PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 8 (1987), p. 103–110

ESTIMATION OF THE PERIODIC FUNCTION IN THE MULTIPLICATIVE INTENSITY MODEL

BY

JACEK LEŚKOW (WROCŁAW)

Abstract. Given a point process $\{N(t), t \geqslant 0\}$ with the stochastic intensity $\lambda(t)$ of the form $\lambda(t) = \alpha_0(t) Y(t)$, it is shown that using the sieves technique one can construct a strongly consistent maximum likelihood estimator of the functional factor $\alpha(t)$. The latter is assumed to be periodic with the known period $T = 1$, and the "censoring process" $Y(t)$ fulfills some mild regularity assumptions. As an easy consequence it follows that the maximum likelihood estimator (MLE) can similarly be computed if $\{N^{(i)}(t), t \in [0, 1], i = 1, 2, \ldots\}$ are not independent and identically distributed but satisfy some mixing conditions.

This paper extends the results of Karr [13].

1. Introduction. We consider the following problem: on a probability space $(\Omega, \mathscr{F}, P)$ we observe a point process $\{N(t), t \geqslant 0\}$ adapted to a standard filtration $\{\mathscr{F}_t, t \geqslant 0\}$. The stochastic intensity $\lambda(t)$ of the process $N(t)$ (for the definition, see [1]) is assumed to be of the form

$$\lambda(t) = \alpha_0(t) Y(t), \qquad t \in \boldsymbol{R}^+, \tag{1.1}$$

where $\alpha_0(t)$ is a deterministic, unknown periodic function integrable on $[0, 1]$ with the known period equal to 1. The stochastic process $Y(t)$, sometimes called "the censoring process", satisfies the usual assumptions of $\{\mathscr{F}_t, t \geqslant 0\}$ predictability and nonnegativity.

The aim of this paper is to show that maximum likelihood estimator of the function α_0 is strongly consistent if α_0 is periodic, and a single realization of the process $N(t)$ on $\boldsymbol{R}^+$ is observed.

Multiplicative intensity model (1.1) was introduced by Aalen [1]. Karr [13] applied the theory to estimate $\alpha_0(t)$ when $\{N^{(i)}(t), t \in [0, 1], i = 1, 2, \ldots\}$ is a family of i.i.d processes. Our approach, based on a single realization of the process $N(t)$ and assumptions of periodicity of $\alpha_0(t)$ (somewhat artificial at the first glance), is quite natural when considering such phenomena as, e.g., arrivals of customers to a store, a work-load of a service station or arrivals of patients at an Intensive Care Unit.

Recently Rolski [18] has demonstrated how to compute characteristics in queues at which arrivals are according to Poisson process with a periodic intensity function $\alpha(t)$. Vere-Jones [19] and Mardia [16] considered a cyclic Poisson process as well. The former author proposed a maximum likelihood estimate for such processes. Lewis [14] presented data collected at an Intensive Care Unit of a hospital. The "time-of-day" effect, described above, had allowed to assume that the intensity was periodic with the known period equal to 24 hours. The same author had also given another example concerning thunderstorm severity in Great Britain that showed a tendency to have a seasonal effect.

In above-mentioned papers $\lambda(t)$ (see (1.1)) was assumed to be nonrandom, i.e. $N(t)$ was a nonhomogeneous Poisson process. The estimation problem was a parametric one. Our model is more general because $\lambda(t)$ may well be random. Moreover, the setup, being nonparametric, makes applications more feasible.

In Section 2 we formulate our central results together with remarks. Section 3 carries the task of proving the results.

