PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 14, Fasc. 1 (1993), pp. 11–24

SIEVE-BASED MAXIMUM LIKELIHOOD ESTIMATOR FOR ALMOST PERIODIC STOCHASTIC PROCESS MODELS*

BY

JACEK LEŚKOW (WROCŁAW)

Abstract. Assume that the point process $\{N(t);\ t \geqslant 0\}$ is observed with stochastic intensity of the form $\lambda(t) = \lambda_0(t) \cdot Y(t)$, where λ_0 is an unknown almost periodic nonnegative function and $Y(t)$ is an observable nonnegative stochastic process. It is shown that the sieve-based maximum likelihood estimator of λ_0 is consistent in the appropriate metric of the space of uniformly almost periodic (UAP) functions. The same technique establishes the consistency of the sieve-based maximum likelihood estimator of a UAP drift function in a stochastic differential equation.

1. Introduction. Let (Ω, F, P) be a probability space on which we observe a point process $\{N(t);\ t \geqslant 0\}$ with history $\{F_t;\ t \geqslant 0\}$, where F_t are increasing sub-σ-fields of F. In the sequel it is assumed that the stochastic intensity $\lambda(t)$ of the process $N(t)$ is in the multiplicative form

$$\lambda(t) = \lambda_0(t) \cdot Y(t), \tag{1.1}$$

where $Y(t)$ is an observable stochastic process, satisfying the predictability conditions (see, e.g., [1], [12] or [13]). The function $\lambda_0(t)$ is unknown, deterministic, continuous, nonnegative and uniformly almost periodic (UAP) on $[0, \infty)$.

Alternatively, it will be assumed that a diffusion process $X(t)$ is observed, which is a strong solution of the stochastic differential equation

$$dX(t) = \lambda_0(t) \cdot a(t, X)dt + dW(t), \tag{1.2}$$

where, as previously, λ_0 is an unknown deterministic continuous UAP function on $[0, \infty)$, $a(t, X)$ is F_t-measurable for each $t \geqslant 0$ and $W(t)$ is a Brownian motion.

* Research supported by the Air Force Office of Scientific Research, Contract No. F49620 85C 0144.

We recall that a real continuous function λ on $[0, \infty)$ is *uniformly almost periodic* (UAP) if for any $\varepsilon > 0$ there exists $L > 0$ such that in any subinterval of $[0, \infty)$ with the length greater than L there exists a τ_ε belonging to this subinterval such that

$$\sup_{x \geqslant 0} |\lambda(x+\tau_\varepsilon)-\lambda(x)| < \varepsilon.$$

Sums and products of UAP functions are UAP functions. For more information on UAP functions see, e.g., [4] or [2]. Note also that the space B of all UAP functions on $\boldsymbol{R}^+$ with the norm

$$\|\lambda\| = \lim_{T\to\infty} (1/T) \int_0^T |\lambda(s)|\,ds, \qquad \lambda \in B, \tag{1.3}$$

is a metric space. The space B may be equipped with other norms, such as sup norm

$$\|\lambda\|_{\sup} = \sup_{s \geqslant 0} |\lambda(s)|$$

or L^2-norm

$$\|\lambda\|_2 = \lim_{T\to\infty} (1/T) [\int_0^T |\lambda(s)|^2\,ds]^{1/2}.$$

We recall that $\|\lambda\|_{\sup} \geqslant \|\lambda\|_2 \geqslant \|\lambda\|_1$ and that the space B with the norm $\|\lambda\|_2$ is a nonseparable Hilbert space.

The space B of UAP functions with the norm (1.3) contains continuous and periodic functions λ for which

$$\|\lambda\| = (1/\tau) \int_0^\tau |\lambda(s)|\,ds,$$

where τ is the period λ. This equality shows the obvious equivalence between the norm (1.3) and the $L^1[0, \tau]$-norm for periodic λ.

This paper deals with the consistent estimation of the unknown function λ_0 in models (1.1) and (1.2) using the norm (1.3). In both models it is assumed that a single realization of a stochastic point process $N(t)$ or the diffusion process $X(t)$ is observed over an increasing time interval.

The nonparametric estimation of λ_0 from point process data in the multiplicative intensity model was considered in the pioneering paper of Aalen [1]. A detailed study of the statistical theory point processes may be found, e.g., in [12] or [13]. However, the methods of Aalen [1] are applicable to the observations of multiple copies of a point process on a fixed time interval exclusively.

The problem of estimating a periodic function λ_0 from a single realization of a point process was considered, e.g., by Krickeberg [14], Pons and

Turckheim [24], Lewis [20] or Leśkow [15]–[17]. The assumption of periodicity of λ_0 is quite natural when applied to such phenomena as: earthquake occurrences [27], [28], arrivals at an intensive care unit [20], distributional patterns of plants [26] or the number of particles entering a Geiger counter [16]. Lewis [20] presented data that showed the "time-of-day" effect and allowed to assume that the underlying intensity was periodic with known period equal to 24 hours. The same author has also given another example concerning the thunderstorm severity in Great Britain which had a "seasonal effect". The above papers presented methods of constructing strongly consistent and asymptotically normal estimators of the unknown function λ_0 for periodic point process models.

