Probab. Th. Rel. Fields 74, 55–69 (1987)

Probability Theory and Related Fields

Estimation of Palm Measures of Stationary Point Processes*

Alan F. Karr

Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, USA

Summary. Estimators of the Palm measure of a stationary point process on a finite-dimensional Euclidean space are developed and shown to be strongly uniformly consistent. From them, similarly consistent estimators of reduced moment measures, the spectral measure, the spectral density function and the underlying probability measure itself are derived. Normal and Poisson approximations to distributions of estimators are presented. Application is made to the problem of combined inference and linear state estimation.

0. Introduction

"Frequency domain" statistical inference for stationary point processes has a reasonably lengthy but also somewhat sporadic history. Such early papers as Bartlett (1963, 1964, 1967) proposed methods of estimation for spectral density functions that are analogous to techniques used for ordinary time series; asymptotic properties, however, were often only incompletely described. With the exception of Brillinger (1972, 1975) relatively little development has occurred since. At the same time, enormous strides have taken place in the theory of stationary point processes, most notably concerning the fundamental role of Palm measures. In addition, recently derived spatial ergodic theorems permit one to obtain new kinds of consistency results.

One purpose of this paper is to apply these recent developments in order to establish strong uniform consistency and asymptotic normality of estimators of spectral measures and spectral density functions of stationary point processes on $\mathbb{R}^d$. These we obtain as consequences of results that extend and refine consistency properties of estimators of Palm measures demonstrated in Kricke-

* Research supported by Air Force Office of Scientific Research, AFSC, grant 82-0029C. The United States Government is authorized to reproduce and distribute reprints for governmental purposes

berg (1982). Our more powerful consistency theorems also allow estimation of the law P of the point process, rather than only the Palm measure P^*.

Organization of the paper is as follows. Section 1 contains background material on stationary point processes, Palm measures, and ordinary and reduced moment and cumulant measures, together with a spectral representation theorem for stationary point processes. In Sect. 2 we present a variety of strong consistency theorems: for Palm measures, reduced second moment measures, spectral measures, spectral density functions and the probability law of the point process; the first four of these establish strong uniform consistency. The setting is completely nonparametric and the underlying space, while Euclidean, is of arbitrary finite dimension. Normal and Poisson process approximations to (scaled) stationary point processes are described in Sect. 3. Our central limit theorem generalizes that of Jolivet (1981). Finally, Sect. 4 applies consistency theorems to the problem of combined statistical inference and linear state estimation.

1. Preliminaries

Let $E=\mathbb{R}^d$, where the dimension d is arbitrary but fixed and for each $x\in E$ let τ_x be the translation operator $y\to\tau_x y=y-x$. Lebesgue measure on E is denoted by dx or λ, as convenient. Let the sample space Ω be the set M_p of locally finite, simple point measures on E, endowed with the Borel σ-algebra $\mathscr{G}$ engendered by the vague topology (see Kallenberg 1983); let N be the coordinate point process $N(\omega)=\omega$. For each x define θ_x: $\Omega\to\Omega$ by

$$N\circ\theta_x=N\tau_x^{-1}\ ; \tag{1.1}$$

note that $\theta_0=I$, the identity mapping on Ω, and that $\theta_x\theta_y=\theta_{x+y}$ for each x and y. Given a probability P on $(\Omega,\mathscr{G})$, N is *stationary* with respect to P if the flow (θ_x) is measure-preserving for P: $P\theta_x^{-1}=P$ for all $x\in E$. The formulation is due to Neveu (1977), to which the reader is referred for details; equivalently, N is P-stationary if and only if $N\tau_x^{-1}=N$ in P-distribution for each x. An event $\Gamma\in\mathscr{G}$ is *invariant* if $\theta_x^{-1}\Gamma=\Gamma$ for all x; the probability P is *ergodic* if $P(\Gamma)=0$ or 1 for every invariant event Γ. Our consistency results are proved within the statistical model $\mathscr{P}$ of ergodic probabilities under which N is stationary. For asymptotic normality we require further assumptions – finiteness of reduced cumulant measures – that bring about only weak dependence of distantly separated portions of N. In all cases our asymptotics pertain to single realizations of N observed over increasingly large compact, convex subsets of $\mathbb{R}^d$.

For each probability P under which N is stationary there exists (Neveu 1977, Theorem II.4) a unique σ-finite measure P^* on Ω, the *Palm measure* of P, satisfying

$$E[\int H(\theta_x\omega,x)\,N(\omega,dx)]=E^*[\int H(\omega,x)\,dx] \tag{1.2}$$

for every bounded, measurable function H: $\Omega\times E\to\mathbb{R}$, where E^* denotes "expectation" with respect to P^*. When $P^*(\Omega)<\infty$ the probability $P_0(\cdot)$

$=P^*(\cdot)/P^*(\Omega)$ is the *Palm distribution;* its heuristic interpretation is that $P_0(\cdot)=P\{\cdot|N(\{0\})=1\}$, i.e., P_0 is the distribution of N conditional on the (null) event that a point is located at the origin.

Moment measures are also important. For each k, let N^k be the k-fold product measure

$$N^k(dx_1, \ldots, dx_k) = N(dx_1) \ldots N(dx_k);$$

then N admits a moment of order k with respect to P if the measure

$$\mu^k(dx_1, \ldots, dx_k) = E[N^k(dx_1, \ldots, dx_k)] \tag{1.3}$$

is *locally* finite, in which case μ^k is termed the *moment measure* of order k. For $z \in E^{k-1}$ let λ_z be the image of Lebesgue measure λ under the mapping $x \to (z_1+x, \ldots, z_{k-1}+x, x)$ of E into E^k. By stationarity, each extant moment measure μ^k admits a disintegration (see Krickeberg 1974, 1982)

$$\mu^k = \int_{E^{k-1}} \lambda_z \, \mu_*^k(dz), \tag{1.4}$$

where μ_*^k, a measure on E^{k-1}, is the *reduced moment measure* of order k. In particular, for $k=1$, $E^{k-1}=\{0\}$ and $\mu_*^1 = \nu \varepsilon_0$, with the scalar ν known as the *intensity* of N.

