MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

STATISTICAL PROBLEMS ON POINT PROCESSES

KLAUS KRICKEBERG

U.E.R. de Mathématiques. Logique Formelle et Informatique, Université René Descartes, Paris, France

The purpose of these lectures is to report on some typical ideas and methods employed in the statistical analysis of point processes. The emphasis is on methods which work in general spaces. Thus we exclude topics based on martingale theory and tied to the structure of the real line; for this the reader should consult papers [0]–[4], [13]. Nor do we treat the extensive theory of the estimation of the spectrum, of the moments, etc., of stationary point processes: papers [5]–[7] give an excellent account of this area. Also, for lack of time, we cannot talk about applications of Palm distributions to statistical inference. Instead we try to illustrate certain basic ideas by describing in some detail two areas: inference in the case of a family of Poisson processes (chapter 2), and filtering problems of Cox (doubly stochastic Poisson) processes (chapter 3). For further reading the references [8], [9], [12], [26], [27] are recommendet.

The first chapter introduces or recalls the concepts and notations used. It turns out that the functional analytic set up simplifies many reasonings and their writing, and makes many things appear in a clearer light. It is, of course, especially appropriate for lectures at the Banach Centre.

Proofs are sometimes sketched and in a few cases given in detail; for the rest there exist sufficiently complete references.

1. Basic concepts

Everything will take place in a basic space X which is assumed to be locally compact with a countable base. We will work with the following spaces of real-valued functions on X or classes of subsets of X or Radon measures in X where "measurable" always means "Borel measurable", that is, Baire:

$\mathscr{C}$ all bounded continuous functions;
$\mathscr{K}$ all bounded continuous functions with compact carrier;
$\mathscr{H}$ all bounded measurable functions;

$\mathscr{H}_0$ all bounded measurable functions with a compact carrier;
$\mathscr{B}$ all Borel sets;
$\mathscr{B}_0$ all bounded, that is, relatively compact, Borel sets;
$\mathscr{M}_+$ all positive measures;
$\mathscr{M}^{\cdot\cdot}$ all positive point measures, that is, measures of the form

$$\mu = \sum_j \varepsilon_{x_j}, \tag{1.1}$$

where the $x_j \in X$ need not be distinct and ε_x is the unit mass located in x. Then the carrier of μ is the locally finite set

$$\operatorname{carr}\mu = \{x_1, x_2, \ldots\};$$

$\mathscr{M}^e$ all finite point measures

$$\mu = \sum_{j=1}^{n} \varepsilon_{x_j}; \tag{1.2}$$

$\mathscr{M}^{\cdot}$ all counting measures, also called *simple point measures*, that is, measures of the form (1.1) with distinct x_j. In this case $\mu(A) = \# (A \cap \operatorname{carr}\mu)$, and μ may be used as a convenient representation of its carrier. Any countable locally finite set is the carrier of a simple point measure;
$\mathscr{M}^0$ all diffuse positive measures, that is, measures in $\mathscr{M}_+$ without atoms;
$\mathscr{P}$ all probability measures.

Some of these entities will also appear with X replaced by some other space, Y say, which need not be locally compact; we will then write for example $\mathscr{C}(Y)$.

If $f \in \mathscr{H}_0$, the integral

$$\mu(f) = \int_X f(x)\mu(dx) \tag{1.3}$$

is defined for every $\mu \in \mathscr{M}_+$, and we denote this as a function of μ by ζ_f; thus $\zeta_f(\mu) = \mu(f)$. We write ζ_A for ζ_{1_A}; here 1_A is the indicator function of $A \subseteq X$. For the point measure (1.1) we have

$$\mu(f) = \sum_j f(x_j).$$

Note that $f \mapsto \zeta_f$ is linear.

We endow $\mathscr{M}_+$ as usual with the vague topology which is the coarsest topology which makes all functions ζ_f with $f \in \mathscr{K}$ continuous. The sigma-algebra $\mathscr{B}(\mathscr{M}_+)$ of the Borel sets in $\mathscr{M}_+$ for this topology can also be characterized without direct reference to a topology in $\mathscr{M}_+$: it is the sigma-algebra generated by all ζ_f with $f \in \mathscr{H}_0$ as well as that generated by all ζ_f with $f \in \mathscr{K}$, but also that generated by all ζ_A with $A \in \mathscr{B}_0$ (see [14]).