2. Consistent maximum likelihood estimation. We consider a countable family of the points processes $\{N^{(i)}(t),\ t\in[0, 1],\ i = 1, 2, \ldots\}$ obtained by splitting a single realization of $N(t)$ on R^+. Formally, let

$$N^{(i)}(t) = N(t+i-1)-N(i-1), \quad t\in[0, 1],\ i = 1, 2, \ldots \tag{2.1}$$

Using the recipe

$$Y^{(i)}(t) = Y(t+i-1), \tag{2.2}$$

we can divide the process $Y(t)$ into a countable number of pieces $\{Y^{(i)}(t),\ t\in[0, 1],\ i = 1, 2, \ldots\}$. Since the function $\alpha_0(t)$ is periodic with the known period equal to 1, the stochastic intensity of the process $N^{(i)}(t)$ can be written in the following form:

$$\lambda^{(i)}(t) = \alpha_0(t)\, Y^{(i)}(t). \tag{2.3}$$

Observe that we can consider processes $\{N^{(i)}(t),\ t\in[0, 1],\ i = 1, 2, \ldots\}$ that do not correspond to a single realization $\{N(t),\ t \geqslant 0\}$ but fulfill certain mixing conditions. Since the methods and calculations are identical in that case as in the former one, we restrict ourselves to the case where observations are engendered by the point process $\{N(t),\ t \geqslant 0\}$.

A natural model for observations engendered by the point process $\{N(t),\ t \geqslant 0\}$ is a family of probability measures $\mathscr{P}^0 = \{P^0_{\tilde{\alpha}},\ \tilde{\alpha}\in\tilde{I}\}$, where $\tilde{I}$ is a set of periodic, nonnegative functions that are left-continuous with right-hand limits and period equal to 1. It is well known [13] that if $\tilde{\alpha}$ is integrable on every bounded interval of R^+, then the family is dominated, i.e. there exists a measure P^0 such that $P^0_{\tilde{\alpha}} \ll P^0$ for every $\tilde{\alpha} \in \tilde{I}$ and P^0 corresponds to $\alpha(t) \equiv 1$.

On the other hand, we can make use of the construction of processes $N^{(i)}(t)$ and $Y^{(i)}(t)$ and periodicity of α to consider an alternative model $\mathscr{P} = \{P_\alpha, \alpha \in I\}$, where I consists of nonnegative left-continuous functions with right-hand limits defined on $[0, 1]$. We also assume that the underlying functions are integrable on $[0, 1]$. Thus, the set I can be equipped with the usual norm

$$\|\alpha\| = \int_0^1 |\alpha(s)|\, ds. \tag{2.4}$$

Let $N^n = \sum_{i=1}^{n} N^{(i)}$ and $Y^n = \sum_{i=1}^{n} Y^{(i)}$, where $N^{(i)}$ and $Y^{(i)}$ were defined above. From Theorem 19.7 of [15] and the considerations above we have the following

LEMMA 1. *Let $P^0_{\tilde{\alpha}} \in \mathscr{P}^0$ and let $P^0_{\tilde{\alpha}}$ correspond to the process $\{N(t), t \geqslant 0\}$, where $\tilde{\alpha}(s) = \alpha(s)$, $s \in [0, 1]$ and $\alpha \in I$. Then*

$$\frac{dP^0_{\tilde{\alpha}}}{dP^0}(n) = \exp\Big[\int_0^1 Y^n(s)\big(1-\alpha(s)\big)\, ds + \int_0^1 \log \alpha(s)\, dN^n(s)\Big].$$

Proof. Observe that

$$\begin{aligned}\frac{dP^0_{\tilde{\alpha}}}{dP^0}(n) &= \exp\Big[\int_0^n Y(s)\big(1-\tilde{\alpha}(s)\big)\, ds + \int_0^n \log \tilde{\alpha}(s)\, dN(s)\Big] \\ &= \exp\Big[\sum_{i=1}^{n} \Big\{\int_{i-1}^{i} Y(s)\big(1-\tilde{\alpha}(s)\big)\, ds + \int_{i-1}^{i} \log \tilde{\alpha}(s)\, dN(s)\Big\}\Big] \\ &= \exp\Big[\int_0^1 Y^n(s)\big(1-\alpha(s)\big)\, ds + \int_0^1 \log \alpha(s)\, dN^n(s)\Big].\end{aligned}$$

The likelihood function

$$L_n(\alpha) = \int_0^1 Y^n(s)\big(1-\alpha(s)\big)\, ds + \int_0^1 \log \alpha(s)\, dN^n(s) \tag{2.5}$$

can therefore be parametrized by functions from the functional space I.