However, in nonparametric approaches to the estimation of the periodic factor of the intensity of a point process it was assumed that the true period of the estimated function is known (see, e.g., [24], [14] or [15], [17], [18]). On the other hand, in the parametric case there are several methods of estimating the period (see, e.g., [27], [28]). This paper presents a nonparametric method of estimating a periodic function from a stochastic point or diffusion process without prior knowledge of the period.

Observe that the sum of two periodic functions with periods (say) 1 and $\sqrt{2}$, respectively, is not periodic but it is almost periodic (see [4]). On the other hand, in a formal setting, it is quite desirable that the space of functions of interest be equipped with a linear structure. In this context, the choice of the space of UAP functions appears quite naturally.

The statistical motivation for the selection of the space of UAP functions is the following. There is an interest in estimating the unknown periodic function λ_0 with an unknown period A_{λ_0}. In practice, however, it is possible to indicate the countable set A of the real numbers such that $A_{\lambda_0} \in A$. Now, in the space B of all UAP functions on the real line for each A_i from A and a given length of the observation interval, say n, we can find the compact set $O(i, m_n)$ of appropriately normalized trigonometric polynomials with the period A_i, where m_n is the number of terms in the polynomial. The sequence m_n tends to infinity for n tending to infinity. Therefore, the set

$$K = \overline{\bigcup_{n=1}^{\infty} K_n}, \quad \text{where } K_n = \bigcup_{i=1}^{n} O(i, m_n),$$

where the closure is in the sup norm, is a separable subset of B since $O(i, m_n)$ are separable. Moreover, $\lambda_0 \in K$, so we can proceed with our nonparametric inferential investigations on K instead of on B. This makes our estimation problem feasible since the set K, unlike the whole space B, is separable. The detailed construction of the sets $O(i, m_n)$ will be presented in Section 2.

In a general context the above statistical motivation enables us to assume that the unknown λ_0 belongs to the set K which is separable, i.e. $K = \overline{\bigcup_n K_n}$ and K_n are compact.

Despite the broad applicability of periodic point process models and the number of theoretical results in this field (see e.g. [20], [27] or [15]), an almost periodic analogue of the theory does not exist. It should be pointed out, however, that many physical and demographical phenomena could be successfully modelled with the help of periodic approach had the true period of the phenomena been known.

There is also a vast literature on the estimation of a drift function, say λ_0, in a stochastic differential equation like (1.2). The maximum likelihood method based on sieves was used by Geman and Hwang [9] to obtain a consistent estimator of the unknown function λ_0 in the model

$$dX(t) = \lambda_0(t) + dW(t), \tag{1.4}$$

where W is a Wiener process and the observations are generated by independent copies of the process $X(t)$. A more general version of the model (1.4) for the independent data was considered by Beder [3] and Nguyen and Pham [22]. The almost periodic models for stochastic differential equations of the type (1.2) have been considered by Dorogovtsev [7] under the assumption that the estimated function λ_0 belongs to a compact (hence finite-dimensional) subset of the space of UAP functions on the positive half-line. In Section 3 we present an infinite-dimensional analogue of the results of Dorogovtsev for the model (1.2).

Section 2 is devoted to the maximum likelihood method based on a sieve and its application to the model (1.1). Here, the unknown function λ_0 is assumed to belong to a countably compact subset of the space B and the consistency of the maximum likelihood estimator in the model (1.1) is demonstrated. Section 3 contains a result on the consistency of the maximum likelihood sieve-based estimator of the function λ_0 in the model (1.2).

Methods based on the assumption of almost periodicity and (1.2) are frequently used in signal processing context. For example, in the recent paper of Dandawate and Giannakis [6] the model (1.2) was used, where $a(t, X)$ described an information signal, $\lambda_0(t)$ — deterministic modulating function, and $W(t)$ denoted a noise. In the mentioned paper the assumption of almost periodicity of the modulating function λ_0 was used to model nonstationary signals. It is also known that the statistical methods based on periodicity or almost periodicity are widely applied in modelling AM and FM radio signals in the underwater environment (see, e.g., [8]).

2. Maximum likelihood estimation in point process models. In this section we assume that the observations come from a point process $\{N(t);\ t \geqslant 0\}$ with a stochastic intensity $\lambda(t)$ of the form (1.1). The unknown function λ_0 is assumed to belong to a separable subset K of the space B with the norm (1.3).

