Z. Wahrscheinlichekitstheorie verw. Gebiete
66, 81-96 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Combined Nonparametric Inference and State Estimation for Mixed Poisson Processes*

Alan F. Karr

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

Summary. Given independent, identically distributed copies of a mixed Poisson process N on a LCCB space E, i.e., a Cox process whose directing measure is of the form αm^*, where $\alpha \geqq 0$ is a random variable with distribution σ and m^* is a measure on E, we construct strongly consistent and asymptotically normal estimators of m^* and the Laplace transform l_σ. Methods are presented for estimating the directing measure of the $(n+1)^{\text{st}}$ process by combining the data for that process with estimates of appropriate quantities, the latter based on the first n processes. The case where different processes are observed over different sets is addressed.

1. Introduction and Problem Formulation

In this paper we analyze methods for effecting nonparametric estimation of defining objects and state estimation using estimated attributes of the process, for mixed Poisson processes, which constitute an important class of Cox processes. The possibly unfamiliar term "state estimation" will be explained below. We assume that the data comprise a sequence of i.i.d. realizations of a basic underlying process, and consider mainly asymptotic properties of estimators of the quantities defining the process and of state estimators constructed using estimated properties of the process.

Our setting is the following. Let E be a locally compact Hausdorff space with Borel σ - algebra $\mathscr{E}$. Let m^* be a fixed, locally finite measure (= Radon measure) on $\mathscr{E}$ and let α be a nonnegative random variable with distribution function σ. Regard $M = \alpha m^*$ as a random measure on $\mathscr{E}$. Then a point process N on E is said to be a *mixed Poisson* process directed by αm^* provided that conditional on α, N is a Poisson point process on E with mean measure αm^*;

* Research supported in part by the Air Force Office of Scientific Research, USAF, grant number AFOSR 82-0029. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes

see Kallenberg (1976). In particular, it follows that for each A, $N(A)$ has the *mixed Poisson distribution* (called by some authors a compound Poisson distribution)

$$(1.1)\qquad P\{N(A)=k\}=\frac{m^*(A)^k}{k!}\int\sigma(dx)\,e^{-xm^*(A)}\,x^k,\qquad k\geqq 0.$$

Mixed Poisson processes are a special case of Cox processes (=doubly stochastic Poisson processes); see Kallenberg (1976) for details.

We assume that m^* and σ are unknown (except that they satisfy certain mild hypotheses below) and that they and other attributes of the individual processes are to be determined from the data. We work with data that are mixed Poisson processes $N_1, N_2, \ldots$ directed by i.i.d. random measures $M_1=\alpha_1 m^*$, $M_2=\alpha_2 m^*, \ldots$, where the α_i are i.i.d. with distribution σ. Thus, each M_i is a deterministic measure m^* common to all processes times a random scalar multiplier α_i.

Our analysis focusses on two main problems. The first is more classical, although our approach is entirely nonparametric: estimation of the measure m^* and the probability measure σ that define the distribution of the N_i. This problem is treated in Sect. 2, where we propose estimators of m^* and of the Laplace transform

$$l_\sigma(t)=\int\sigma(dx)\,e^{-tx},$$

which are shown to be strongly consistent and jointly asymptotically normal.

Our second class of problems deals with *state estimation.* Consider for a moment a single mixed Poisson process N directed by the random measure $M(A)=\alpha m^*(A)$ and suppose that m^* and σ were known. The random variable α is not directly observable but often (see, e.g., the discussion below of potential applications to modeling of cancer) is of paramount interest. Thus one must estimate α, realization-by-realization, based on observation only of N, possibly over only a subset A of E. Regarding α as an unobservable state of nature leads to the term "state estimation". Optimality in the sense of minimum mean-squared error is attained by conditional expectations $E[\alpha\,|\,\mathscr{F}_A^N]$, where $\mathscr{F}_A^N=\sigma(N(B)\colon B\subset A)$ is the σ-algebra representing observation of N over A. It is shown in Karr (1983) that

$$(1.2)\qquad E[\alpha\,|\,\mathscr{F}_A^N]=\frac{\int\sigma(dx)\,e^{-xm^*(A)}\,x^{1+N(A)}}{\int\sigma(dy)\,e^{-ym^*(A)}\,y^{N(A)}};$$

consequently, M itself is estimated by

$$(1.3)\qquad E[M\,|\,\mathscr{F}_A^N]=\frac{\int\sigma(dx)\,e^{-xm^*(A)}\,x^{1+N(A)}}{\int\sigma(dy)\,e^{-ym^*(A)}\,y^{N(A)}}\,m^*,$$

where the conditional expectation is in the sense of Karr (1976).

Obviously (1.2)–(1.3) can be implemented only if m^* and σ are known. Nonetheless the underlying problems of state estimation are of equal importance when m^* and σ are unknown. In Sect. 3 we treat this class of questions in the following formulation: i.i.d. copies of N are observed one-by-one. Suppose that processes $N_1, \ldots, N_n$ have already been observed. To effect state estimation based on partial observation of the process N_{n+1} over the set A (i.e., to approximate $E[\alpha_{n+1} | \mathscr{F}_A^{N_{n+1}}]$ or $E[M_{n+1} | \mathscr{F}_A^{N_{n+1}}]$) we invoke the *principle of separation* long used in engineering. That is, the data $N_1, \ldots, N_n$ are used to form estimates of the functionals of m^* and σ that appear in (1.2)–(1.3), while the partial observations of N_{n+1} are substituted for those of N. This procedure yields a "pseudo-state estimator" that approximates the "true" state estimator. Many of our results are asymptotic comparisons of "true" and "pseudo" state estimators.