Cumulant and reduced cumulant measures are defined analogously. If N admits a moment of order k then the measure

$$\gamma^k \left(\prod_{j=1}^{k} f_j\right) = \sum (-1)^{|\mathcal{J}|-1} (|\mathcal{J}|-1)! \prod_{\ell} \mu^{J_\ell} \left(\prod_{j \in J_\ell} f_j\right) \tag{1.5}$$

is the *cumulant measure* of order k. In (1.5), $(\prod f_j)(x) = \prod f_j(x_j)$ and the summation is over all partitions $\mathcal{J} = \{J_1, \ldots, J_{|\mathcal{J}|}\}$ of $\{1, \ldots, k\}$. The *reduced* cumulant measure γ_*^k satisfies

$$\gamma^k = \int_{E^{k-1}} \lambda_z \, \gamma_*^k(dz); \tag{1.6}$$

it is a signed measure in general, but whereas moment measures, reduced moment measures and cumulant measures are not finite except in trivial cases, reduced cumulant measures may be finite, in the sense of having finite total variation; this finiteness (which is analogous to integrability of the covariance function of a stationary process on $\mathbb{R}$) implies that distant parts of N are nearly independent.

The *covariance measure* ρ is the reduced cumulant measure of order two; for it the disintegration (1.6) becomes

$$\rho_*(dx) = \mu_*^2(dx) - \nu^2 \, dx, \tag{1.7}$$

so that estimation procedures applicable to reduced moment measures yield – by substitution – estimates of the reduced covariance measure; these are applied in Sect. 4 to the problem of combined inference and linear state estimation. When N is a Poisson process with intensity ν, then $\rho_* = \nu \varepsilon_0$, a manifestation of the independent increments property of N.

The crucial relation between Palm measures and reduced moment measures is that the latter are ordinary moment measures with respect to the former.

Lemma 1.1. *Let N be a stationary point process admitting moment of order $k \geqq 2$. Then*

$$\mu_*^k = E^*[N^{k-1}]. \tag{1.8}$$

The proof is given, for example, in Krickeberg (1982), as is that of Lemma 1.2 below. Under the convention that $E^0 = \{0\}$, so that $\mu_*^1 = \nu \varepsilon_0$, (1.8) obtains for $k=1$ as well, becoming $\nu = P^*(\Omega)$. Thus $P^*(\Omega) < \infty$ if and only if N admits a moment of order 1.

Frequency domain analysis of a stationary point process is based on the spectral representation. Denote by $\mathscr{D}$ the class of infinitely differentiable functions on E with compact support and denote the Fourier transform of $\psi \in \mathscr{D}$ by

$$\tilde{\psi}(v) = \int e^{i\langle v, x\rangle} \psi(x)\, dx, \tag{1.9}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{R}^d$; $\tilde{\psi}$ belongs to the class $\mathscr{S}$ of rapidly decreasing functions on E (see Rudin 1973, or Yoshida 1968). The inverse Fourier transform of $\psi \in \mathscr{S}$ is

$$\hat{\psi}(x) = (1/2\pi) \int e^{-i\langle v, x\rangle} \psi(v)\, dv. \tag{1.10}$$

If N is a stationary point process admitting moment of order two, there exists (see Itô 1955, and also Daley 1971 and Vere-Jones 1974) a unique complex-valued random measure Z on E, with orthogonal increments, such that for $\psi \in \mathscr{D}$,

$$N(\psi) = \int_E \tilde{\psi}(v)\, Z(dv); \tag{1.11}$$

this is the *spectral representation* of N. The measure

$$F(dv) = E[|Z(dv)|^2] \tag{1.12}$$

is the *spectral measure* of N.

As might be anticipated, the spectral measure and reduced second moment measure are linked intimately; this link is used in order to estimate spectral measures.

Lemma 1.2. *Let N be a stationary point process with moment of order two. Then the reduced second moment measure and spectral measure fulfill the Parseval relations*

$$\mu_*^2(\psi) = F(\tilde{\psi}) \tag{1.13}$$

and

$$F(\psi) = \mu_*^2(\hat{\psi}) \tag{1.14}$$

for $\psi \in \mathscr{D}$.

We conclude the section by describing the nonparametric estimators whose properties are developed in the remaining sections. Let N be a stationary point process with unknown probability law P, and suppose that a single realization of N has been observed over a compact, convex subset K of $\mathbb{R}^d$. The fundamental estimators are those (see Krickeberg 1982), of the Palm measure: for

$H: \Omega \to \mathbb{R}_+$, $P^*(H) = \int H \, dP^*$ is estimated by

$$\hat{P}^*(H) = \frac{1}{\lambda(K)} \int_K H(N \tau_x^{-1}) \, N(dx). \tag{1.15}$$

While a "method of moments" justification stems from (1.2):

$$E[\hat{P}^*(H)] = \lambda(K)^{-1} E[\int_K H(N \circ \theta_x) \, N(dx)]$$
$$= \lambda(K)^{-1} E^*[H(N) \int_K dx] = E^*[H(N)] = P^*(H),$$

in other words these estimators are unbiased, a more compelling justification is based on the conditioning interpretation of P^*. For each $x \in K$ that is a point of N – and only such x contribute to the integral in (1.15) – $N \tau_x^{-1}$ has a point at the origin, and therefore $\hat{P}^*(H)$ is just a weighted average of H-values of translations of N placing each point in turn at the origin. In (1.15) and throughout the paper we suppress dependence on the "sample size" K, but this interpretation is only rather loose because $\hat{P}^*(H)$ may not be measurable with respect to the σ-algebra $\mathscr{F}^N(K) = \sigma(N(B)\colon B \subset K)$ representing observation of N over K. In many specific cases, there is a bounded set A (depending on H) such that $H(\mu) = H(\mu_A)$, where μ_A is the restriction of μ to A, which renders $\hat{P}^*(H)$ measurable with respect to $\mathscr{F}^N(K + A)$. This is not true, however, of the estimators $\hat{F}$ and $\hat{f}_\rho$ below.