Let P_λ^T be the distribution of the point process $\{N(t);\ 0 \leqslant t \leqslant T\}$ indexed by $\lambda \in B$, where T is finite. It is well known (see, e.g., [21], Theorem 19.7, or [13], Theorem 5.2, p. 170) that the family of measures $\boldsymbol{P}^T = \{P_\lambda^T;\ \lambda \in B\}$ is dominated, i.e., there exists $P_1^T \in \boldsymbol{P}^T$ such that $P_\lambda^T \ll P_1^T$ for any $P_\lambda^T \in \boldsymbol{P}^T$. The measure P_1^T may be chosen to correspond to a Poisson process with intensity 1 on $[0, T]$.

The density of P_λ^T with respect to P_1^T may be represented in the form

$$\frac{dP_\lambda^T}{dP_1^T} = \exp\Big\{\int_0^T Y(s)(1-\lambda(s))\,ds + \int_0^T \log\lambda(s)\,dN(s)\Big\}. \tag{2.1}$$

The log-likelihood function will be defined as

$$\begin{aligned} L(\lambda, T) &= T^{-1}\ln\{dP_\lambda^T/dP_1^T\} \\ &= T^{-1}\int_0^T Y(s)(1-\lambda(s))\,ds + T^{-1}\int_0^T \log\lambda(s)\,dN(s). \end{aligned} \tag{2.2}$$

For the technical convenience the "entropy" is defined as

$$\begin{aligned} H(\lambda, T) &= -\mathrm{E}_{\lambda_0} L(\lambda, T) \\ &= -T^{-1}\int_0^T \mathrm{E}\{Y(s)\}\cdot\{1-\lambda(s)+\lambda_0(s)\log\lambda(s)\}\,ds. \end{aligned} \tag{2.3}$$

In the nonparametric setting the direct maximization of the likelihood function $L(\lambda, T)$ fails. A way to circumvent such difficulties is to introduce a *sieve*, i.e., a family of increasing compact subsets $\{K_n\}$ of the set K such that $\overline{\bigcup_n K_n} = K$ (see, e.g., [10], [9] or [15]). This idea is consistent with the practical need for separable subsets indicated in the Introduction. Let us now define the family of subsets $O(i, m_n)$ on which the sieve $\{K_n\}$ will be built.

Let us introduce

$$W(i, m_n, t) = \sum_{k=-m_n}^{m_n} \big(\alpha_{i,k}\sin(2\pi kt/A_i) + \beta_{i,k}\cos(2\pi kt/A_i)\big),$$

where $\alpha_{i,k}$ and $\beta_{i,k}$ are real coefficients, $A_i \in A$, and A is the countable set containing the period A_{λ_0} of the unknown λ_0.

Now, put

$$O(i, m_n) = \big\{\lambda \in B\colon\ \lambda(t) = \min\big(m_n, \max(m_n^{-1}, W(i, m_n, t))\big)\big\},$$

where m_n is integer and $m_n \to \infty$ for $n \to \infty$.

To see better the formula for the set $O(i, m_n)$ it helps to observe that if $\lambda \in O(i, m_n)$, then $\lambda(t) = W(i, m_n, t)$ for $m_n^{-1} \leqslant W(i, m_n, t) \leqslant m_n$, $\lambda(t) = m_n^{-1}$ for $W(i, m_n, t) < m_n^{-1}$ and $\lambda(t) = m_n$ for $W(i, m_n, t) > m_n$.

Observe that $O(i, m_n)$ is compact in the space B with topology generated by the norm (1.3) (see also [7]). We will put now

$$K_n = \bigcup_{i=1}^{n} O(i, m_n),$$

where $\lim_{n\to\infty} m_n = \infty$ and m_n is usually called the *size of the sieve*. Obviously, the set K_n is also compact, so we can call the family $\{K_n\}$ a *sieve*.

It is easy to see now that for a function $\lambda \in K_n$

$$m_n^{-1} \leqslant \lambda(s) \leqslant m_n \tag{2.4}$$

and

$$|\lambda'(s)| \leqslant m_n \lambda(s), \tag{2.5}$$

where m_n is the sequence tending to infinity (a size of the sieve), and λ' denotes the derivative of the function λ.

In the sequel, we will put $m_n = [\max(T^\alpha, 1)]$, where $[\cdot]$ denotes the integer part and $\alpha > 0$. It is also understood that $n = [\max(T, 1)]$.

The sets of functions having the property (2.4) and (2.5) are compact in B in the topology generated by the norm (1.3) (see, e.g., [7]). Moreover, the function $L(\lambda, T)$ is continuous on K_n, so the maximum likelihood estimator $\hat{\lambda}_T$ may be defined by

$$L(\hat{\lambda}_T, T) = \max_{\lambda \in K_n} L(\lambda, T), \qquad T > 0. \tag{2.6}$$

The following assumptions will be used in the sequel:

(A.1) The process $Y(s)$ is φ-mixing with the mixing rate $\varphi(s) = O(s^{-2})$ for s outside the neighbourhood of zero.