Our interest in these mixed Poisson processes and the questions addressed here was stimulated by the paper of Bartoszyński, Brown, McBride, Thompson (1981), especially the proportional hazards model in their Sect. 5. The basic problem they treat is estimation (from i.i.d. realizations) of the intensity function of a nonhomogeneous Poisson process on $\mathbb{R}_+$. The context is metastasis in the growth of malignant tumors: different processes correspond to different patients and events of one patient's process are times of metastases. Their techniques are based on penalized maximum likelihood estimation. In the proportional hazards model (based on Cox (1972)) they permit finitely many covariates $z_1, \ldots, z_d$ that are deterministic and observable, and influence the intensity through a factor $\exp(\sum \beta_i z_i)$, where $\beta \in \mathbb{R}^d$ is unknown. Our model replaces the covariates by multipliers α_n that are random and unobservable, and generated by the unknown probability distribution σ; thus while related the two models apply to differing physical situations. State estimation for α_{n+1} in effect seeks to estimate for that process and each realization, the contribution of the unobservable "covariates".

In reality, of course, not all patients are observed for the same length of time, let alone over the same time interval, so in Sect. 4 we extend some of the results of earlier sections to the case where the process N_i is observed over a *deterministic* set A_i. These results do not include the case of randomly censored data, although extension to a censoring mechanism independent of the processes should be straightforward.

Mixed Poisson processes on $\mathbb{R}_+$ were introduced by Lundberg (1940) in the context of insurance (m^* = Lebesgue measure). When α has a gamma distribution the process is called a Pólya process. The definitive mathematical characterization of these Cox processes is due to Kallenberg (1975): when $m^*(E) = \infty$, N is a mixed Poisson process with parameters α, m^* if and only if N is symmetrically distributed with respect to m^* in the sense that whenever $A_1, \ldots, A_k$ are disjoint (bounded) sets, the random variables $N(A_1), \ldots, N(A_k)$ are interchangeable. For further details, including an analogous characterization when $m^*(E) < \infty$, see Kallenberg (1975, 1976). State estimation techniques for these processes are developed in Karr (1983).

Estimation from i.i.d. samples of the mixing distribution of a single mixed Poisson distribution has been treated by Tucker (1963) and, using maximum likelihood/convexity methods, by Simar (1976). Albrecht (1982) deals with mixed Poisson processes on $\mathbb{R}$, with m^* assumed to be Lebesgue measure. Virtually all aspects of our problem are more general. The basic question of estimating m^* and σ involves a whole family, indexed by $\mathscr{E}$, of mixed Poisson distributions, but with the special structure indicated by (1.1). In Sect. 2 we exploit this structure in order to effectively estimate m^* and σ. For the state estimation problems of Sect. 3, the crucial quantities to be estimated are integrals $K_A(k)$ of the form (3.2) below, which we estimate using the estimator for m^* together with the empirical process associated with $(N_n(A))$; these estimators are related to the estimators obtained by Simar (1976) for a single mixed Poisson distribution. (Simar essentially estimates the mixing distribution μ by estimating integrals of the form $\int \mu(dx)e^{-x}x^k$, by which μ is uniquely determined.)

For the most part, our estimators are devised *ad hoc*, with their reasonableness justified by virtue of strong consistency and asymptotic normality. The generality in which we work, with E a general space and m^* and σ both unknown, seemingly precludes use of maximum likelihood methods.

2. Nonparametric Estimation of m^* and σ

The following assumptions will be in force *throughout* the paper:

(2.1a) $$E \text{ is compact;}$$

(2.1b) $$m^* \text{ is diffuse;}$$

(2.1c) $$\sigma(0)=0;$$

(2.1d) $$\int \sigma(dx)\,x=1.$$

None of these is very restrictive; the role of (2.1d) is to fix the value of an otherwise unidentifiable constant that could be shifted between m^* and the α_i.

For *this section only* we assume for the sake of exposition that $E=[0, T]$ is a compact interval in $\mathbb{R}_+$. Although the arguments below seemingly make heavy use of this assumption, it can be relaxed, in view of (2.1b), by constructing (cf. Kallenberg (1976, Chapter 8)) increasing sets (A_x), $0 \leqq x \leqq 1$, that play the same role as do the intervals $[0, x]$ below.

Let $N_1, N_2, \ldots$ be the mixed Poisson processes on E directed by $M_1=\alpha_1 m^*$, $M_2=\alpha_2 m^*, \ldots$, respectively, where the α_i are i.i.d. with distribution σ; each N_i is observed over all of E. Note that

(2.2) $$E[N(A)]=m^*(A),$$

(this uses (2.1d)) and

(2.3) $$P\{N(A)=0\}=l_\sigma(m^*(A)),$$

both expressions holding for all $A \in \mathscr{E}$.

Below we use distribution function notation where convenient and denote by m^{**} the right-continuous inverse of m^*.

We now introduce estimators for m^* and for l_σ (the latter is considered only on the interval $[0, m^*(T)]$, because of (2.3)). Let

(2.4) $$\hat{m}^*=\frac{1}{n}\sum_{i=1}^{n} N_i$$

and let

$$\hat{l}_\sigma(t) = \tilde{l}(\hat{m}^{**}(t)), \qquad t \leqq \hat{m}^*(T), \tag{2.5}$$

where $\hat{m}^{**}$ is the right-continuous inverse of $\hat{m}^*$ and

$$\tilde{l}(x) = \frac{1}{n} \sum_{i=1}^{n} 1(N_i(x) = 0), \qquad x \leqq T. \tag{2.6}$$

The motivation for (2.4) is evidently (2.2), while (2.5) is motivated by the fact that $\tilde{l}(x)$ estimates $l_\sigma(m^*(x))$, which implies that $\hat{l}_\sigma(t) = \tilde{l} \circ \hat{m}^{**}(t)$ estimates $l_\sigma(t) = l_\sigma(m^*(m^{**}(t)))$. Here and below estimators are distinguished by the caret and dependence of them on n (the sample size) is suppressed.

We now examine properties of these estimators.