A random measure in X is a probability measure, P say, on $\mathscr{B}(\mathscr{M}_+)$, thus $P \in \mathscr{P}(\mathscr{M}_+)$. Intuitively speaking, we select a measure μ in X at random, following the law P. For any $f \in \mathscr{H}_0$ or $A \in \mathscr{B}_0$, the measurable functions ζ_f or ζ_A can then be considered as random variables on the probability space $(\mathscr{M}_+, \mathscr{B}(\mathscr{M}_+), P)$: they represent the integral of f, or the measure of A, as a function of chance.

It will be convenient to take the functional analytic point of view also when describing P. We may define P as a real-valued functional on $\mathscr{C}(\mathscr{M}_+)$ which is linear, positive, normed, that is $P(1) = 1$, and sigma-continuous: $\varphi_n \in \mathscr{C}(\mathscr{M}_+)$ and $\varphi_n \searrow 0$ imply $P(\varphi_n) \searrow 0$. We can then extend P to $\mathscr{H}(\mathscr{M}_+)$ uniquely so as to preserve these properties, and by Ulam's theorem P will be automatically tight, that is, continuous in the loose topology: if $\varphi_n \in \mathscr{H}(\mathscr{M}_+)$ is uniformly bounded and converges to 0 uniformly on every compact set in $\mathscr{M}_+$, then $P(\varphi_n) \to 0$ (see [24], p. 29).

The functions ζ_f with $f \in \mathscr{H}_0$ are continuous but in general not bounded. If all of them are P-integrable, P is said to be of the *first order*, and the measure in X defined by

$$\nu_P(f) = P(\zeta_f) = \int_{\mathscr{M}_+} \mu(f)P(d\mu), \quad \text{or} \quad \nu_P = \int_{\mathscr{M}_+} \mu P(d\mu) \text{ for short},$$

is called the *first order moment measure*, or *intensity measure*, of P. Thus, for example, $\nu_P(A)$ is just the average measure of A. Analogously, the kth order moment measure of P is the measure in X^k given by

$$\nu_P^{(k)}(f_1 \otimes \ldots \otimes f_k) = P(\zeta_{f_1} \cdot \ldots \cdot \zeta_{f_k}), \quad \text{or} \quad \nu_P^{(k)} = \int_{\mathscr{M}_+} \mu^k P(d\mu) \text{ for short}, \tag{1.4}$$

where $(f_1 \otimes \ldots \otimes f_k)(x_1, \ldots, x_k) = f_1(x_1) \ldots f_k(x_k)$ and μ^k is the k-fold product measure $\mu \otimes \ldots \otimes \mu$, provided that P is of the kth order, that is, the integrals (1.4) exist. In particular,

$$\nu_P^{(k)}(A_1 \times \ldots \times A_k) = P(\zeta_{A_1} \cdot \ldots \cdot \zeta_{A_k}).$$

Thus $\nu_P^{(k)}$ is the intensity measure of the random measure P^k in X^k whose law is the image of P by the measurable map $\mu \mapsto \mu^k$ of $\mathscr{M}_+(X)$ into $\mathscr{M}_+(X^k)$; in other words, $P^{(k)}$ has the realization μ^k if P realizes itself by μ.

The cumulant measures of P are derived from the moment measures by the formula

$$\gamma_P^{(k)}(f_1 \otimes \ldots \otimes f_k) = \sum_{\{J_1, \ldots, J_m\}} (-1)^{m-1}(m-1)! \prod_{l=1}^{m} \nu_P^{(k_l)}\Big(\bigotimes_{j \in J_l} f_j\Big) \tag{1.5}$$

where $\{J_1, \ldots, J_m\}$ runs through all partitions of $\{1, \ldots, k\}$ into mutually disjoint non-empty sets, and $k_l = \# J_l$. They play an important rôle for the estimation of the spectrum of stationary random measures on the line [5], but we will only need the second cumulant measure, or covariance measure, $\gamma_P = \gamma_P^{(2)}$ of P defined by

$$\gamma_P(f \otimes g) = \operatorname{cov}_P(\zeta_f, \zeta_g) = \nu_P^{(2)}(f \otimes g) - \nu_P(f)\nu_P(g). \tag{1.6}$$