(A.2) For $p = 1$ or $p = 2$ the p-th moments of the process Y satisfy the following inequalities:

$$0 < \inf_{s>0} \mathrm{E}Y^p(s) \leqslant \sup_{s>0} \mathrm{E}Y^p(s) < \infty.$$

THEOREM 2.1. *Assume that conditions* (A.1) *and* (A.2) *are fulfilled and that the* MLE $\hat{\lambda}_T$ *is defined by* (2.6) *over the sieve* K_n *with the properties* (2.4) *and* (2.5). *Assume also that* $0 < \alpha < 1/4$, *i.e., the sequence* m_n *tends to infinity slower than* $n^{1/4}$. *Then the estimator* $\hat{\lambda}_T$ *is consistent, i.e.,*

$$\|\hat{\lambda}_T - \lambda_0\| \to 0 \textit{ in probability} \quad \textit{as } T \to \infty,$$

where $\|\cdot\|$ *is the norm defined in* (1.3).

For the proof we need the following two lemmas.

LEMMA 2.2. *Given* $\lambda_0 \in K$ *and* $\delta > 0$ *there exist* $N(\delta)$ *and* $\lambda(\delta)$ *such that* $\|\lambda(\delta) - \lambda_0\| < \delta$ *and* $\lambda(\delta) \in K_n$ *for* $n \geqslant N(\delta)$.

Proof. Since $\lambda_0 \in K = \bigcup_n K_n$ and the sets K_n are increasing, there exists N_1 such that $\lambda_0 \in K_n$ for $n \geqslant N_1$. Moreover, the sets K_n are compact in the topology generated by the norm (1.3), so there exists a δ-net $\{\lambda_1^n, \ldots, \lambda_p^n\}$ such that $\|\lambda_i^n - \lambda_0\| < \delta$ for $i = 1, \ldots, p$ and $n \geqslant N_1$. To obtain the assertion it suffices now to take $\lambda(\delta)$ from $\{\lambda_1^n, \ldots, \lambda_p^n\}$ and put $N(\delta) = N_1$. ∎

Observe also that for any $\varepsilon > 0$ it is possible to find $N(\varepsilon)$ such that for $\lambda(\delta) \in K_n$, $n \geqslant N(\varepsilon)$, we have $|H(\lambda(\delta), T) - H(\lambda_0, T)| < \varepsilon$ for sufficiently large T (see, e.g., [15] for similar computations).

The next lemma is based on the proof of Theorem 1, Chapter 8.2, of [10].

LEMMA 2.3. *If for fixed* $\omega \in \Omega$,

$$\lim_{T \to \infty} |H(\hat{\lambda}_T(\omega), T) - H(\lambda_0, T)| = 0,$$

then $\|\hat{\lambda}_T(\omega) - \lambda_0\| \to 0$ *as* $T \to \infty$.

Proof. Keeping in mind a fixed $\omega \in \Omega$ we drop it from the notation. Observe that

$$(2.7) \quad H(\hat{\lambda}_T, T) - H(\lambda_0, T) = T^{-1} \int_0^T \mathrm{E} Y(s)(\hat{\lambda}_T(s) - \lambda_0(s))\,ds - T^{-1} \int_0^T \mathrm{E} Y(s) \lambda_0(s) \log(\hat{\lambda}_T(s)/\lambda_0(s))\,ds.$$

A simple application of the ideas of the proof of the above-mentioned theorem of [10] (see also [15]) yields the following implication:

$$(2.8) \quad \text{If } T^{-1} \int_0^T \mathrm{E} Y(s) \lambda_0(s) h\big(-1 + (\hat{\lambda}_T(s)/\lambda_0(s))\big)\,ds < \varepsilon \quad \text{for } h(y) = y - \log(1+y),$$

then

$$T^{-1} \int_0^T |\hat{\lambda}_T(s) - \lambda_0(s)|\,ds < \eta(\varepsilon), \quad \text{where } \eta(\varepsilon) \to 0 \text{ for } \varepsilon \to 0.$$

Note that the assumption in (2.8) is exactly what we have obtained in (2.7).

To prove the assertion of the lemma suppose, conversely, that there exists $\varepsilon^* > 0$ such that $\|\hat{\lambda}_T - \lambda_0\| > \varepsilon^*$ for large T, i.e., there exists a subsequence $\{T_k\}$, $T_k \to \infty$, such that $\lim_{k\to\infty} \|\hat{\lambda}_{T_k} - \lambda_0\| > \varepsilon^*$. Thus we could find k_0 such that, for any $k > k_0$,

$$T_k^{-1} \int_0^{T_k} |\hat{\lambda}_{T_k}(s) - \lambda_0(s)|\,ds > \varepsilon^*$$

which, on the behalf of the implication (2.8), contradicts the assumption of the lemma. ∎

Proof of Theorem 2.1. To prove Theorem 2.1 it suffices now to show that

$$\lim_{T \to \infty} |H(\hat{\lambda}_T, T) - H(\lambda_0, T)| = 0 \text{ in probability}.$$

Using the same techniques as in [15] note that

$$(2.9)\quad \big(H(\hat{\lambda}_T, T)-H(\lambda_0, T)\big) \leqslant \big(H(\hat{\lambda}_T, T)+L(\hat{\lambda}_T, T)\big) \\ +\big(H(\lambda(\delta), T)+L(\lambda(\delta), T)\big)+\big(H(\lambda(\delta), T)-H(\lambda_0, T)\big),$$

where $\lambda(\delta)$ is chosen as in Lemma 2.2.