For the sake of technical convenience we define "formal entropy"

$$H_n(\alpha) = -E_{\alpha_0} L_n(\alpha) = -\int_0^1 m^n\big(\alpha_0(s)\big)\big(1-\alpha(s)+\alpha_0(s)\log\alpha(s)\big)\, ds, \tag{2.6}$$

where

$$m^n\big(\alpha_0(s)\big) = \sum_{i=1}^{n} E_{\alpha_0} Y^{(i)}(s), \tag{2.7}$$

and the function $\alpha_0(s)$ is the true unknown parameter. Note that α_0 minimizes $H_n(\alpha)$ for $\alpha \in I$.

It is known that the unconstrained optimization of $L_n(\alpha)$ does not lead to worthy results (see e.g. [8] for numerous examples). Grenander [11] proposed a method of sieves, based on maximization of (2.5) over a subspace S_n of I, as a remedy in this situation.

A family $\{S_n, n = 1, 2, \ldots\}$ of subsets of I is called a *sieve* if S_n is compact for every n, S_n is increasing in n and $\bigcup S_n$ is dense in I. Note that different sieves give rise to different ML estimators (see [8] and [9] for numerous examples).

Let $\{a_n\}$ be an arbitrary sequence of positive numbers such that $a_n \to 0$. For the technical convenience we denote by $S(a_n)$ the subset of I (instead of S_n).

Following Karr [13] and Grenander [11] we define $S(a_n)$ to be the set of absolutely continuous functions satisfying

$$a_n \leqslant \alpha(s) \leqslant 1/a_n, \tag{2.8}$$

$$|\alpha'(s)| \leqslant \alpha(s)/a_n, \quad s \in [0, 1]. \tag{2.9}$$

It can be shown [13] that $S(a_n)$ is compact in L^1, $\bigcup S(a_n)$ is dense for $a_n \to 0$, and that in $S(a_n)$ there exists a function $\hat{\alpha}(n, a_n)$ such that

$$L_n(\hat{\alpha}(n, a_n)) \geqslant L_n(\alpha) \quad \text{for } \alpha \in S(a_n). \tag{2.10}$$

A function $\hat{\alpha}(n, a_n)$, fulfilling (2.10), will be called the *maximum likelihood estimator* (MLE).

Furthermore, let $M^{(i)}(s)$ be a martingale corresponding to the point process $N^{(i)}(s)$, and $[M^{(i)}](s)$ be the quadratic variation process of $M^{(i)}(s)$ (for the definition see [4]). In the sequel we suppress the argument $s \in [0, 1]$ and the double index (n, a_n). We have the following

THEOREM 1. *Assume that* (A.1)-(A.4) *hold true, where*:

(A.1) $C_1 \leqslant n^{-1} m^n(\alpha_0) \leqslant C_2$ *for* m^n *defined in* (2.7);

(A.2) $n^{-1} \int_0^1 \operatorname{Var} Y^n(s)\, ds \leqslant C_3$;

(A.3) $E\big(M^{(i)}(1) M^{(j)}(1)\big) = 0$ *and* $n^{-1} E_{\alpha_0}[\sum_{i=1}^{n} M^{(i)}]^2(1) \leqslant C_4$ *for* $i \neq j$;

(A.4) *the process* $Y(s)$ *is strongly mixing with the mixing function* $\varphi(s) = O(s^{-1/2} \log^{-2} s)$.

$C_1, \ldots, C_4$ *are constants not depending on* $s \in [0, 1]$ *and* n.