The sets $\mathscr{M}^{\cdot\cdot}$, $\mathscr{M}^{\cdot}$, $\mathscr{M}^e$, and $\mathscr{M}^0$ belong to $\mathscr{B}$ (see [14]). The random measure P is called a *point process* or *simple point process* if $P(\mathscr{M}^{\cdot\cdot}) = 1$ or $P(\mathscr{M}^{\cdot}) = 1$, respectively. Intuitively speaking, a simple point process consists in selecting at random, following the law P, a countable and locally finite subset of X, and ζ_A is the number of points of this set which fall into A, considered as a random variable.

A convenient tool for the study of many aspects of random measures is its characteristic functional defined by

$$(1.7)\qquad \hat{P}f = P(\exp(i\zeta_f)) = \int_{\mathcal{M}_+} \exp(i\mu(f))P(d\mu), \quad f \in \mathcal{H}_0.$$

P is completely determined by $\hat{P}f$ for $f \in \mathcal{K}$, and we will use $\hat{P}$ mainly to compute the finite-dimensional distributions of P, that is, the joint laws of the random vectors $(\zeta_{f_1}, \ldots, \zeta_{f_k})$ with $f_1, \ldots, f_k \in \mathcal{H}_0$. In fact, the characteristic function of this joint law is simply

$$(1.8)\qquad (t_1, \ldots, t_k) \mapsto \hat{P}(t_1 f_1 + \ldots + t_k f_k) = P\big((\exp i(t_1\zeta_{f_1} + \ldots + t_k\zeta_{f_k}))\big)$$

with $t_1, \ldots, t_k \in R$. Hence $\zeta_{f_1}, \ldots, \zeta_{f_k}$ are independent if and only if

$$(1.9)\qquad \hat{P}(t_1 f_1 + \ldots + t_k f_k) = \prod_{l=1}^{k} \hat{P}(t_l f_l) \quad \text{for all } t_1, \ldots, t_k.$$

We are going to construct a few examples.

1. Let $\varrho \in \mathcal{M}_+$ and $0 < \varrho(X) < \infty$. Consider a single random point distributed in X according to the law $\varrho/\varrho(X)$. The law of this point process is, by definition, the image of $\varrho/\varrho(X)$ under the map $x \mapsto \varepsilon_x$; hence

$$P\varphi = \frac{1}{\varrho(X)} \int_X \varphi(\varepsilon_x)\varrho(dx), \quad \varphi \in \mathcal{H}(\mathcal{M}_+).$$

2. Consider now n independent random points distributed in X according to $\varrho/\varrho(X)$. The corresponding point process P is given by

$$P\varphi = \frac{1}{\varrho(X)^n} \int_{X^n} \varphi(\varepsilon_{x_1} + \ldots + \varepsilon_{x_n})\varrho(dx_1) \ldots \varrho(dx_n).$$

For $n = 0$ we have $P\varphi = \varphi(0)$.

3. *The Poisson process* P_ϱ. We select first an integer at random according to the Poisson law with parameter $\varrho(X)$, that is, n is selected with probability $(1/n!)\varrho(X)^n \exp(-\varrho(X))$, and we then distribute n points in X as in example 2. Therefore

$$(1.10)\qquad P_\varrho\varphi = \exp(-\varrho(X)) \sum_{n=0}^{\infty} \frac{1}{n!} \int_{X^n} \varphi(\varepsilon_{x_1} + \ldots + \varepsilon_{x_n})\varrho(dx_1) \ldots \varrho(dx_n).$$

Obviously, P_ϱ is linear, positive, normed and sigma-continuous, hence it is a random measure. By setting $\varphi = 1_{\mathcal{M}^e}$ we see that P_ϱ is a point process carried by the set of all finite point measures. By (1.7) and (1.10):

$$(1.11)\qquad \hat{P}_\varrho(f) = \exp\big(\varrho[\exp(if) - 1]\big).$$

Since $\operatorname{carr}[\exp(if) - 1] \subseteq \operatorname{carr} f$, the right-hand side of this expression makes sense even when $\varrho(X) = +\infty$, and in fact it can easily be shown by various extension procedures ([14], [15]) that in this case, too, there exists one and only one random measure P_ϱ such that (1.11) holds for every $f \in \mathcal{H}(\mathcal{M}_+)$. It is called the *Poisson process with intensity* ϱ.