On account of Lemma 2.2 the third term of the right-hand side of (2.9) can be made arbitrarily small for large T and the desired convergence would follow if we show the asymptotical neglibility of the first two terms. Following the same technical considerations as in [15] note that

$$(2.10)\quad |H(\hat{\lambda}_T, T)+L(\hat{\lambda}_T, T)| \leqslant \Big|m_n T^{-1}\int_0^T \big(Y(s)-\mathrm{E}Y(s)\big)ds\Big| \\ +m_n T^{-1}\Big|N(T)-\int_0^T \lambda_0(s)\,Y(s)ds\Big|+m_n T^{-1}\sup_{0\leqslant t\leqslant T}\Big|N(t)-\int_0^t \lambda_0(s)\,Y(s)ds\Big|.$$

We analyze the three terms separately.

The Chebyshev inequality applied to the first term of the right-hand side of (2.10) yields that

$$P\Big\{\Big|m_n T^{-1}\int_0^T \big(Y(s)-\mathrm{E}Y(s)\big)ds\Big| > \varepsilon\Big\} < \varepsilon^{-2}\operatorname{Var}\Big(m_n T^{-1}\int_0^T \big(Y(s)-\mathrm{E}Y(s)\big)ds\Big).$$

Note that

$$(2.11)\quad \operatorname{Var}\Big(\int_0^T \big(Y(s)-\mathrm{E}Y(s)\big)ds\Big) = \mathrm{E}\iint_{K_T}\big(Y(s)-\mathrm{E}Y(s)\big)\cdot\big(Y(v)-\mathrm{E}Y(v)\big)dvds \\ +\mathrm{E}\iint_{K_T^c}\big(Y(s)-\mathrm{E}Y(s)\big)\cdot\big(Y(v)-\mathrm{E}Y(v)\big)dvds,$$

where $K_T = \{(s, v)\colon |s-v| < 1, 0 \leqslant s, v \leqslant T\}$ and K_T^c is the complement of K_T in the set $[0, T]\times[0, T]$.

Observe now that the first term of the right-hand side of (2.11), due to the assumption (A.2), is of the order $O(T)$. Similarly, the second term of the right-hand side of (2.11) is of the order $O(T\ln T)$. Hence we get the convergence in probability of the first term of (2.10).

The second term of the inequality (2.10) will be analyzed with the help of the SLLN for martingales (see [23]). In particular, it suffices to show that for some $a > 1$

$$(2.12)\qquad \mathrm{E}\Big(\int_a^\infty (m_n^2 T^{-2})\,d\langle M\rangle(T)\Big) < \infty,$$

where $\langle M\rangle(t)$ is the predictable variation process corresponding to the martingale M. Applying now the assumption (A.2) and the fact that $m_n = [T^\alpha]$ for $0 < \alpha < 1/4$ we easily get the inequality (2.12) for any $a > 1$.

To the third term we apply the inequality of Burkholder, i.e., for any square-integrable martingale M

$$\mathrm{E}\{\sup_{0\leqslant t\leqslant T}|M(t)|^4\}\leqslant C\mathrm{E}\{\langle M\rangle(T)\}^2, \tag{2.13}$$

where the constant C does not depend on the martingale M.

Observe now that due to (2.13) we have

$$P\{\sup_{0\leqslant t\leqslant T} M(t)>\varepsilon\}\leqslant(\varepsilon^{-2}T^{-4}m_n^4)\cdot C\mathrm{E}\{\langle M\rangle(T)\}^2. \tag{2.14}$$

The sequence on the right-hand side of (2.14) is summable when $0<\alpha<1/4$.

The application of the Borel–Cantelli lemma yields the P-a.e. convergence for the first term in (2.9). The second term of the expression (2.9) may be analyzed analogously. This completes the proof of Theorem 2.1 ■

To obtain strong consistency of the estimator $\hat{\lambda}_T$ we need the following assumption:

(A.3) The process $\{Y(t); t\geqslant 0\}$ is stationary and uniformly bounded, i.e., there exists a constant C such that $Y(s)\leqslant C$ for any $s\geqslant 0$ and $\omega\in\Omega$.

COROLLARY 2.4. *Assume that the conditions* (A.1), (A.2) *and* (A.3) *hold. Then the maximum likelihood estimator* $\hat{\lambda}_T$ *of* λ_0 *in the model* (1.1) *is strongly consistent in the norm* (1.3), *i.e.* $\|\hat{\lambda}_T-\lambda_0\|\to 0$ *almost surely as* $T\to\infty$.