The remaining estimators are derived from $\hat{P}^*$ by substitution. For estimation of the intensity, taking $H \equiv 1$ in (1.15) yields the (obvious) estimator $\hat{\nu} = N(K)/\lambda(K)$, which is $\mathscr{F}^N(K)$-measurable. For estimation of the reduced second moment measure, choosing $H(N) = N(f)$, where $f \in C_+(E)$ (the set of positive, continuous functions on E with compact support), gives

$$\hat{\mu}_*^2(f) = \frac{1}{\lambda(K)} \int_K N(dx) \int f(y-x) \, N(dy) \tag{1.16}$$

as $\mathscr{F}^N((\operatorname{supp} f) - K)$-measurable estimator of $\mu_*^2(f) = \int f \, d\mu_*^2$. Using (1.14) and (1.16) we then obtain

$$\hat{F}(\psi) = \hat{\mu}_*^2(\hat{\psi}) = \frac{1}{\lambda(K)} \int_K N(dx) \int \hat{\psi}(y-x) \, N(dy) \tag{1.17}$$

as estimator of the spectral measure F.

The appropriate spectral density function is not that of F, but rather that of the covariance spectral measure F_ρ, which satisfies the Parseval relation

$$\rho_*(\psi) = F_\rho(\tilde{\psi}). \tag{1.18}$$

It follows that the covariance *spectral density function* $f_\rho(v) = (dF_\rho/d\lambda)(v)$ satisfies

$$f_\rho(v) = \frac{1}{2\pi} E^*[\int e^{-i\langle v, y\rangle} (N - \nu \lambda)(dy)], \tag{1.19}$$

and therefore we have substitution estimators

$$\hat{f}_\rho(v)=\frac{1}{2\pi\,\lambda(K)}\int_K N(dx)\int e^{-i\langle v,y-x\rangle}(N-v\,\lambda)\,(dy). \tag{1.20}$$

A shared shortcoming of the estimators $\hat{F}$ and $\hat{f}_\rho$ is that they cannot be calculated from observation of N over compact sets. One method of approximation by estimators requiring only observation over compact sets is truncation. For example, the estimator $\hat{F}$ would be replaced by

$$\hat{F}'(\psi)=\frac{1}{\lambda(K)}\int_K N(dx)\int_K \hat{\psi}(y-x)\,N(dy).$$

Since

$$|\hat{F}(\psi)-\hat{F}'(\psi)|\leqq\frac{1}{\lambda(K)}\int_K N(dx)\int_{\mathbb{R}^d\setminus K}|\hat{\psi}(y-x)|\,N(dy),$$

and since $\hat{\psi}$ is rapidly decreasing, asymptotic properties of the $\hat{F}'$ will - possibly under mild additional assumptions - be those of the $\hat{F}$.

Finally we investigate estimation of P itself, using estimators

$$\hat{P}(H)=\frac{1}{v\,\lambda(K)^2}\int_K N(dx)\int_K H(N\,\tau_{x-y}^{-1})\,dy. \tag{1.21}$$

For these and all of the other estimators, we establish strong consistency.

2. Strong Consistency Theorems

For K bounded and convex, let $\delta(K)$ be the supremum of the radii of Euclidean balls contained in K. In order to have the "infinitely much" data necessary for consistent estimation we shall require that $\delta(K)\to\infty$; the crucial geometric consequence of convexity is that convex sets grow more rapidly than their boundaries: for every $\varepsilon>0$, $\lambda((\partial K)^\varepsilon)/\lambda(K)\to 0$, where $(\partial K)^\varepsilon$ is the set of points within distance ε of the boundary of K.

Here is our main consistency theorem, for the estimators $\hat{P}^*$.

Theorem 2.1. *Assume that P is ergodic and that the intensity v is positive and finite, and let $\mathscr{H}$ be a uniformly bounded set of continuous functions on Ω that is compact in the topology of uniform convergence on compact subsets. Then almost surely with respect to P,*

$$\lim_{\delta(K)\to\infty}\ \sup_{H\in\mathscr{H}}|\hat{P}^*(H)-P^*(H)|=0, \tag{2.1}$$

where $\hat{P}^*(H)$ is given by (1.15).

Proof. We combine appeal to the spatial ergodic theorem of Nguyen and Zessin (1979) (as in Krickeberg 1982) with arguments adapted from Karr (1985). First let H be a fixed element of $C(\Omega)$; then the random measure

$$M(A)=\int_A H(N\,\tau_x^{-1})\,N(dx)=\int_A H(N\circ\theta_x)\,N(dx)$$

is covariant with respect to (θ_x), i.e., $M(\tau_x A)\circ\theta_x = M(A)$ for all A and x. It follows that Proposition 4.23 of Nguyen and Zessin (1979), which even though the argument provided there requires minor emendations, *is* correct, applies, with the consequence that

$$\lim_{\delta(K)\to\infty} \hat{P}^*(H) = \lim_{\delta(K)\to\infty} M(K)/\lambda(K) = E[M([0,1]^d)] = P^*(H) \tag{2.2}$$

almost surely (and in $L^1(P)$ as well, but we do not pursue this aspect). In particular, from the choice $H\equiv 1$ we infer strong consistency of the estimators $\hat{\nu} = N(K)/\lambda(K)$:

$$\lim_{\delta(K)\to\infty} \hat{\nu} = \lim_{\delta(K)\to\infty} N(K)/\lambda(K) = \nu \tag{2.3}$$

almost surely with respect to P.