By (1.8) and (1.11) the distribution of the random variable ζ_f on the probability space $(\mathcal{M}_+, \mathcal{B}(\mathcal{M}_+), P_\varrho)$ has the characteristic function whose logarithm is given by

$$(1.12)\qquad t \mapsto \log \hat{P}_\varrho(tf) = \varrho[\exp(itf) - 1];$$

hence it is the compound Poisson law determined by f and ϱ. Therefore, $P_\varrho(\zeta_f) = \varrho(f)$, that is, P_ϱ has the intensity measure ϱ as already implied by its name. Later on we will make essential use of the fact that (1.12) is linear in ϱ.

It also follows from (1.8) and (1.11) that $\zeta_{f_1}, \ldots, \zeta_{f_k}$ are independent if $f_1, \ldots, f_k$ have mutually disjoint carriers. In particular, ζ_A has the Poisson distribution with parameter $\varrho(A)$, and $\zeta_{A_1}, \ldots, \zeta_{A_k}$ are independent whenever $A_1, \ldots, A_k$ are disjoint: this is the usual definition of P_ϱ. From here, or from (1.10), it follows that P_ϱ is a point process and that P_ϱ is simple if and only if ϱ is diffuse. Moreover, by writing equation (1.3), valid for every realization μ of the process, in the form of a stochastic integral

$$\zeta_f = \int_X f(x)\zeta_{dx},$$

we recognize immediately the intuitive meaning of the name "compound Poisson law".

Finally, by repeated differentiation of (1.8) we obtain [16] the kth cumulant measure of P_ϱ according to definition (1.5):

$$(1.13)\qquad \gamma_{P_\varrho}^{(k)}(f_1 \otimes \ldots \otimes f_k) = \varrho(f_1 \cdot \ldots \cdot f_k),$$

that is,

$$\gamma_{P_\varrho}^{(k)}(A_1 \times \ldots \times A_k) = \varrho(A_1 \cap \ldots \cap A_k).$$

In particular, its covariance measure is equal to

$$(1.14)\qquad \operatorname{cov}_{P_\varrho}(\zeta_f, \zeta_g) = \varrho(f \cdot g), \quad \operatorname{cov}_{P_\varrho}(\zeta_A, \zeta_B) = \varrho(A \cap B).$$

4. *The Cox process.* Let W be a random measure in X. We select a point measure by a double random mechanism: first we obtain a measure ϱ at random following the law W; then we produce a point measure at random following the law P_ϱ. In other words, the conditional distribution of the point process P_W thus constructed, given the realization ϱ of the underlying random measure W, coincides with P_ϱ. Therefore,

$$(1.15)\qquad P_W(\varphi) = \int_{\mathcal{M}_+} P_\varrho(\varphi)\, W(d\varrho), \quad \varphi \in \mathcal{H}(\mathcal{M}_+),$$

or for short $P_W = \int_{\mathcal{M}_+} P_\varrho W(d\varrho)$, that is, P_W is the mixture of the P_ϱ's with weight $W(d\varrho)$. It is called the *Cox process* built on W.

We can use equation (1.15) for a rigorous definition of P_W. In fact, $\varrho \mapsto P_\varrho(\varphi)$ is $\mathcal{B}(\mathcal{M}_+)$-measurable (see [14]); hence the integral exists, and the functional P_W

thus defined is obviously linear, positive, normed and sigma-continuous. By choosing $\varphi = 1_{\mathscr{M}}$. or $\varphi = 1_{\mathscr{M}^{\cdot}}$. we see that P_W is a point process and P_W is simple if and only if $W(\mathscr{M}^0) = 1$. Moreover, P_W determines W completely ([14], [16]); however, any proof of this assertion is based on fairly deep analytic results about Laplace transforms or the moment problem.