Proof. The proof follows along the lines of the proof of Theorem 2.1. It suffices, therefore, to demonstrate that

$$m_n T^{-1}\int_0^T (Y(s)-\mathrm{E}Y(s))ds\to 0 \text{ a.s.} \tag{2.15}$$

We will show that

$$\mathrm{E}\Big[\int_0^T (Y(s)-\mathrm{E}Y(s))ds\Big]^4 = O(T^2), \tag{2.16}$$

which, due to the Borel–Cantelli lemma and the fact that $m_n=[T^\alpha]$, will suffice to complete the proof of Corollary 2.4. To see that (2.16) holds note that on the account of the stationarity of the process Y we have

$$\begin{aligned}(2.17)\quad & \mathrm{E}\Big[\int_0^T (Y(s)-\mathrm{E}Y(s))ds\Big]^4 \\ & \leqslant C_1\cdot T\iiint_K |\mathrm{E}\bar{Y}(0)\bar{Y}(v_1)\bar{Y}(v_1+v_2)\bar{Y}(v_1+v_2+v_3)|dv_1dv_2dv_3,\end{aligned}$$

where $K = \{(v_1, v_2, v_3)\colon\ v_1+v_2+v_3 \leqslant T,\ v_1, v_2, v_3 > 0\}$, $\bar{Y}(s) = Y(s)-\mathrm{E}Y(s)$ and C_1 is a finite constant. We are now ready to use Lemma 4, p. 172, of [5]. From this lemma it follows that

$$(2.18)\qquad \iiint_K |\mathrm{E}\bar{Y}(0)\bar{Y}(v_1)\bar{Y}(v_1+v_2)\bar{Y}(v_1+v_2+v_3)|\,dv_1\,dv_2\,dv_3$$
$$\leqslant C_2\Big\{\iiint_{K_1}\varphi(v_1)dv_1\,dv_2\,dv_3+\iiint_{K_2}(\varphi(v_1)\varphi(v_3)+\varphi(v_2))dv_1\,dv_2\,dv_3$$
$$+\iiint_{K_3}\varphi(v_3)dv_1\,dv_2\,dv_3\Big\},$$

where $K_i = \{(v_1, v_2, v_3)\colon\ v_j, v_k \leqslant v_i,\ j, k = 1, 2, 3,\ j \neq k,\ j \neq i,\ k \neq i\} \cap K$, $i = 1, 2, 3$, φ is the mixing function connected with the process Y (see (A.2)) and C_2 is a positive constant. Combining now the inequalities (2.17), (2.18) and the assumption (A.2) on the speed of convergence of the function φ we get the inequality (2.16) which proves the almost sure convergence of (2.15). ∎

Remark 2.5. The assumption (A.3) on the boundedness of the process Y may be replaced by a stronger mixing property assumption for Y. Suppose that the process $\{Y(s);\ s \geqslant 0\}$ is stationary and ψ-mixing with the rate $\psi(s) = O(s^{-2})$. Then the following inequality holds:

$$(2.19)\qquad \mathrm{E}|\bar{Y}(s)\bar{Y}(s+v)| \leqslant \psi(v)\mathrm{E}|\bar{Y}(s)|\mathrm{E}|\bar{Y}(v)|.$$

For the definition of ψ-mixing and the proof of the inequality (2.19) see, e.g., [25].

Straightforward calculations, based on the inequalities (2.18), (2.19) applied to the process Y show that the equality (2.16) still holds in the ψ-mixing case.

The mixing assumption is quite understandable in practice when modelling processes with the long-term independence property. When such a property holds, then it is usually assumed that the dependence vanishes after a finite number of observations. Therefore, the assumptions of φ-mixing or ψ-mixing and related polynomial rates of convergence for φ and ψ are not really very restrictive.

Remark 2.6. For finite n and, hence, finite m_n the estimator $\hat{\lambda}_T$ may be computed by using maximization algorithms for the function $L(\hat{\lambda}_T, T)$. Such algorithms provide methods of computing the coefficients of the trigonometric polynomial $W(i, m_n, t)$ defined at the beginning of this section. This issue will be studied in a subsequent research on properties of the estimator $\hat{\lambda}_T$ in the almost periodic stochastic process models.

3. Nonparametric maximum likelihood estimation in diffusion models. In this section it is assumed that the observations are generated by a diffusion process $\{X(t);\ t>0\}$ being a strong solution of the stochastic differential equation

$$dX(t) = \lambda_0(t)\cdot a(t, X)dt + dW(t),$$

where λ_0 is the unknown UAP function, a is the nonanticipative functional, and W is a Wiener process.