(2.7) **Theorem** (Consistency). *Assume that* (2.1) *holds. Then almost surely*

a) $\hat{m}^* \to m^*$ *uniformly on* $[0, T]$;

b) $\hat{l}_\sigma \to l_\sigma$ *uniformly on* $[0, m^*(T)]$.

Proof. a) By Theorem (2.1) of Karr (1979), using the assumption that the N_i are i.i.d. point processes with mean measure m^*, we infer that almost surely $\hat{m}^* \to m^*$ in the sense of vague convergence of Radon measures on $[0, T]$. Since (2.1a) implies that m^* is continuous, the convergence is uniform by Billingsley (1968, p. 21) or Chung (1974, p. 86).

b) First of all, we claim that

$$\tilde{l}(x) \to l_\sigma(m^*(x)) \tag{2.8}$$

uniformly in $x \in [0, T]$, almost surely. Indeed,

$$\tilde{l}(x) = \frac{1}{n} \sum_{i=1}^{n} 1(T_{i,1} > x),$$

where $T_{i,1}$ is the time of the first event in N_i, and by (2.3),

$$l_\sigma(m^*(x)) = P\{T_1 > x\},$$

where T_1 is the first event in a generic copy N. Therefore, (2.8) follows from the Glivenko/Cantelli theorem; cf. Chung (1974).

Since $l_\sigma \circ m^*$ is uniformly continuous

$$\|\hat{l}_\sigma - l_\sigma\|_\infty \leqq \|\tilde{l} - l_\sigma \circ m^*\|_\infty + \text{constant} \times \|\hat{m}^{**} - m^{**}\|_\infty,$$

and b) now follows from (2.8) and a). $\square$

For simplicity we assume in the following result that $m^*(T) = 1$; the general case is obtained by a scaling argument entailing properly placed multiplications by proper powers of $m^*(T)$.

(2.9) **Theorem** (Asymptotic normality). *Assume that* (2.1) *holds, that* $m^*(T) = 1$ *and that* m^* *is strictly increasing and differentiable on* $[0, T]$. *Then for every* $\varepsilon > 0$

$$(2.10)\qquad \begin{bmatrix} n^{\frac{1}{2}}(\hat{m}^* - m^*) \\ n^{\frac{1}{2}}(\hat{l}_\sigma - l_\sigma) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} U \\ \tilde{X} - \tilde{Z} \end{bmatrix}$$

as processes on $[0, T] \times [0, 1-\varepsilon]$, *where* $\xrightarrow{d}$ *denotes convergence in distribution and where*

i) $U \stackrel{d}{=} W_U^0(m^*) + Vm^*$ *on* $[0, T]$ *with* W_U^0 *a Brownian bridge and* V *a random variable independent of* W_U^0 *with distribution* $N(0, \tau^2)$, *where*

$$(2.11)\qquad \tau^2 = \operatorname{Var}(N_1(T));$$

ii) $\tilde{Z}(y) = (l'_\sigma(y)/(m^{**})'(y))\, Z(y)$ *on* $[0, 1-\varepsilon]$, *where* $Z(y) = -W_U^0(m^{**}(y)) - V(m^{**})'(y)\, y$, *with* W_U^0, V *as in* i) *above;*

iii) $\tilde{X} \stackrel{d}{=} W_X^0(1 - l_\sigma)$ *on* $[0, 1-\varepsilon]$, *where* W_X^0 *is a Brownian bridge and the process*

$$(2.12)\qquad X = \tilde{X}(m^*)$$

on $[0, m^*(T)]$ *satisfies the following covariance relations:*

$$(2.13)\qquad \begin{aligned} \operatorname{Cov}(U(x), X(y)) &= -m^*(x)\, l_\sigma(m^*(y)) && \text{if } x \leqq y \\ &= (m^*(x) - m^*(y))\, K_{[0,y]}(1) \\ &\quad - m^*(x)\, l_\sigma(m^*(y)) && \text{if } x > y. \end{aligned}$$

Proof. 1) To begin, consider the sequence of processes

$$\begin{bmatrix} U_n \\ X_n \\ Z_n \end{bmatrix} = \begin{bmatrix} n^{\frac{1}{2}}(\hat{m}^* - m^*) & \text{on } [0, T] \\ n^{\frac{1}{2}}(\hat{l} - l_\sigma \circ m^*) & \text{on } [0, T] \\ n^{\frac{1}{2}}(\hat{m}^{**} - m^{**}) & \text{on } [0, 1-\varepsilon] \end{bmatrix}.$$

By the continuous mapping theorem and an argument based on a Taylor expansion, it suffices to show that

$$(2.14)\qquad (U_n, X_n, Z_n) \xrightarrow{d} (U, X, Z)$$

and then to calculate relevant covariance relationships. But it is also apparent that each X_n is a functional of U_n, so we need only establish asymptotic normality of (U_n, Z_n). The composition of $\hat{l}$ and $\hat{m}^{**}$ in the definition of $\hat{l}_\sigma$ is dealt with by Theorem (2.7) and Billingsley (1968, pp. 144–145).

2) Restricting attention to (U_n) alone for the moment, we have

$$(2.15)\qquad U_n \xrightarrow{d} W_U^0(m^*(\cdot)) + Vm^*(\cdot),$$

where W_U^0 is a Brownian bridge and V is a normally distributed random variable independent of W_U^0 with mean 0 and variance τ^2 given by (2.11) above. To establish (2.15) we use the representation (see Kallenberg (1976) or Matthes/Kerstan/Mecke (1978))

$$N_i = \sum_{k=1}^{N_i(T)} \varepsilon_{X_{ik}},$$

where ε_x denotes the point mass at x, and the X_{ik} are i.i.d. with distribution m^* and independent of the $N_i(T)$. Putting $S_n = \sum_1^n N_i(T)$ we can write

$$U_n = n^{-\frac{1}{2}} \left(\sum_{i=1}^{S_n} \varepsilon_{X_i} - nm^* \right) \tag{2.16}$$

with the X_i independent of (S_n) and themselves i.i.d. with distribution m^*. Thus,

$$U_n = n^{-\frac{1}{2}} \left(\sum_1^{S_n} \varepsilon_{X_i} - S_n m^* \right) + n^{-\frac{1}{2}} (S_n - n) m^*$$

$$\sim S_n^{-\frac{1}{2}} \left(\sum_1^{S_n} \varepsilon_{X_i} - S_n m^* \right) + n^{-\frac{1}{2}} (S_n - n) m^*$$

(since $S_n/n \to m^*(T) = 1$ a.s.)

$$\xrightarrow{d} W_U^0(m^*) + Vm^*$$

by independence of (S_n) and (X_i), the continuous mapping theorem, standard theory of empirical processes and the ordinary central limit theorem.