Many properties of W can thus be translated into properties of P_W. We will need only a few of them. First, W is of the kth order if and only if P_W is;

$$\nu_{P_W} = \nu_W, \quad \gamma_{P_W}(f\otimes g) = \gamma_W(f\otimes g)+\nu_W(f\cdot g), \tag{1.16}$$

in particular,

$$\mathrm{var}_{P_W}(\zeta_f) = \mathrm{var}_W(\zeta_f)+\nu_W(f^2), \quad \mathrm{var}_{P_W}(\zeta_A) = \mathrm{var}_W(\zeta_A)+\nu_W(A).$$

Therefore, $\mathrm{var}_{P_W}(\zeta_A) \geqslant W(\zeta_A)$ and equality holds for every A if and only if W-almost surely ζ_A is a constant for every A, that is, if P_W is a Poisson process; thus P_W is "overdispersed" in the general case.

Finally, by (1.7), (1.12), and (1.15):

$$\hat{P}_W(f) = \hat{W}(i[1-\exp(if)]). \tag{1.17}$$

5. *The mixed Poisson process*. This is a particular case of the preceding example: W is carried by the set of all measures of the form $\vartheta\sigma$ with $\vartheta \geqslant 0$ where σ is a fixed measure in X. Hence we can write

$$P_W = \int_0^\infty P_{\vartheta\sigma}dF(\vartheta) \tag{1.18}$$

where F is a cumulative distribution function on $R_+ = [0, +\infty[$.

2. Inference about Poisson processes

Consider first two finite measures ϱ and σ in X; thus $\varrho, \sigma \in \mathscr{M}_+$ and $\varrho(X), \sigma(X) < \infty$. Set $P = P_\varrho$ and $Q = P_\sigma$.

PROPOSITION. *If $\varrho \ll \sigma$ and $h = d\varrho/d\sigma$, then $P \ll Q$ and*

$$\frac{dP}{dQ}(\mu) = \begin{cases} \exp\big(-(\varrho(X)-\sigma(X))\big)h(x_1)\dots h(x_n) & \text{if} \quad \mu = \varepsilon_{x_1}+\dots+\varepsilon_{x_n}, \\ 0 & \text{if } \mu \text{ is not of this form;} \end{cases} \tag{2.1}$$

recall that Q-almost surely μ has in fact the form (1.2).

Proof. For every $\varphi \in \mathscr{H}(\mathscr{M}_+)$ we have by (1.10)

$$P\varphi = \exp\big(-(\varrho(X)-\sigma(X))\big)\exp(-\sigma(X))\times$$
$$\times\sum_{n=0}^\infty \frac{1}{n!}\int_{X^n}\varphi(\varepsilon_{x_1}+\dots+\varepsilon_{x_n})h(x_1)\dots h(x_n)\sigma(dx_1)\dots\sigma(dx_n) = Q(\varphi\cdot\eta)$$

where η is the function on $\mathscr{M}_+$ defined by the right-hand side of (2.1).

Note that we can write (2.1) in the form

$$\log\eta = -\varrho(X)+\sigma(X)+\zeta_{\log h} \tag{2.2}$$

where

$$\zeta_{\log h}(\mu) = \log h(x_1)+\dots+\log h(x_n)$$

for μ given by (1.2).

Consider next two finite measures ϱ_0 and ϱ_1 which are absolutely continuous with respect to the finite measure σ and set

$$h_0 = \frac{d\varrho_0}{d\sigma}, \quad h_1 = \frac{d\varrho_1}{d\sigma}, \quad P_0 = P_{\varrho_0}, \quad P_1 = P_{\varrho_1}, \quad \xi = \zeta_{\log(h_1/h_0)}.$$

By (2.2) the Neyman–Pearson test of the hypothesis H_0: $P = P_0$ against the alternative H_1: $P = P_1$ looks like this: having observed the realization (1.2), accept H_1 if

$$\xi(\mu) = \sum_{j=1}^n (\log h_1(x_j)-\log h_0(x_j)) > c \tag{2.3}$$

where c is determined by the given level of the test.

