpha[N(A_{mj})]\right|>\varepsilon\right\}<\infty,$$

for every $\varepsilon > 0$, which completes the proof. □

Survival analysis. Let $X_1, X_2, \ldots$ be i.i.d. random variables whose unknown distribution function F admits hazard function $h = F'/(1 - F)$, which is to be estimated from the X_i. Substitution methods based on empirical distribution functions are not effective, since the latter do not admit hazard functions. To estimate h from the counting process viewpoint, one observes that for each n the point process $N_t^n = \sum_{i=1}^n 1(X_i \leq t)$ has P_h-stochastic intensity

$$\lambda_t^n(h) = h(t) \sum_{i=1}^n 1(X_i \geq t);$$

thus we have a multiplicative intensity model with baseline stochastic intensity

$$\lambda_t^n = \sum_{i=1}^n 1(X_i \geq t) = n - N_{t-}^n.$$

The methods described above apply, as do results in Sections 3 and 4. The central limit theorem implied by Theorem 4.5 is the same as that described, e.g., in Gill (1980), Jacobsen (1982) and Karr (1986) for the martingale estimators $\hat{B}_t = \int_0^t (n - N_{s-}^n)^{-1} \, dN_s^n$; the Gaussian limit martingale has variance function $F(t)/[1 - F(t)]$.

Acknowledgments. It is a pleasure to thank the referees and Associate Editor for numerous helpful suggestions, which have improved the paper materially.

REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

BORGAN, O. (1984). Maximum likelihood estimation in a parametric counting process model, with application to censored failure time data and multiplicative models. *Scand. J. Statist.* **11** 1–16.

BRÉMAUD, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, New York.

BURKHOLDER, D. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.

GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

GILL, R. D. (1980). *Censoring and Stochastic Integrals*. Math. Centre Tracts 124. Mathematisch Centrum, Amsterdam.

GRAVES, L. M. (1929). Riemann integration and Taylor's theorem in general analysis. *Trans. Amer. Math. Soc.* **27** 163–177.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statist.* **12**. Springer, New York.

JACOBSEN, M. (1984). Maximum likelihood estimation in the multiplicative intensity model—a survey. *Internat. Statist. Rev.* **52** 193–207.

JACOD, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31** 235–253.

JACOD, J. (1979). *Calcul Stochastique et Problèmes de Martingales. Lecture Notes in Math.* **714**. Springer, New York.

JOHANSEN, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51** 165–174.

KARR, A. F. (1986). *Point Processes and their Statistical Inference*. Dekker, New York.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 27–98.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **13** 171–179.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes* **2**. Springer, New York.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1980). A functional central limit theorem for semimartingales. *Theory Probab. Appl.* **25** 667–688.

LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods.* Wiley, New York.

MCKEAGUE, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14** 579–589.

MILLAR, P. W. (1983). The minimax principle in asymptotic statistical theory. *Lecture Notes in Math.* **976** 75–265. Springer, New York.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

REBOLLEDO, R. (1977). Sur les applications de la théorie des martingales a l'étude statistique d'une famille de processus ponctuels. *Lecture Notes in Math.* **636** 27–70. Springer, New York.

REBOLLEDO, R. (1980). Central limit theorem for local martingales. *Z. Wahrsch. verw. Gebiete* **51** 269–286.

Q. W. C. WHITING SCHOOL OF ENGINEERING
THE JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218