Turning to the set $\mathscr{H}$, given $\varepsilon>0$ there exists by finiteness of P^* (recall that $P^*(\Omega)=\nu$) a compact subset Γ of Ω such that $P^*(\Omega\setminus\Gamma)<\varepsilon$; we may without loss of generality suppose that Γ is a P^*-continuity set, i.e., $P^*(\partial\Gamma)=0$. Moreover, there exist $\tilde{H}_1,\ldots,\tilde{H}_L\in\mathscr{H}$ such that to each $H\in\mathscr{H}$ there corresponds $\ell(H)\in\{1,\ldots,L\}$ for which $\|H-\tilde{H}_{\ell(H)}\|_\Gamma$ $(=\sup\{|H(\omega)-\tilde{H}_{\ell(H)}(\omega)|:\ \omega\in\Gamma\})<\varepsilon$. Then assuming, as we may, that (2.2) holds for $\tilde{H}_1,\ldots,\tilde{H}_L$, and that (2.3) holds as well, in the decomposition

$$\begin{aligned}\sup_{H\in\mathscr{H}}|\hat{P}^*(H)-P^*(H)| \leqq &\sup_{H\in\mathscr{H}}|\hat{P}^*(H)-\hat{P}^*(\tilde{H}_{\ell(H)})|\\ &+\max_{\ell}|\hat{P}^*(\tilde{H}_\ell)-P^*(\tilde{H}_\ell)|\\ &+\sup_{H\in\mathscr{H}}|P^*(\tilde{H}_{\ell(H)})-P^*(H)|,\end{aligned} \tag{2.4}$$

the second term converges to zero almost surely by (2.2). Concerning the third, for each H,

$$\begin{aligned}|P^*(\tilde{H}_{\ell(H)})-P^*(H)| &\leqq |P^*(\tilde{H}_{\ell(H)}1_\Gamma)-P^*(H\,1_\Gamma)|\\ &\quad+|P^*(\tilde{H}_{\ell(H)}1_{\Gamma^c})-P^*(H\,1_{\Gamma^c})|\\ &\leqq \varepsilon P^*(\Gamma)+P^*(\Gamma^c)\sup_{H\in\mathscr{H}}\|H\|_\infty\\ &\leqq \varepsilon(\nu+\sup_{H\in\mathscr{H}}\|H\|_\infty),\end{aligned} \tag{2.5}$$

so that this term can be made arbitrarily small by proper choice of ε. Finally, by straightforward arguments (Karr 1979), (2.4) implies that almost surely $\hat{P}^*\to P^*$ vaguely as Radon measures on Ω, and hence also weakly, by appeal to (2.3). Because Γ is a P^*-continuity set, almost surely $\hat{P}^*(\Gamma^c)<\varepsilon$ for all sufficiently large K; consequently

$$\begin{aligned}\sup_{H\in\mathscr{H}}|\hat{P}^*(H)-\hat{P}^*(\tilde{H}_{\ell(H)})| &\leqq \varepsilon\hat{P}^*(\Gamma)+\hat{P}^*(\Gamma^c)\sup_{H\in\mathscr{H}}\|H\|_\infty\\ &\leqq \varepsilon(2\nu+\sup_{H\in\mathscr{H}}\|H\|_\infty)\end{aligned}$$

once $\delta(K)$ is large enough, which completes the proof. □

Strong consistency of the estimators $\hat{\mu}_*^2$ of (1.16) of the reduced second moment measure will be shown in two forms, the first more intrinsically useful for estimation of μ_*^2 and the second directed at estimation of the spectral measure.

Theorem 2.2. *Assume that P is ergodic and that under P, N admits a moment of order two, and let $\hat{\mu}_*^2$ be given by* (1.16). *Let $\mathscr{K}$ be a compact, uniformly bounded subset of $C_+(E)$, each element of which is supported in the same compact subset K_0 of E. Then almost surely*

$$\lim_{\delta(K)\to\infty} \sup_{f\in\mathscr{K}} |\hat{\mu}_*^2(f)-\mu_*^2(f)|=0. \tag{2.6}$$

Proof. Given $f\in\mathscr{K}$, define $H_f\colon \Omega\to\mathbb{R}$ by $H_f(\mu)=\mu(f)$. Then H_f is continuous, and the proof will be effected by showing that the mapping $f\to H_f$ of $C_+(E)$ into the set of bounded, continuous functions on Ω is itself continuous, for then $\mathscr{H}=\{H_f\colon f\in\mathscr{K}\}$ is the continuous image of a compact set, and is hence compact; at this point, (2.6) follows from (2.1). Let Γ be a compact subset of Ω; then $a=\sup\{\mu(K_0)\colon \mu\in\Gamma\}$ is finite, and consequently for f, $g\in\mathscr{K}$

$$\sup_{\mu\in\Gamma}|H_f(\mu)-H_g(\mu)|=\sup_{\mu\in\Gamma}|\mu(f)-\mu(g)|\leqq a\|f-g\|_\infty,$$

which verifies the requisite continuity. □

Theorem 2.3. *Assume that P is ergodic and that N admits a moment of order two. Let $\mathscr{K}$ be a compact subset of $\mathscr{D}$, all of whose elements are supported in the same compact subset K_0 of E. Then almost surely*

$$\lim_{\delta(K)\to\infty} \sup_{\psi\in\mathscr{K}} |\hat{\mu}_*^2(\hat{\psi})-\mu_*^2(\hat{\psi})|=0. \tag{2.7}$$