Then, for $a_n = n^{-1/4+\delta}$, $\delta > 0$, *and MLE satisfying* (2.10) *we have, for* $n \to \infty$, $\|\alpha_n - \alpha_0\| \to 0$ P_{α_0}*-almost everywhere.*

As we have mentioned, the result of Theorem 1 remains valid when the observations $\{N^{(i)}(t), t \in [0, 1], i = 1, 2, \ldots\}$ do not correspond to a single realization of the point process $N(t)$ on $\boldsymbol{R}^+$. However, the very formulation should be adjusted to that case. Namely, (A.4) is replaced by

(A.4′) *The processes* $Y^{(i)}(s)$ *are strongly mixing with the mixing function* $\varphi(i) = O(i^{-1/2}(\log i)^{-2})$ *for* $s \in [0, 1]$, $i = 1, 2, \ldots$

We have then the following

THEOREM 2. *Let* (A.1)–(A.3) *and* (A.4′) *be fulfilled. Then MLE* $\hat{\alpha}_n$ *is strongly consistent, i.e., for* $n \to \infty$, $\|\hat{\alpha}_n - \alpha_0\| \to 0$ P_{α_0}*-a.e., where* $\hat{\alpha}_n$ *is obtained by the method of sieves* (2.10).

Under (A.1)–(A.3) a weak consistency result for $\hat{\alpha}_n$ in a single realization model can be obtained. Since the proof of that result follows closely the patterns of proofs of our Theorem 1 and that of Karr [13], we present the assertion only.

THEOREM 3. *Assume that* (A.1)–(A.3) *are satisfied and the rate of growth of the sieve* a_n *is* $a_n = n^{-1/2}(\log\log n)^{1/2}$. *Then, for MLE* $\hat{\alpha}_n$, *we have* $\|\hat{\alpha}_n - \alpha_0\| \to 0$ *in* P_{α_0} *probability for* $n \to \infty$.

Remarks. (i) It will be seen from the proof in Section 3 that the rate of convergence is related to the growth of c_i^2, where

$$c_i^2 = E_{\alpha_0}|N^{(i)}(1) - \int_0^1 \alpha_0(s)\, Y^{(i)}(s)\, ds|^2.$$

If $c_n^2 = O(n^{\beta-2})$, where $3 > \beta > 2$, then we can take $a_n = n^{-\gamma}$ and $\gamma = (3-\beta-\varepsilon)/4$, $\varepsilon > 0$.

(ii) Assumption (A.4) seems to be very restrictive. However, it is well known (see e.g. [12] and [6]) that if $Y(s)$ is a Markov process with stationary transition probabilities satisfying Doeblin condition, then $Y(s)$ is strongly mixing with $\varphi(s)$ decreasing geometrically. This enables us to enlarge the class of possible applications of the method of sieves to the processes whose stochastic intensity fulfills mixing condition (A.4). Such considerations were not possible within the frames of the theory developed in [13]. Thus, our model seems to be adequate when observing the periodic phenomena $(\alpha(t))$ with random censorship $Y(t)$.

(iii) In Theorem 2 the sieve $S(a_n)$ is allowed to grow faster than that of Theorem 1.

(iv) Since, for fixed n, the likelihood function L_n is continuous in L_+^1 on $S(a_n)$ and is concave, the solution of (2.10) is unique (for details see e.g. [7], Proposition 1.1 and Proposition 1.2, p. 35).

3. Proof of Theorem 1. Let us forst note that if $\{X_n, \mathscr{F}_n, n \geq 1\}$ is a strong mixing sequence with a mixing function $\varphi(n)$, then $\{X_n, \mathscr{F}_n\}$ is a mixingale with constants $c_n = 2(EX_n^2)^{1/2}$ and $\psi_m = \varphi^{1/2}(m)$. Thus, in view of (A.4), the sequence $\{Y^n - m^n\}$ is mixingale with $c_n = 2(\operatorname{Var} Y^n)^{1/2}$ and ψ_m as

(A.4). For the definition of the strong mixing sequence and that of a mixingale we refer the reader to [3] and [12], respectively.