3) Now consider (Z_n) alone. From the representation (2.16) we have

$$Z_n = n^{\frac{1}{2}} \left[\sum_{k=0}^{S_n - 1} (X_{S_n, (k+1)} - X_{S_n, (k)})\, \varepsilon_{k/n} - m^{**} \right]$$

(where the $X_{S_n, (k)}$ are the order statistics from $X_1, \dots, X_{S_n}$)

$$\sim S_n^{\frac{1}{2}} \left[\sum_{k=0}^{S_n - 1} (X_{S_n, (k+1)} - X_{S_n, (k)})\, \varepsilon_{k/S_n} - m^{**} \right]$$

$$+ n^{\frac{1}{2}} \left[\sum_0^{S_n - 1} (X_{S_n, (k+1)} - X_{S_n, (k)}) (\varepsilon_{k/n} - \varepsilon_{k/S_n}) \right]$$

$$= I + II.$$

By independence of (S_n) and (X_i) and theory of empirical processes (Shorack (1972), e.g.), for a Brownian bridge W_Z^0,

$$I \xrightarrow{d} W_Z^0(m^{**}) \tag{2.17}$$

on $[0, 1-\varepsilon]$ for every $\varepsilon > 0$. Applying II to a smooth function f gives

$$n^{\frac{1}{2}} \left[\sum_0^{S_n - 1} (X_{S_n, (k+1)} - X_{S_n, (k)}) \left(f\left(\frac{k}{n}\right) - f\left(\frac{k}{S_n}\right) \right) \right] \tag{2.18}$$

$$\sim n^{\frac{1}{2}} \left[\sum_0^{S_n - 1} (X_{S_n, (k+1)} - X_{S_n, (k)}) f'\left(\frac{k}{S_n}\right) \frac{k}{S_n} \frac{S_n - n}{n} \right]$$

$$\xrightarrow{d} V \int f'(y)\, y m^{**}(dy)$$

by Taylor's theorem (since $S_n/n \to 1$ a.s.), independence of (S_n) and (X_i), (2.17) and the ordinary central limit theorem, with V as above. To obtain ii) one uses

(2.18), integration by parts and a straightforward approximation of indicator functions by smooth functions. Finally, validity of the relationship

$$W_Z^0 = -W_U^0$$

is a consequence of Shorack (1972). This completes the proof of asymptotic normality of (U_n, Z_n).

4) That $X = \tilde{X}(m^*)$ has the indicated form follows by arguments in 2) above, since

$$\tilde{l}(x) = \frac{1}{n} \sum_{i=1}^{n} 1(T_{i,1} > x)$$

is the ordinary empirical process associated with the probability distribution $1 - l_\sigma(m^*(x))$. □

3. Estimation of the Directing Measure

Recall that the conditions (2.1) are in force; however, E is otherwise a general LCCB space. Let $N_1, N_2, \ldots$ be i.i.d. copies of the mixed Poisson process N directed by $M = \alpha m^*$, where α has distribution σ. Suppose that $N_1, \ldots, N_n$ have been observed over all of E and N_{n+1} over a subset A. Consider the problem of reconstructing the directing measure M_{n+1}, in the sense of state estimation, from these data. According to (1.3), if m^* and σ were known, the appropriate state estimators would be the conditional expectations

$$E[M_{n+1}(B) \mid \mathscr{F}_A^{N_{n+1}}] = \frac{K_A(N_{n+1}(A)+1)}{K_A(N_{n+1}(A))}\, m^*(B), \tag{3.1}$$

where

$$K_A(k) = \int \sigma(dx)\, e^{-x m^*(A)} x^k = \frac{k!}{m^*(A)^k}\, P\{N(A) = k\}. \tag{3.2}$$

However, if m^* and σ are unknown, it is necessary to replace m^* and K_A by estimates thereof; our strategy is to construct the estimates $\hat{m}^*$, $\hat{K}_A$ from $N_1, \ldots, N_n$, then to replace m^*, K_A in (3.1) by them.

We begin the section by developing properties of the estimators

$$\hat{m}^* = \frac{1}{n} \sum_{i=1}^{n} N_i \tag{3.3}$$

and

$$\hat{K}_A(k) = \frac{1}{n} \sum_{i=1}^{n} \frac{k!}{\hat{m}^*(A)^k}\, 1(N_i(A) = k). \tag{3.4}$$

We proceed to study the difference between the "true" state estimator $E[M_{n+1} \mid \mathscr{F}_A^{N_{n+1}}]$ of (3.1) and the "pseudo-state estimator"

$$\hat{E}[M_{n+1} \mid \mathscr{F}_A^{N_{n+1}}] = \frac{\hat{K}_A(N_{n+1}(A)+1)}{\hat{K}_A(N_{n+1}(A))}\, \hat{m}^*, \tag{3.5}$$

where both sides of (3.5) are viewed as random measures on E; see Karr (1976). The estimator $\hat{m}^*$ appeared in Sect. 2, where motivation for it was given; the choice of $\hat{K}_A$ is motivated by (3.2).