Proof. Since $\hat{\psi}$ does not have compact support, (2.7) *does not follow* from (2.6). Instead, we follow with minor modifications the *reasoning* used to prove Theorem 2.2. In view of (2.3) we first replace $\Omega=M_p$ in the proof of Theorem 2.1 by $M_p(v)=\{\mu\colon \lim\mu(K)/\lambda(K)=v\}$. As in the proof of Theorem 2.2, define $H_\psi(\mu)=\mu(\hat{\psi})$, $\psi\in\mathscr{K}$; then it suffices to show that the mapping $\psi\to H_\psi$ is continuous, for then we may appeal to a minor alteration of Theorem 2.1 in order to conclude the proof. By compactness of $\mathscr{K}$ and continuity of the inverse Fourier transform (Rudin 1973, Theorem 7.7), given $\varepsilon>0$ there is a compact subset K_1 of $\mathbb{R}^d$ such that

$$\sup_{\psi\in\mathscr{K}} \int_{K_1^c} |\hat{\psi}(x)|\,dx<\varepsilon. \tag{2.8}$$

Given a compact subset Γ of $M_p(v)$, for all $\psi, \phi\in\mathscr{K}$

$$\begin{aligned}\sup_{\mu\in\Gamma}|H_\psi(\mu)-H_\phi(\mu)|&=\sup_{\mu\in\Gamma}|\mu(\hat{\psi})-\mu(\hat{\phi})|\\&\leqq\sup_{\mu\in\Gamma}\left|\int_{K_1}\hat{\psi}\,d\mu-\int_{K_1}\hat{\phi}\,d\mu\right|\\&\quad+\sup_{\mu\in\Gamma}\left|\int_{K_1^c}\hat{\psi}\,d\mu-\int_{K_1^c}\hat{\phi}\,d\mu\right|\\&\leqq[\sup_{\mu\in\Gamma}\mu(K_1)]\,\|\hat{\psi}-\hat{\phi}\|_\infty+2\varepsilon\end{aligned}$$

(by (2.3) and (2.8))

$$\leq [\sup_{\mu\in\Gamma} \mu(K_1)] \, \|\psi-\phi\|_1 + 2\varepsilon$$

(Rudin 1973, Theorem 7.5)

$$\leq [\sup_{\mu\in\Gamma} \mu(K_1)] \, \|\psi-\phi\|_\infty \lambda(K_0) + 2\varepsilon,$$

which gives the necessary continuity since K_1 depends on neither ψ nor ϕ. □

Corollary 2.4. *Let $\mathscr{K}$ be as in Theorem* 2.3 *and let $\hat{F}$ be the estimator of the spectral measure given by* (1.17). *Then almost surely*

$$\lim_{\delta(K)\to\infty} \sup_{\psi\in\mathscr{K}} |\hat{F}(\psi) - F(\psi)| = 0. \tag{2.9}$$

By the same pattern of reasoning, that is, because the mapping of $v \in \mathbb{R}^d$ into the functional

$$H_v(\mu) = \int e^{-i\langle v, y\rangle} (\mu - v\lambda)(dy)$$

on $M_p(v)$ is continuous, we obtain the following consistency theorem for the estimators $\hat{f}_\rho(v)$ of the spectral density function.

Theorem 2.5. *Assume that P is ergodic and that under P, N admits covariance spectral density function f_ρ satisfying* (1.19). *Let $\hat{f}_\rho$ be given by* (1.20); *then for each compact subset K_0 of E, almost surely*

$$\lim_{\delta(K)\to\infty} \sup_{v\in K_0} |\hat{f}_\rho(v) - f_\rho(v)| = 0. \tag{2.10}$$

It is instructive to compare the estimators $\hat{f}_\rho$ with periodogram estimators commonly used in statistical analysis of stationary point processes (see for example Brillinger 1975; or Cox and Lewis 1966). In our setting the periodogram is given by

$$\hat{f}(v) = \frac{1}{2\pi\lambda(K)} \left|\int_K e^{-i\langle v, x\rangle} N(dx)\right|^2.$$

Even for $E = \mathbb{R}$ and $K = [0, T]$ with $T \to \infty$, the periodogram is not a consistent estimator of the spectral density function, which is related to second moment properties of the periodogram. By contrast, the estimators $\hat{f}_\rho$ of (1.20) are strongly uniformly consistent in the sense of (2.10) - but at the price (even if truncation is imposed) that their computation is quadratic in $N(K)$, rather than linear. Thus neither estimator seems clearly superior.

Finally we consider estimation of the probability law P itself. Even though P is uniquely determined by the Palm measure P^* and even though, as the preceding development confirms, many of the main functionals of P of interest in inference are easily expressed as functionals of P^* as well, estimation of P remains an important problem. Our estimators $\hat{P}(H)$, given by (1.21), are motivated by the identity

$$E[N(K) H(N)] = E^*\left[\int_K H(N\tau_{-y}^{-1}) \, dy\right], \tag{2.11}$$

which follows at once from (1.2).

In the following theorem we establish strong - but not uniform - consistency of the estimators (1.21); indeed strong consistency in Theorem 2.6 requires the full force of uniformity in Theorem 2.1.

Theorem 2.6. *Assume that P is ergodic and that the intensity is positive and finite, and let H be a bounded, continuous function on Ω. Then for the estimators*

$$\hat{P}(H)=\frac{1}{\nu\,\lambda(K)^2}\int_K N(dx)\int_K H(N\,\tau_{x-y}^{-1})\,dy,$$

we have $\hat{P}(H)\to P(H)=\int H\,dP$ almost surely as $\delta(K)\to\infty$.