Note that since the sieve is dense in the space I, given the "true" parameter α_0 and any $\delta > 0$, one can find a_n sufficiently small such that $\|\alpha^* - \alpha_0\| < \delta$, where $\alpha^* \in S(a_n)$. Therefore, due to (2.6), (A.1) and the inequality $\log x \leqslant x - 1$ for $x > 0$, we have

$$\frac{1}{n}|H_n(\alpha^*) - H_n(\alpha_0)| < \mathrm{const}\cdot\delta. \tag{3.1}$$

Moreover, it can easily be seen that, for fixed n,

$$\begin{aligned}
&\frac{1}{n}(H_n(\hat{\alpha}) - H_n(\alpha_0)) \\
&\quad\leqslant \frac{1}{n}\big[(H_n(\hat{\alpha}) + L_n(\hat{\alpha})) - L_n(\hat{\alpha}) + (H_n(\alpha^*) - H_n(\alpha^*)) - H_n(\alpha_0)\big] \\
&\quad\leqslant \frac{1}{n}\big[(H_n(\hat{\alpha}) + L_n(\hat{\alpha})) - L_n(\alpha^*) - H_n(\alpha^*) + H_n(\alpha^*) - H_n(\alpha_0)\big] \\
&\quad\leqslant \frac{1}{n}\big[|H_n(\hat{\alpha}) + L_n(\hat{\alpha})| + |H_n(\alpha^*) + L_n(\alpha^*)| + |H_n(\alpha^*) - H_n(\alpha_0)|\big].
\end{aligned} \tag{3.2}$$

In the second inequality in (3.2) we applied the inequality $L_n(\hat{\alpha}) \geqslant L_n(\alpha^*)$ which is true for any $\alpha^* \in S(a_n)$. It is known (see [11]) that if $\lim_{n\to\infty} n^{-1}|H_n(\hat{\alpha}_n) - H_n(\alpha_0)| = 0$, i.e.

$$\lim_{n\to\infty} \int_0^1 h\left(\frac{\alpha_n(s)}{\alpha_0(s)} - 1\right)\alpha_0(s)\,ds = 0, \qquad P_{\alpha_0}\ \text{a.e.},$$

then $\hat{\alpha}_n \to \alpha_0$ P_{α_0} a.e. in I for $h(y) = y - \log(1+y)$.

Therefore, to show that the right-hand side of (3.2) tends to zero observe that the third term can be made arbitrarily small when the sieve grows. In order to estimate the other two terms we follow Karr [13]. We have

$$\begin{aligned}
\left|\frac{H_n(\hat{\alpha})}{n} + \frac{L_n(\hat{\alpha})}{n}\right| &\leqslant \frac{1}{a_n}\left|\int_0^1 \frac{1}{n}(Y^n(s) - m^n)\,ds\right| + \\
&+ \frac{1}{na_n}\Big|N^n(1) - \int_0^1 \alpha(s)\,Y^n(s)\,ds\Big| + \frac{1}{na_n}\sup_{t\leqslant 1}\Big|N_t^n - \int_0^t \alpha(s)\,Y^n(s)\,ds\Big|.
\end{aligned} \tag{3.3}$$

We prove now that the three terms of (3.3) converge to zero almost everywhere. Due to (A.4) and the strong law of large numbers for mixingales [12], the first term in (3.3) tends to zero for fixed $s \in [0, 1]$, namely

$$\frac{1}{na_n}\left|Y^n(s)-m^n\big(\alpha_0(s)\big)\right| \to 0\ P_{\alpha_0}\text{-a.e.}$$

if $a_n = n^{-1/4+\delta}$, $\delta > 0$.

Applying now Fubini's theorem (see e.g. formula (35.5), Chapter V, in [17]) we have P_{α_0}-a.e.

$$\frac{1}{na_n}|Y^n - m^n(\alpha_0)| \to 0 \text{ a.e. on } [0, 1]$$

with respect to the Lebesgue measure. Taking into account (A.2) and the dominated convergence theorem, we have

$$\frac{1}{na_n}\int_0^1 \left|Y^n(s)-m^n\big(\alpha_0(s)\big)\right| ds \to 0\ \ P_{\alpha_0}\text{-a.e.}$$

for the prescribed choice of a_n.