The first results of the section establish consistency and asymptotic normality of the basic estimators.

(3.6) **Theorem** (Consistency). *For each $A \in \mathscr{E}$, almost surely*

$$\frac{\hat{K}_A(\cdot+1)}{\hat{K}_A(\cdot)} \to \frac{K_A(\cdot+1)}{K_A(\cdot)} \quad \text{in } \mathbb{R}_+^\infty.$$

We omit the straightforward proof.

(3.7) **Theorem** (Asymptotic normality). *Assume that $m^*(A)>0$ and that $\operatorname{Var}(\alpha)<\infty$. Then*

$$\begin{bmatrix} n^{\frac{1}{2}}(\hat{m}^*-m^*) \\ n^{\frac{1}{2}}(\hat{K}_A-K_A) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mu \\ Z \end{bmatrix}, \tag{3.8}$$

where μ is a Gaussian random measure (parameterized by the set $C(E)$ of bounded continuous functions on E) and Z is a Gaussian sequence, whose covariance function is given for $g, h \in C(E)$ and $k, j \in N$ by

$$\Gamma(g,h)=m^*(gh)+\operatorname{Var}(\alpha)\, m^*(g)\, m^*(h), \tag{3.9a}$$

$$\Gamma(g,k)=\frac{k!\,P\{N(A)=k\}}{m^*(A)^k}\left[k-m^*(g)-\frac{m^*(g1_A)+\operatorname{Var}(\alpha)\, m^*(A)\, m^*(g)}{m^*(A)}\right], \tag{3.9b}$$

$$\begin{aligned} \Gamma(k,j)=&P\{N(A)=k\}\,P\{N(A)=j\} \\ &\times\left[-\frac{k!\,j!}{m^*(A)^{k+j}}-\frac{k!\,j!\,j(k-m^*(A))}{m^*(A)^{k+2j}}-\frac{k!\,j!\,k(j-m^*(A))}{m^*(A)^{2k+j}}\right. \\ &\left.+\left(\frac{k!\,k}{m^*(A)^{2k}}\right)\left(\frac{j!\,j}{m^*(A)^{2j}}\right)[m^*(A)+\operatorname{Var}(\alpha)\, m^*(A)^2]\right] \\ &+P\{N(A)=k\}\left(\frac{k!}{m^*(A)^k}\right)^2 1(k=j), \end{aligned} \tag{3.9c}$$

where $m^*(f)=\int f\, dm^*$ and 1_A is the indicator function of A.

Proof. By the Cramér/Wold device and Theorem 4.2 of Kallenberg (1976) it suffices to show that for $L \geqq 1$, $c_1, \ldots, c_{L+1} \in \mathbb{R}$ and $g \in C(E)$ the linear combination

$$\left[\sum_{l=1}^{L} c_l\, n^{\frac{1}{2}}(\hat{K}_A(l)-K_A(l))\right]+c_{L+1}\, n^{\frac{1}{2}}(\hat{m}^*(g)-m^*(g)) \tag{3.10}$$

has a limiting normal distribution, and then to calculate the covariance function. Concerning asymptotic normality of (3.10), we observe first that for each k

(3.11) $$n^{\frac{1}{2}}(\hat{K}_A(k)-K_A(k))$$
$$=\frac{k!}{m^*(A)^k}\, n^{-\frac{1}{2}} \sum_{i=1}^{n} [1(N_i(A)=k)-P\{N(A)=k\}]$$
$$+k!\left(\frac{1}{n}\sum_{i=1}^{n} 1(N_i(A)=k)\right)\frac{1}{\hat{m}^*(A)^k\, m^*(A)^k}\, n^{\frac{1}{2}}(m^*(A)^k-\hat{m}^*(A)^k).$$

By virtue of (3.11), Theorem (3.6), a standard Taylor expansion and Slutsky's theorem (cf. Billingsley (1968)), asymptotic normality of (3.10) follows from that of the random vectors

(3.12) $$\begin{bmatrix} n^{-\frac{1}{2}}\sum_{i=1}^{n}[1(N_i(A)=1)-P\{N(A)=1\}] \\ \vdots \\ n^{-\frac{1}{2}}\sum_{i=1}^{n}[1(N_i(A)=L)-P\{N(A)=L\}] \\ n^{-\frac{1}{2}}\sum_{i=1}^{n}[N_i(A)-m^*(A)] \\ n^{-\frac{1}{2}}\sum_{i=1}^{n}[N_i(g)-m^*(g)] \end{bmatrix};$$

but asymptotic normality of (3.12) is evident, since the N_i are i.i.d. □

Theorem (3.7) is not directly relevant to the problem at hand, but the following consequence of it is.

(3.13) **Theorem** (Asymptotic normality). *Under the assumptions of Theorem* (3.7),

(3.14) $$\begin{bmatrix} n^{\frac{1}{2}}(\hat{m}^*-m) \\ n^{\frac{1}{2}}\left(\frac{\hat{K}_A(\cdot+1)}{\hat{K}_A(\cdot)}-\frac{K_A(\cdot+1)}{K_A(\cdot)}\right)\end{bmatrix} \xrightarrow{d} \begin{bmatrix}\mu \\ Z^*\end{bmatrix},$$

where (μ, Z^*) *is a Gaussian process with covariance function* Σ *given by*

(3.15a) $\Sigma(g, h)=\Gamma(g, h)$,

(3.15b) $$\Sigma(g, k)=\frac{1}{K_A(k)}\,\Gamma(g, k+1)-\frac{K_A(k+1)}{K_A(k)^2}\,\Gamma(g, k),$$

and

(3.15c) $$\Sigma(k, j)=\frac{1}{K_A(k)\,K_A(j)}\,\Gamma(k+1, j+1)-\frac{K_A(k+1)}{K_A(k)^2\,K_A(j)}\,\Gamma(k, j+1)$$
$$-\frac{K_A(j+1)}{K_A(k)\,K_A(j)^2}\,\Gamma(k+1, j)+\frac{K_A(k+1)\,K_A(j+1)}{K_A(k)^2\,K_A(j)^2}\,\Gamma(k, j),$$

where Γ *is given by* (3.9).