Proof. For each K, by (2.11)

$$\begin{aligned}&\left|\frac{1}{\nu\,\lambda(K)^2}\int_K N(dx)\int_K H(N\,\tau_{x-y}^{-1})\,dy-P(H)\right|\\ &\quad\leqq\left|P(H)-\frac{1}{\nu}E[H(N)\,N(K)/\lambda(K)]\right|\\ &\quad\quad+\frac{1}{\nu}\left|\frac{1}{\lambda(K)}\int_K E^*[H(N\,\tau_y^{-1})]\,dy-\frac{1}{\lambda^2(K)}\int_K N(dx)\int_K H(N\,\tau_{x-y}^{-1})\,dy\right|.\end{aligned}\tag{2.12}$$

Since $N(K)/\lambda(K)\to\nu$ in L^1 (see discussion in the proof of Theorem 2.1), $\nu^{-1}E[H(N)N(K)/\lambda(K)]\to E[H(N)]=P(H)$, so that the first term in (2.12) converges to zero. By Theorem 2.1 applied to the family $\mathscr{H}=\{\mu\to H(\mu\,\tau_{-y}^{-1})\}$ (we omit the straightforward verification of the hypotheses) we infer that

$$\frac{1}{\lambda(K)}\int_K H(N\,\tau_{x-y}^{-1})\,N(dx)\to E^*[H(N\,\tau_{-y}^{-1})]\tag{2.13}$$

uniformly in y, which by an analytical argument (for which uniformity in (2.13) is crucial) implies that the second term in (2.12), whose components are $\lambda(K)^{-1}$ times the dy-integrals of the two sides of (2.13) over K, converges to zero almost surely. □

3. Normal and Poisson Approximations

In this section we present a central limit theorem and Poisson approximation theorem complementing the consistency theorems of Sect. 2. For brevity we work only in the context of Theorem 2.1, which is, after all, the principal result in Sect. 2.

Our central limit theorem extends the conclusion of the central limit theorem of Jolivet (1981) for estimators or reduced moment measures (of which the estimators (1.16) correspond to the reduced second moment measure), but does not weaken the hypotheses.

Theorem 3.1. *Suppose that P is ergodic, that under P moments of N of every order exist, and that each reduced cumulant measure γ_*^k of order $k\geqq 2$ has finite* *total variation. Then there exists a centered Gaussian process* $\{G(H): H \in C(\Omega)\}$ *such that for each* H,

$$\lambda(K)^{\frac{1}{2}}[\hat{P}^*(H) - P^*(H)] \xrightarrow{d} G(H), \tag{3.1}$$

where $\xrightarrow{d}$ *denotes convergence in* (P-) *distribution.*

Proof. Consider the class A of functions H of the form

$$H(\mu) = \sum_{j=1}^{n} c_j \mu^{k_j - 1}(f_j)^{d_j}, \tag{3.2}$$

where the c_j are real constants, k_j and d_j are positive integers and $f_j \in C(E^{k_j - 1})$. This class A of "polynomials" is a vector space and an algebra (i.e., is closed under pointwise multiplication) and evidently separates the points of Ω; consequently by the Stone-Weierstrass theorem, the (uniform) closure of A is $C(\Omega)$. It suffices, therefore, to show that (3.1) holds whenever H has the form (3.2). By the continuous mapping theorem, this last assertion holds if for each n, $k_1, \ldots, k_n$ and $f_1, \ldots, f_n$,

$$\lambda(K)^{\frac{1}{2}}[(\hat{\mu}_*^{k_1}(f_1), \ldots, \hat{\mu}_*^{k_n}(f_n)) - (\mu_*^{k_1}(f_1), \ldots, \mu_*^{k_n}(f_n))] \xrightarrow{d} (G(H_1), \ldots, G(H_n)), \tag{3.3}$$

where $H_j(\mu) = \mu^{k_j - 1}(f_j)$ and where $\hat{\mu}_*^k(f) = \lambda(K)^{-1} \int_K N^{k-1} \circ \theta_x(f)\, N(dx)$. Using the Cramér-Wold device we can reduce (3.3) to show that for each k and f,

$$\lambda(K)^{\frac{1}{2}}[\hat{\mu}_*^k(f) - \mu_*^k(f)] \xrightarrow{d} G(H_f), \tag{3.4}$$

where $H_f(\mu) = \mu^{k-1}(f)$, but (3.4) holds, in the presence of our hypotheses, by Jolivet (1981), Theorem 1. □

The hypotheses of Theorem 3.1 are rather severe, in part because of the generality of the convergence condition that $\delta(K) \to \infty$. Another shortcoming of normal approximations in general is that they are ineffective for estimation of small probabilities. Poisson approximations, by contrast *can* estimate small probabilities and, moreover, have a lengthy history (see for example, Çinlar 1972) in the context of point processes. Unfortunately, however, in the following theorem the severity of the assumptions in Theorem 3.1 is not mitigated.

For each $r > 0$ let B_r be the closed ball of radius r centered at the origin.

Theorem 3.2. *Let* Γ_r, $r > 0$, *be decreasing events for which there exists a finite measure* ξ *on* Ω *such that*

$$\lim_{r \to \infty} r^d P^*(\cdot \cap \Gamma_r) = \xi(\cdot) \tag{3.5}$$

in the sense of weak convergence. For each r *let* N_r *be the point process on* $\Omega \times B_1$ *defined by*

$$N_r = \sum 1(N \tau_{X_i}^{-1} \in \Gamma_r)\, 1(X_i \in B_r)\, \varepsilon_{(N \tau_{X_i}^{-1}, X_i/r)}. \tag{3.6}$$

If the hypotheses of Theorem 3.1 are satisfied, then as $r \to \infty$, $N_r \xrightarrow{d} \bar{N}$, where $\bar{N}$ is a Poisson process with mean measure $\eta(\Gamma \times B) = \xi(\Gamma)\lambda(B)$.