To find an MLE of the UAP function λ_0 the technique of Section 2 will be applied. As previously, let P_λ^T denote the distribution of the process $\{X(t);\ 0 \leqslant t \leqslant T\}$ for $\lambda \in K$, K being a countable compact subset of the space B. It is well known that in such a case the family $P^T = \{P_\lambda^T;\ \lambda \in B\}$ is dominated by a measure P_0^T corresponding to a Wiener process on $[0, T]$ and the likelihood function $L(\lambda, T)$ is of the form

$$L(\lambda, T) = T^{-1}\int_0^T \lambda(t)\cdot a(t, X)dX(t) - (2T)^{-1}\int_0^T \lambda^2(t)\cdot a^2(t, X)dt. \tag{3.1}$$

For the details see, e.g., [21], Theorem 7.7, or [19].

The MLE of the UAP function λ_0 will be defined following the method presented in Section 2. As previously, the sieve $\{K_n\}$ is defined by

$$K_n = \bigcup_{i=1}^{n} O(i, m_n),$$

where the sets $O(i, m_n)$ were defined in Section 2.

The maximum likelihood sieve-based estimator $\hat{\lambda}_T$ is defined as

$$L(\hat{\lambda}_T, T) = \max_{\lambda\in K_n} L(\lambda, T), \qquad T > 0, \tag{3.2}$$

where $L(\lambda, T)$ was defined in (3.1).

Before stating the consistency result the following assumptions are introduced:

(B.1) The process $a(t, X)$ is φ-mixing with mixing function $\varphi(t) = O(t^{-2})$ for t outside a neighbourhood of zero.

(B.2) There exists $\eta > 0$ such that $\inf_{t>0} \mathrm{E}a^2(t, X) > \eta$. Moreover, the fourth moments of a are bounded, i.e., $\sup_{t>0} \mathrm{E}a^4(t, X) < \infty$.

THEOREM 3.1. *Suppose that the conditions* (B.1) *and* (B.2) *are fulfilled. Moreover, let* m_n, *the size of the sieve, be of the order* $[T^\alpha]$, $T > 0$ *for* $0 < \alpha < 1/4$. *Then the maximum likelihood estimator* $\hat{\lambda}_T$ *defined in* (3.2) *and based on the sieve* $\{K_n\}$ *is weakly consistent, i.e.,* $\|\hat{\lambda}_T - \lambda_0\| \to 0$ *in probability as* $T \to \infty$, *where* $\|\cdot\|$ *is the norm defined in* (1.3).

Proof. The line of argument is virtually the same as in the proof of Theorem 2.1. Observe first that

$$H(\lambda, T) = T^{-1}\int_0^T \lambda(t)\lambda_0(t)\mathrm{E}a^2(t, X)dt + (2T)^{-1}\int_0^T \lambda^2(t)\mathrm{E}a^2(t, X)dt. \tag{3.3}$$

Therefore, for an arbitrary UAP function λ we have

$$|H(\lambda, T)-H(\lambda_0, T)| = (2T)^{-1}\int_0^T \mathrm{E}a^2(t, X)(\lambda(t)-\lambda_0(t))^2\,dt.$$

It is now very easy to see that, by the assumption (B.2), $\|\lambda-\lambda_0\| < \eta(\varepsilon)$ for $|H(\lambda, T)-H(\lambda_0, T)| < \varepsilon$, where $\eta(\varepsilon)\to 0$ for $\varepsilon\to 0$. After computations similar to those of Theorem 2.1 it is clear that to show Theorem 3.1 it suffices to prove that $(H(\hat{\lambda}_T, T)-H(\lambda_0, T))$ tends to zero in probability. Observe that

$$\text{(3.4)}\quad H(\lambda, T)+L(\lambda, T) = T^{-1}\int_0^T \lambda(t)\lambda_0(t)(a^2(t, X)-\mathrm{E}a^2(t, X))dt$$

$$+(2T)^{-1}\int_0^T \lambda^2(t)(a^2(t, X)-\mathrm{E}a^2(t, X))dt+T^{-1}\int_0^T \lambda(t)a(t, X)dW(t).$$

From the SLLN for martingales (see [23]) it is straightforward to see that the third term of the right-hand side of (3.4) tends to zero almost everywhere. The variance of the first term of the right-hand side of (3.4) may be represented as

$$\text{(3.5)}\qquad T^{-2}\int_0^T\int_0^T \lambda(u)\lambda_0(u)\lambda(v)\lambda_0(v)\mathrm{E}\{\bar{a}^2(t, X)\cdot\bar{a}^2(s, X)\}\,dsdt,$$

where $\bar{a}^2(t, X) = a^2(t, X)-\mathrm{E}a^2(t, X)$.

Applying now the assumptions (B.1) and (B.2) and the same considerations as in the proof of Theorem 2.1 it is easy to see that the term

$$\int_0^T\int_0^T \mathrm{E}\{\bar{a}^2(t, X)\cdot\bar{a}^2(s, X)\}\,dsdt$$

is of the order $O(T)$. Therefore, we get the convergence in probability of the first term of the right-hand side of (3.4). The proof of Theorem 3.1 is now complete since the second term of the right-hand side of (3.4) may be handled analogously. ∎

Remark 3.2. The condition (B.1) for the functional $a(t, X)$ seems to be restrictive, nevertheless, it is fulfilled for the models considered, e.g., by Ibragimov and Has'minskii [11], Nguyen and Pham [22] and Dorogovtsev [7].