We now take up the question of asymptotic behavior of the difference

$$\hat{E}[M_{n+1}\,|\,\mathscr{F}_A^{N_{n+1}}]-E[M_{n+1}\,|\,\mathscr{F}_A^{N_{n+1}}]$$

between the pseudo-state estimator given by (3.5) and the "true" state estimator given by (3.1). Recall that the former is based on estimates of m^* and the K_A obtained from observation of the previous processes $N_1, \ldots, N_n$, whereas the latter is applicable when m^* and σ are known. The following result is the most important in this section.

(3.16) **Theorem.** *Under the assumptions of Theorem* (3.7),

$$n^{\frac{1}{2}}(\hat{E}[M_{n+1} \mid \mathscr{F}_A^{N_{n+1}}] - E[M_{n+1} \mid \mathscr{F}_A^{N_{n+1}}]) \xrightarrow{d} \eta \tag{3.17}$$

as random signed measures on E, where η is a mixture, with mixing distribution (1.1), *of the sequence (ρ_k) of (centered) Gaussian random measures on E having covariance function Λ given by*

$$\begin{aligned}\Lambda(g_k, g_j) &= \mathrm{Cov}(\rho_k(g_k), \rho_j(g_j)) \\ &= m^*(g_k)\, m^*(g_j)\, \Sigma(k,j) + m^*(g_k)\, \frac{K_A(j+1)}{K_A(j)}\, \Sigma(g_k, j) \\ &\quad + m^*(g_j)\, \frac{K_A(k+1)}{K_A(k)}\, \Sigma(g_j, k) + \frac{K_A(k+1)}{K_A(k)}\, \frac{K_A(j+1)}{K_A(j)}\, \Sigma(g_k, g_j),\end{aligned} \tag{3.18}$$

where Σ is given by (3.15).

Proof. Once again appealing to Slutsky's theorem, we find that the processes

$$\rho_n = n^{\frac{1}{2}} \left(\frac{\hat{K}_A(\cdot+1)}{\hat{K}_A(\cdot)}\, \hat{m}^*(\cdot) - \frac{K_A(\cdot+1)}{K_A(\cdot)}\, m^*(\cdot) \right)$$

asymptotically satisfy

$$\begin{aligned}\rho_n(g, k) &= m^*(g)\, n^{\frac{1}{2}} \left(\frac{\hat{K}_A(k+1)}{\hat{K}_A(k)} - \frac{K_A(k+1)}{K_A(k)} \right) \\ &\quad + \frac{K_A(k+1)}{K_A(k)}\, n^{\frac{1}{2}}(\hat{m}^*(g) - m^*(g)).\end{aligned}$$

Together with Theorem (3.13), this implies that $\rho_n \xrightarrow{d} \rho$, where ρ is Gaussian with covariance function Λ given by (3.18).

For each n, $N_{n+1}(A)$ is independent of $\{N_1, \ldots, N_n\}$ and hence of $\{\hat{m}^*, \hat{K}_A\}$; therefore (here $N(A)$ has distribution (1.1))

$$(\rho_n, N_{n+1}(A)) \xrightarrow{d} (\rho, N(A)), \tag{3.19}$$

where ρ and $N(A)$ are independent. The mapping $H((m_n), j) \to m_j$, where the m_n are Radon measures on E, is trivially continuous, so (3.17) follows from (3.19) by the continuous mapping theorem. □

Theorem (3.16) dealt with estimating the directing measure M_{n+1}. However, in some applications (e.g., the cancer model discussed in Sect. 1) estimation of α_{n+1} may be the principal interest. (Possibly m^* is even known.) Analogous but simpler arguments yield the following result concerning the difference between the pseudo-state estimator

$$\hat{E}[\alpha_{n+1} \mid \mathscr{F}_A^{N_{n+1}}] = \frac{\hat{K}_A(N_{n+1}(A)+1)}{\hat{K}_A(N_{n+1}(A))}$$

and the true state estimator

$$E[\alpha_{n+1} \mid \mathscr{F}_A^{N_{n+1}}] = \frac{K_A(N_{n+1}(A)+1)}{K_A(N_{n+1}(A))}.$$

(3.20) **Theorem.** *Under the assumptions of Theorem* (3.7),

(3.21) $$n^{\frac{1}{2}}(\hat{E}[\alpha_{n+1} \mid \mathscr{F}_A^{N_{n+1}}] - E[\alpha_{n+1} \mid \mathscr{F}_A^{N_{n+1}}]) \xrightarrow{d} Y,$$

where Y has the mixed normal distribution obtained by mixing according to the distribution (1.1) *the Gaussian sequence Z with covariance function $\Sigma(k,j)$ given by* (3.15c).

To conclude the section we observe that in some sense the rates of convergence in Theorems (3.16) and (3.20) are best possible since they match the rates of convergence for the estimators themselves.

4. Different Processes Observed over Different Sets

For many situations the assumption in Sect. 3 that each process be observed over the same set A is excessively restrictive; the fact that the distribution of $N(A)$ depends on A only through $m^*(A)$ indicates that the assumption can be relaxed. In this section we deal with behavior of the difference

$$\hat{E}[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}] - E[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}],$$

where N_i is observed over the (deterministic) set A_i. In order to obtain results we will need to assume that there is a set A such that $m^*(A_i) \to m^*(A)$ in some sense (we consider several). The difference above then takes the form

$$\frac{\hat{K}(N_{n+1}(A_{n+1})+1)}{\hat{K}(N_{n+1}(A_{n+1}))} - \frac{K_{A_{n+1}}(N_{n+1}(A_{n+1})+1)}{K_{A_{n+1}}(N_{n+1}(A_{n+1}))}$$

for suitable estimators $\hat{K}$ introduced below.