Proof. We verify first that $E[N_r]\to\eta$ in the sense of weak convergence; indeed for Γ a ξ-continuity set,

$$E[N_r(\Gamma\times B)]=E[\int_{B_r} 1(N\tau_x^{-1}\in\Gamma\cap\Gamma_r)\,1(x\in rB)\,N(dx)]$$

(where $rB=\{rx\colon x\in B\}$)

$$=P^*(N\in\Gamma\cap\Gamma_r)\,\lambda(B_r\cap rB)$$

(by (1.2))

$$=P^*(N\in\Gamma\cap\Gamma_r)\,r^d\,\lambda(B_1\cap B)$$

$$=P^*(N\in\Gamma\cap\Gamma_r)\,r^d\,\lambda(B)\to\xi(\Gamma)\,\lambda(B).$$

It follows from this computation (cf. Kallenberg 1983, Lemma 4.5) that (N_r) is tight and hence it suffices to show that for any "subsequence" $(N_{r'})$ converging in distribution the limit is Poisson with mean measure η. For this, it is sufficient by (Kallenberg 1983, Theorem 4.7) to verify that

$$P\{N_r(\Gamma\times B)=0\}\to e^{-\eta(\Gamma\times B)} \tag{3.7}$$

for Γ a ξ-continuity set and B a Borel subset of B_1. Under the assumptions of Theorem 3.1, in the manner of Jolivet (1981), one may use (3.5) and the computational rules of Leonov and Shiryayev (1959) to evaluate the cumulants of $N_r(\Gamma\times B)$; with computational details omitted, the result is that for each ℓ^{th}-order cumulant c_ℓ of $N_r(\Gamma\times B)$ converges to $\eta(\Gamma\times B)$. Consequently (3.7) holds. □

4. Combined Inference and Linear State Estimation

In this section we apply the estimation procedures developed in Sects. 1 and 2 to construct approximations to minimum mean squared error (MMSE) linear state estimators of unobserved portions of N, when the probability P – under which N is stipulated to be stationary – is unknown.

Let us first suppose that P were known, and introduce the centered process

$$M(f)=N(f)-\nu\,\lambda(f)=\int f\,dN-\nu\int f\,d\lambda, \tag{4.1}$$

where ν is the P-intensity of N. The linear state estimation problem is this: given data $\mathscr{F}^N(A)$ representing observation of N over a bounded set A with $\lambda(A)>0$ and a function f (without loss of generality, vanishing on A), calculate that function $\hat f$ on A for which

$$E[(M(f)-M(\hat f))^2]\leqq E[(M(f)-M(g))^2] \tag{4.2}$$

for every function g on A. Thus $M(\hat f)$ is the optimal (in the MMSE sense) linear predictor of the unobservable random variable $M(f)$ given the observations $\mathscr{F}^N(A)$, and hence $M(\hat f)+\nu\,\lambda(\hat f)$ is the optimal linear state estimator of $N(f)$. With P known, derivation of $\hat f$ is straightforward.

Proposition 4.1. *Assume that N admits a moment of order two under P and let ρ_* be the reduced covariance measure. Given f vanishing on A, the function $\hat f$*

satisfying (4.2) *is the unique function such that*

$$\int_A \hat{f}(y)\,\rho_*(dy-x) = \int f(y)\,\rho_*(dy-x) \tag{4.3}$$

for λ-almost all $x \in A$.

Proof. By standard Hilbert space theory, $M(\hat{f})$ is the projection of $M(f)$ onto the linear space spanned by $\{M(g)\colon g \in L^2(A)\}$, and consequently the unique solution of the normal equations.

$$E[(M(\hat{f})-M(f))\,M(g)]=0, \qquad g \in L^2(A). \tag{4.4}$$

With ρ the ordinary covariance measure,

$$\begin{aligned} E[M(f)\,M(g)] &= \int g(x)\,f(y)\,\rho(dx,dy) \\ &= \int_A g(x)\,[\int f(y)\,\rho_*(dy-x)]\,dx; \end{aligned}$$

consequently (4.4) is equivalent to

$$0=\int_A g(x)\,[\int_A \hat{f}(y)\,\rho_*(dy-x) - \int f(y)\,\rho_*(dy-x)]\,dx,$$

confirming (4.3). □

Suppose now that P is not known, specifically that ρ_* is unknown; nevertheless state estimation may be equally as important as when P is known. We shall construct "pseudo"-state estimators that approximate the "true" state estimators $M(\hat{f})$ arising from (4.3) and describe their asymptotic behavior. More precisely, let $f \in C(E)$ be fixed (recall that f has compact support) and suppose that N is observed over compact, convex sets K such that $K \cap (\operatorname{supp} f) = \emptyset$. We then construct, using estimators $\hat{\rho}_* = \hat{\mu}_*^2 - \hat{\nu}^2\lambda$ of ρ_* based on the observations $\mathscr{F}^N(K)$, estimators $\tilde{f}$ of the solution to (4.3) with $A=K$, and establish that $M(\hat{f}-\tilde{f}) \to 0$ (in an appropriate sense) as $\delta(K) \to \infty$.

The estimator

$$\hat{\rho}_* = \hat{\mu}_*^2 - \hat{\nu}^2\,\lambda \tag{4.5}$$

is obtained by substituting into the identity (1.7) the estimators $\hat{\nu} = N(K)/\lambda(K)$ and $\hat{\mu}_*^2$, the latter given by (1.16). Given $\hat{\rho}_*$, let $\tilde{f} = \tilde{f}_K$ be the unique function on K minimizing

$$\|\int \tilde{f}(y)\,\tilde{\rho}_*(dy-\cdot) - \int f(y)\,\tilde{\rho}_*(dy-\cdot)\|_2. \tag{4.6}$$

As pseudo-state estimator based on the observations $\mathscr{F}^N(K)$ we then take

$$\hat{M}(f) = \int_K \tilde{f}\,dN - \hat{\nu}\,\lambda(\tilde{f}\,1_K). \tag{4.7}$$

While the function $\hat{f}$ in Proposition 4.1 seems to depend on A there, in fact it does not: there exists a single function $\hat{f} \in L^2(E)$ such that for each A, $\hat{f}_A = \hat{f}\,1_A$ is the solution to (4.3). Consequently $\hat{M}(f)$ is an approximation to the true state estimator.