Similar considerations can be applied to the second term in (3.3), namely, for $a_n = n^{-1/4+\delta}$,

$$\frac{1}{na_n}|N^n(1) - \int_0^1 \alpha(s)\, Y^n(s)| \to 0\ \ P_{\alpha_0}\text{-a.e.},$$

since $Y^{(i)}$, and hence $N^n(1) - \int_0^1 \alpha(s)\, Y^n(s)$, is also a mixingale with the desired asymptotic properties.

By a theorem of Burkholder [5], (A.3) and an argument analogous to that of Karr [13] we obtain the following bounds for the third term:

$$P_{\alpha_0}\left\{\frac{1}{na_n}\sup_{0\leqslant t\leqslant 1}|N^n(t) - \int_0^t \alpha(s)\, Y^n(s)| > \varepsilon\right\} = O(n^{-2}a_n^4\varepsilon^{-2}).$$

Hence, applying Borel-Cantelli lemma for $a_n = n^{-1/4+\delta}$, we get the convergence to zero a.e. This completes the proof, as the second term in (3.2) can be handled analogously.

Acknowledgement. I am deeply indebted to Dr. Tomasz Rolski for pointing out the problem and for numerous helpful comments and suggestions. I also gratefully acknowledge suggestions of two referees which have led to corrections and substantial improvement of the paper.

REFERENCES

[1] O. O. Aalen, *Nonparametric inference for a family of counting processes*, Ann. Stat. 6 (1978), p. 711–726.

[2] P. K. Andersen and Ø. Borgan, *Counting processes models for life history data* (review), Scand. J. Stat. 12 (1985), p. 97–158.

[3] P. Billingsley, *Convergence of probability measures*, Wiley, New York 1968.
[4] P. Brémaud, *Point processes and queues: Martingale dynamics*, Springer-Verlag, New York 1981.
[5] D. Burkholder, *Distribution function inequalities for martingales*, Ann. Prob. 1 (1973), p. 19–42.
[6] J. Doob, *Stochastic processes*, Wiley, New York 1953.
[7] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland, New York 1976.
[8] S. Geman, *Sieves for nonparametric estimation of densities and regressions*, Rept. in Pattern Analysis 99 (1981), Brown University.
[9] S. Geman and C.-R. Hwang, *Nonparametric maximum likelihood estimation by the method of sieves*, Ann. Stat. 10 (1982), p. 410-414.
[10] R. D. Gill, *Censoring and stochastic integral*, Math. Centre tracts 124, Mathematisch Centrum, Amsterdam 1980.
[11] U. Grenander, *Abstract inference*, Wiley, New York 1981.
[12] P. Hall and J. J. Heyde, *The martingale limit theory and its application*, Academic Press, 1980.
[13] A. F. Karr, *Maximum likelihood estimation in the multiplicative intensity model, via sieves*, Ann. Stat. (1987), to appear.
[14] P. A. W. Lewis, *Statistical analysis of point processes*, in *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, 1972.
[15] R. S. Liptser, and A. N. Shiryayev, *Statistics of random processes II*, Springer-Verlag, New York 1978.
[16] K. V. Mardia, *Statistics of directional data*, J. Royal Stat. Soc. B, 37 (1975), No. 3.
[17] K. R. Parthasarathy, *Introduction to probability and measure*, New Delhi 1980.
[18] T. Rolski, *Approximation of periodic queues* (preprint), Mathematical Institute of the University of Wrocław, Wrocław 1986.
[19] D. Vere-Jones, *On the estimation of frequency of point process data*, in *Essays in Statistical Sciences*, Paper in honour of P. A. P. Moran, J. Appl. Prob., Special volume 19 A (1982).

Polish Academy of Sciences
Institute of Mathematics
ul. Kopernika 18
51-617 Wrocław, Poland

Received on 25. 9. 1985;
revised version on 25. 3. 1986