We assume that (2.1) is satisfied, that N_i is observed over the set A_i and that A is a fixed set. The following conditions will be used below:

(4.1) $$n^{-1} \sum_{i=1}^{n} |m^*(A_i) - m^*(A)| \to 0,$$

(4.2) $$n^{-\frac{1}{2}} \sum_{i=1}^{n} |m^*(A_i) - m^*(A)| \to 0,$$

(4.3) $$\sup_i \frac{m^*(A_i)}{m^*(A)} < \infty.$$

Evidently (4.1) and (4.2) are forms of convergence of $m^*(A_i)$ to $m^*(A)$. Note, however, that they imply nothing about convergence of A_i to A, so that the A_i need resemble A only in terms of the measure m^*.

As replacements for the estimators $\hat{K}_A$ defined by (3.4) we propose

$$\hat{K}(k) = \frac{k!}{\left(\frac{1}{n}\sum_{j=1}^{n} N_j(A_j)\right)^k} \frac{1}{n}\sum_{i=1}^{n} 1(N_i(A_i)=k). \tag{4.4}$$

These estimators have the following properties (recall that the A_i and A are fixed; the latter is suppressed).

(4.5) **Theorem** (Consistency). *Assume that* (4.1) *and* (4.3) *are satisfied and that*

$$\int \sigma(dx)\, x^k < \infty \tag{4.6}$$

for all $k \in \mathbb{N}$. *Then*

$$\frac{\hat{K}(\cdot+1)}{\hat{K}(\cdot)} \to \frac{K_A(\cdot+1)}{K_A(\cdot)} \tag{4.7}$$

in $\mathbb{R}_+^\infty$ *almost surely.*

Proof. 1) Consider first the (uncomputable) estimators

$$\tilde{K}(k) = \frac{k!}{m^*(A)^k} \frac{1}{n}\sum_{i=1}^{n} 1(N_i(A_i)=k),$$

for which we show that

$$\tilde{K} \to K_A \tag{4.8}$$

almost surely. To do so, we begin by showing that

$$\lim_{n\to\infty} E[\tilde{K}(k)] = K_A(k) \tag{4.9}$$

for each k. Indeed,

$$|E[\tilde{K}(k)] - K_A(k)| \leqq c_k\left(\int \sigma(dx)\, x^{k+1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} |m^*(A_i) - m^*(A)|\right)$$
$$+ K_A(k)\left(\frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{m^*(A_i)}{m^*(A)}\right)^k - 1\right|\right)$$

[where by (4.3) c_k does not depend on n]

$$\leqq \text{constant} \times \frac{1}{n}\sum_{i=1}^{n} |m^*(A_i) - m^*(A)|$$

by (4.3) and (4.6); the last expression converges to zero by (4.1).

The random variables

$$Y_i = \frac{k!}{m^*(A)^k} 1(N_i(A_i)=k)$$

are independent with uniformly bounded variances and therefore by Theorem 5.4.1 of Chung (1974)

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - E[Y_i]) \to 0 \tag{4.10}$$

almost surely; (4.8) follows at once from (4.9) and (4.10).

2) By an analogous argument

$$\frac{1}{n}\sum_{i=1}^{n} N_i(A_i) \to m^*(A) \tag{4.11}$$

almost surely.

The theorem now is a consequence of (4.8) and (4.11). □

(4.12) **Theorem** (Asymptotic normality). *Assume that* (4.2), (4.3) *and* (4.6) *are satisfied. Then*

$$n^{\frac{1}{2}}\left(\frac{\hat{K}(\cdot+1)}{\hat{K}(\cdot)} - \frac{K_A(\cdot+1)}{K_A(\cdot)}\right) \xrightarrow{d} Z, \tag{4.13}$$

where Z is a Gaussian sequence with covariance function Σ given by (3.15c) *above* (*with Γ there given by* (3.9c)).

Proof. 1) Since (4.2) implies (4.1) the convergence (4.7) obtains; together with Slutsky's theorem it implies that for fixed k we have (asymptotically)

$$\begin{aligned} n^{\frac{1}{2}}[\hat{K}(k) - K_A(k)] &= \frac{k!}{m^*(A)^k}\, n^{-\frac{1}{2}} \sum_{i=1}^{n} [1(N_i(A_i)=k) - P\{N_i(A_i)=k\}] \\ &\quad - \frac{k!\,kP\{N(A)=k\}}{m^*(A)^{2k}}\, n^{-\frac{1}{2}} \sum_{i=1}^{n} [N_i(A_i) - m^*(A_i)]. \end{aligned}$$

Together with the Cramér/Wold device this computation reduces asymptotic normality of $n^{\frac{1}{2}}(\hat{K} - K_A)$ to that of standardized sums of the random variables

$$Y_i = c_0 N_i(A_i) + \sum_{l=1}^{L} c_l 1\{N_i(A_i) = l\},$$

where $c_0, \ldots, c_L \in \mathbb{R}$.

2) The random variables $Y_i' = Y_i - E[Y_i]$ are independent with mean 0 and therefore $n^{-\frac{1}{2}} \sum_{1}^{n} Y_i'$ has normal limit distribution $N(0, \tau^2)$ provided that

$$\tau_n^2 = \sum_{i=1}^{n} \mathrm{Var}(Y_i') \to \infty, \tag{4.14}$$

that

$$\tau_n^2/n \to \tau^2, \tag{4.15}$$

and that the Lindeberg condition be satisfied (see Billingsley (1968, Theorem 7.2)). Since

$$|Y_i'| \leqq \text{constant} + |c_0| \, |N_i(A_i) - m^*(A_i)|$$

and since the random variables $N_i(A_i) - m^*(A_i)$ have moments of all orders by (4.6) that are uniformly bounded in i by (4.3), it is clear that Lindeberg's condition is satisfied. If $\tau^2 > 0$ in (4.15), then evidently (4.14) holds, so we restrict attention to (4.15). Using (4.2) and computations analogous to those appearing in the proof of Theorem (4.5) it is straightforward to verify that asymptotic variances are precisely those in the case when $A_i = A$ for all i, which correspond to the covariance function Γ of (3.9c).