$$M(\hat{f})=\int_K \hat{f}\,dN-\nu\,\lambda(\hat{f}1_K). \tag{4.8}$$

Theorem 4.2. *Assume that P is ergodic and that the reduced covariance measure ρ_* has finite total variation. Then as $\delta(K)\to\infty$, $\hat{M}(f)-M(\hat{f})\to 0$ almost surely.*

Proof. We begin with the decomposition

$$\hat{M}(f)-M(\hat{f})=\int_K(\tilde{f}-\hat{f})dN+(\nu-\hat{\nu})\,\lambda(\hat{f}1_K)+\hat{\nu}[\lambda(\hat{f}1_K)-\lambda(\tilde{f}1_K)]. \tag{4.9}$$

The second term is dealt with most easily; by ergodicity of P, $\hat{\nu}\to\nu$ almost surely and hence since $\hat{f}\in L^2(E)$ this term converges to zero as $\delta(K)\to\infty$.

We next show that almost surely $\tilde{f}-\hat{f}$ converges to zero in L^2. Indeed,

$$\begin{aligned}
&\|\textstyle\int\tilde{f}(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\rho_*(dy-\cdot)\|_2\\
&\quad\leqq\|\textstyle\int\tilde{f}(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\hat{\rho}_*(dy-\cdot)\|_2\\
&\qquad+\|\textstyle\int f(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\rho_*(dy-\cdot)\|_2\\
&\quad\leqq\|\textstyle\int\hat{f}(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\hat{\rho}_*(dy-\cdot)\|_2\\
&\qquad+\|\textstyle\int f(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\hat{\rho}_*(dy-\cdot)\|_2
\end{aligned}$$

(by (4.6))

$$\begin{aligned}
&\quad\leqq\|\textstyle\int\hat{f}(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\rho_*(dy-\cdot)\|_2\\
&\qquad+2\|\textstyle\int f(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\rho_*(dy-\cdot)\|_2\\
&\quad=\|\textstyle\int\hat{f}(y)\,\hat{\rho}_*(dy-\cdot)-\int\hat{f}(y)\,\rho_*(dy-\cdot)\|_2\\
&\qquad+2\|\textstyle\int f(y)\,\hat{\rho}_*(dy-\cdot)-\int f(y)\,\rho_*(dy-\cdot)\|_2,
\end{aligned}$$

which converges to zero by Theorem 2.1 applied to the estimators $\hat{\rho}_*$.

In view of the preceding paragraph, the third term in (4.9) converges to zero almost surely, and so also does the first, which completes the proof. □

Acknowledgement. Several referees of earlier versions of the paper generously provided numerous suggestions for improving the paper; it is a pleasure to thank them.

References

1. Bartlett, M.S.: The spectral analysis of point processes. J. Roy. Statist. Soc. B **25**, 264–296 (1963)
2. Bartlett, M.S.: The spectral analysis of two-dimensional point processes. Biometrika **51**, 299–311 (1964)
3. Bartlett, M.S.: The spectral analysis of line processes. Proc. Fifth Berkeley Sympos. Math. Statist. Probab. **3**, 135–153. University of California (1967)
4. Brillinger, D.R.: The spectral analysis of stationary interval functions. Proc. Sixth Berkeley Sympos. Math. Statist. Probab. **1**, 483–513. University of California (1972)
5. Brillinger, D.R.: Statistical inference for stationary point processes. In: Puri, M.L. (ed.) Stochastic processes and related topics. New York: Academic Press 1975
6. Çinlar, E.: Superposition of point processes. In: Lewis, P.A.W. (ed.) Stochastic point processes. New York: Wiley 1972
7. Cox, D.R., Lewis, P.A.W.: The statistical analysis of series of events. London: Chapman and Hall 1966

8. Daley, D.J.: Spectral properties of weakly stationary point processes. J. Roy. Statist. Soc. B **33**, 406–428 (1971)
9. Itô, K.: Stationary random distributions. Proc. Mem. Sci. Univ. Kyoto A **28**, 209–223 (1955)
10. Jolivet, E.: Central limit theorem and convergence of empirical processes for stationary point processes. In: Bartfai, P., Tomko, J. (eds.) Point processes and queueing problems. Amsterdam: North-Holland 1981
11. Kallenberg, O.: Random measures, 3rd edn. Berlin: Akademie-Verlag and New York: Academic Press 1983
12. Karr, A.F.: Classical limit theorems for measure-valued Markov processes. J. Multivariate Anal. **9**, 234–247 (1979)
13. Karr, A.F.: Inference for thinned point processes, with application to Cox processes. J. Multivariate Anal. **16**, 368–392 (1985)
14. Krickeberg, K.: Moments of point processes. In: Harding, E.F., Kendall, M.G. (eds.) Stochastic geometry. New York: Wiley 1974
15. Krickeberg, K.: Processus ponctuels en statistique. Lect. Notes Math. **929**, 205–313. Berlin Heidelberg New York: Springer 1982
16. Leonov, V.P., Shiryayev, A.N.: On a method of calculation of semi-invariants. Theor. Probab. Appl. **4**, 319–329 (1959)
17. Neveu, J.: Processus ponctuels. Lect. Notes Math. **598**, 249–447. Berlin Heidelberg New York: Springer 1977
18. Nguyen, X.X., Zessin, H.: Ergodic theorems for spatial processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **48**, 133–158 (1979)
19. Rudin, W.: Functional analysis. New York: McGraw-Hill 1973
20. Vere-Jones, D.: An elementary approach to the spectral theory of stationary random measures. In: Harding, E.F., Kendall, M.G. (eds.) Stochastic geometry. New York: Wiley 1974
21. Yoshida, K.: Functional analysis. Berlin Heidelberg New York: Springer 1968

Received March 26, 1984; in revised form April 7, 1986