Similarly, in the context of the statistical inference for signals the assumption (A.1) or (B.1) guarantees the integrability of the covariance of the signal $a(t, X)$ which is quite essential in applications (see, e.g., [6] or [8]).

REFERENCES

[1] O. O. Aalen, *Statistical inference for a family of counting processes*, Ann. Statist. 6 (1978), pp. 701–726.

[2] L. Amerio and G. Prouse, *Almost-Periodic Functions and Functional Equations*, Van Nostrand Publishing Company, NY, 1971.

[3] J. Beder, *A sieve estimator for the mean of a Gaussian process*, Ann. Statist. 15 (1987), pp. 59–87.

[4] A. S. Besicovitch, *Almost-Periodic Functions*, Dover Publications, NY, 1953.

[5] P. Billingsley, *Convergence of Probability Measures*, Wiley, NY, 1968.

[6] A. V. Dandawate and G. B. Giannakis, *Nonparametric cyclic-polyspectral analysis of AM signals and processes with missing observations*, 26th Conference of Information Sciences and Systems, Princeton University, Princeton, NJ, March 18–20, 1992.

[7] A. Dorogovtsev, *Theory of Estimation of Parameters of Random Processes*, Kiev 1982 (in Russian).

[8] Ya. P. Dragan and I. N. Yavorski, *The rhythmics of the sea waves and the undersea acoustic signals*, Institute of Mechanics and Physics, Academy of Sciences of the Ukrainian Republic, Kiev 1982 (in Russian).

[9] S. Geman and C. R. Hwang, *Nonparametric maximum likelihood estimation by the method of sieves*, Ann. Statist. 10 (1982), pp. 401–414.

[10] U. Grenander, *Abstract Inference*, Wiley, New York 1981.

[11] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation, Asymptotic Theory*, Springer-Verlag, New York 1981.

[12] M. Jacobsen, *Statistical Analysis of Counting Processes*, Lecture Notes in Statist., Vol 12, Springer-Verlag, 1983.

[13] A. F. Karr, *Point Processes and Their Statistical Inference*, M. Dekker Inc., New York 1986.

[14] K. Krickeberg, *Statistical problems on point processes*, Banach Center Publ. Math. Statist., Vol. 6, Institute of Mathematics, Polish Academy of Sciences, 1980, pp. 197–223.

[15] J. Leśkow, *Estimation of a periodic function in the multiplicative intensity model*, Probab. Math. Statist. 8 (1987), pp. 103–110.

[16] — *Sieve methods in estimation of a periodic function of a point process*, Ph. D. Thesis, Institute of Mathematics, Polish Academy of Sciences, Technical Report No 50, Technical University of Wrocław, October 1987.

[17] — *Histogram maximum likelihood estimator of a periodic function in the multiplicative intensity model*, Statist. Decisions 6 (1988), pp. 79–88.

[18] — *A note on kernel regularization of histogram estimator in the multiplicative intensity model*, Statist. Probab. Lett. 7, No 5 (1988).

[19] — and R. Różański, *Estimator of a drift function for diffusion p processes*, Statist. Decisions (1989).

[20] P. A. W. Lewis, *Stochastic Point Processes, Statistical Analysis, Theory and Applications*, Wiley, New York 1972.

[21] R. Liptser and A. Shiryayev, *Statistics of Random Processes*, Moscow 1981.

[22] H. T. Nguyen and D. P. Pham, *Estimation nonparamétrique par la méthode de Grenander dans les modèles de diffusion linéaires et nonstationnaires*, C. R. Acad. Sci. Paris, Sér. A 290 (1980), pp. 197–200.

[23] — *On the law of large numbers for continuous-time martingales and applications to statistics*, Stochastica 6, No 1 (1982), pp. 5–23.

[24] O. Pons and E. Turckheim, *Cox's periodic regression model*, Ann. Statist. 16 (1988), pp. 678–693.

[25] J. Samur, *Convergence of sums of mixing triangular arrays of random vectors with stationary rows*, Ann. Probab. 12 (1984), pp. 390–426.

[26] D. Thompson, *Spatial point processes with application to ecology*, Biometrika 42 (1955), pp. 102–115.

[27] D. Vere-Jones, *On the estimation of frequency in point processes data*, J. Appl. Probab. 19A (1982), pp. 387–394.

[28] — *Inference for earthquake models: a self-correcting model*, Stochastic Process. Appl. 17 (1984), pp. 337–347.

Institute of Mathematics
Technical University of Wrocław
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland

Received on 20.1.1992;
revised version on 27.7.1992