The remainder of the proof is routine. □

We are now able to examine the difference between the true state estimator

$$E[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}] = \frac{K_{A_{n+1}}(N_{n+1}(A_{n+1}) + 1)}{K_{A_{n+1}}(N_{n+1}(A_{n+1}))}$$

and the pseudo-state estimator

$$\hat{E}[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}] = \frac{\hat{K}(N_{n+1}(A_{n+1}) + 1)}{\hat{K}(N_{n+1}(A_{n+1}))}.$$

However, we must require that $m^*(A_n) \to m^*(A)$ at a rate faster than $n^{-\frac{1}{2}}$.

(4.16) **Theorem.** *Assume that* (4.6) *holds and that*

(4.17) $$n^{\frac{1}{2}} |m^*(A_n) - m^*(A)| \to 0.$$

Then

(4.18) $$n^{\frac{1}{2}}(\hat{E}[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}] - E[\alpha_{n+1} \mid \mathscr{F}_{A_{n+1}}^{N_{n+1}}]) \to Y,$$

where Y *has the same mixed normal distribution as does the limit in* (3.22).

Proof. Note that (4.17) implies both (4.2) and (4.3), so that (4.13) holds.

1) Consider first the processes

$$n^{\frac{1}{2}} \left(\frac{\hat{K}(\cdot+1)}{\hat{K}(\cdot)} - \frac{K_{A_{n+1}}(\cdot+1)}{K_{A_{n+1}}(\cdot)} \right) = n^{\frac{1}{2}} \left(\frac{\hat{K}(\cdot+1)}{\hat{K}(\cdot)} - \frac{K_A(\cdot+1)}{K_A(\cdot)} \right) + n^{\frac{1}{2}} \left(\frac{K_A(\cdot+1)}{K_A(\cdot)} - \frac{K_{A_{n+1}}(\cdot+1)}{K_{A_{n+1}}(\cdot)} \right).$$

In view of (4.13) and some calculations, if

(4.19) $$n^{\frac{1}{2}} |K_{A_n}(k) - K_A(k)| \to 0$$

for every k, then

(4.20) $$n^{\frac{1}{2}} \left(\frac{\hat{K}(\cdot+1)}{\hat{K}(\cdot)} - \frac{K_{A_{n+1}}(\cdot+1)}{K_{A_{n+1}}(\cdot)} \right) \xrightarrow{d} Z,$$

where Z is as in Theorem (4.12). However,

$$n^{\frac{1}{2}}|K_{A_n}(k)-K_A(k)|\leqq(\int\sigma(dx)\,x^{k+1})\,n^{\frac{1}{2}}|m^*(A_n)-m^*(A)|$$

and hence (4.19) holds by (4.6) and (4.17).

2) To complete the proof we need only show that

$$N_{n+1}(A_{n+1})\xrightarrow{d} N(A),$$

where $N(A)$ has distribution (1.1), but this follows at once from (4.17). □

Remark. The principal difficulty with conditions (4.1), (4.2) and (4.17) is that when m^* is unknown there may be no way to verify in advance whether they are satisfied. Of course, sufficiently strong assumptions on the A_i and A entail (4.1) or (4.2); for example, (4.1) follows from

$$\frac{1}{n}\sum_{i=1}^{n} 1_{A_i \Delta A}\to 0$$

pointwise on E, and similarly for (4.2). However, there is no corresponding sufficient condition for (4.17). Presumably one would at this point require partial knowledge of m^*; e.g., if $m^*\ll m_0$ for some known measure m_0 and (4.17) holds for m_0, then it holds also for m^*.

Acknowledgement. R.J. Serfling generously contributed numerous comments and suggestions.

References

Albrecht, P.: On some statistical methods connected with the mixed Poisson process. Scand. Actuar. J. 1–14 (1982)

Bartoszyński, R., Brown, B.W., McBride, C., Thompson, J.R.: Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process. Ann. Statist. **9**, 1050–1060 (1981)

Billingsley, P.: Convergence of Probability Measures. New York: Wiley 1968

Chung, K.L.: A Course in Probability Theory, 2nd ed. New York: Academic Press 1974

Cox, D.R.: Regression models and life tables. J. Roy. Statist. Soc. Ser. B **34**, 187–220 (1972)

Kallenberg, O.: On symmetrically distributed random measures. Trans. Amer. Math. Soc. **202**, 105–121 (1975)

Kallenberg, O.: Random Measures. New York: Academic Press 1976

Karr, A.F.: A conditional expectation for random measures. Technical report, The Johns Hopkins University (1976)

Karr, A.F.: Classical limit theorems for measure-valued Markov processes. J. Multivariate Anal. **9**, 234–247 (1979)

Karr, A.F.: State estimation for Cox processes on general spaces. Stochastic Processes Appl. **14**, 209–232 (1983)

Lundberg, O.: On Random Processes and their Application to Sickness and Accident Statistics. Uppsala: Almqvist and Wiksells 1940

Matthes, K., Kerstan, J., Mecke, J.: Infinitely Divisible Point Processes. New York: Wiley 1978

Shorack, G.R.: Functions of order statistics. Ann. Math. Statist. **43**, 412–427 (1972)

Simar, L.: Maximum likelihood estimation of a compound Poisson process. Ann. Statist. **4**, 1200–1209 (1976)

Tucker, H.G.: An estimate of the compounding distribution of a compound Poisson distribution. Theor. Probability Appl. **8**, 195–200 (1963)

Received September 2, 1982; in revised form August 15, 1983