To compute c note first that $h_0(x_1) > 0, \dots, h_0(x_n) > 0$ almost surely under the null hypothesis. Hence under H_0 we have $\xi(\mu) = -\infty$ if and only if $h_1(x_j) = 0$ for at least one j, that is, $\xi > -\infty$ if and only if $\zeta_N = 0$ where $N = \{x: h_1(x) = 0\}$. Therefore,

$$P_0\{\xi > c\} = P_0\{\xi > c|\ \xi > -\infty\} = P_0\{\xi > c|\ \zeta_N = 0\}.$$

Analogously to (1.12) we find that the logarithm of the characteristic function of the conditional distribution of ξ given that $\zeta_N = 0$ equals

$$t \mapsto \int_{X\setminus N}\left[\exp\left(it\log\frac{h_1}{h_0}\right)-1\right]d\varrho_0;$$

hence the distribution in question needed to compute c is the compound Poisson law determined by h_1/h_0 and ϱ_0 restricted to $X\setminus N$. In the same way we can determine the law needed to find the power of the test. Since these laws are in general not tabulated, we will have to rely on asymptotic results for $\varrho_l(X) \to \infty$, $l = 0, 1$, to be taken up later.

Consider now, for fixed finite σ, the class $\mathscr{P}(\sigma)$ of all laws P_ϱ such that $\varrho(X) < \infty$ and $\varrho \ll \sigma$. We are going to mark laws in $\mathscr{P}(\sigma)$ and their intensity measures ϱ, their densities η with respect to $Q = P_\sigma$, and the densities h of their intensity measures with respect to σ, by corresponding indices without further comment.

Suppose that f is any positive measurable function defined on X and $h' \leqslant h''$. Then

$$P'\{\zeta_f > c\} \leqslant P''\{\zeta_f > c\} \quad \text{for all } c, \tag{2.4}$$

that is, ζ_f is stochastically smaller for P' than for P''. In fact, if $f = 1_A$ with $A \in \mathscr{B}_0$, ζ_f has the Poisson distribution with parameters $\varrho'(A)$ and $\varrho''(A)$, respectively, where

$\varrho'(A) \leqslant \varrho''(A)$, and (2.4) is well known. The general statement is then derived by writing f as the limit of an increasing sequence of linear combinations with positive coefficients of such indicator functions.

Assume in particular that $h_0 \leqslant h_1$ and apply the preceding remark to the function $f = \log(h_1/h_0)$ so that $\zeta_f = \xi$. Then we see by (2.4) that the test (2.3) has an increasing power function with respect to the usual order among the functions h. In particular, it is unbiased on the level $P_0\{\xi > c\}$ for the null hypothesis $h \leqslant h_0$ against the alternative $h_0 < h$.

Formula (2.1) also implies that the family $\mathscr{P}(\sigma)$ has monotone likelihood ratios in the obvious generalized sense: if $h_0 \leqslant h_1$, then $\mu \mapsto \eta_1(\mu)/\eta_0(\mu)$ is an increasing function of μ in the part of $\mathscr{M}^e$ where η_0 does not vanish.

We pass to the case where $\log h$ is a linear function, depending on h, of some function T which does not depend on h. In other words, we consider a subfamily $(P_\vartheta)_{\vartheta\in\Theta}$ of $\mathscr{P}(\sigma)$ such that, for every $\vartheta \in \Theta$, we have

$$h_\vartheta = \exp(\langle p_\vartheta, T\rangle)$$

where T is a measurable map of X into a vector space R^l, $p_\vartheta \in R^l$, and $\langle\, ,\rangle$ denotes the standard scalar product in R^l. By (2.2) and the linearity of ζ,

$$\log\eta_\vartheta = -\varrho_\vartheta(X)+\sigma(X)+\langle p_\vartheta, \zeta_T\rangle. \tag{2.5}$$

Thus $(P_\vartheta)_{\vartheta\in\Theta}$ is an exponential family and ζ_T a minimal sufficient statistics, and the usual reasonings about statistical inference on such families apply. In particular, the family is complete for ζ_T if the set $\{p_\vartheta: \vartheta \in \Theta\}$ has an inner point in R^l so that we can then make use, for example, of the optimality of unbiased estimators based on ζ_T and the Rao–Blackwell theorem ([11], p. 121).

Note that for realization (1.2) we have

$$\zeta_T(\mu) = \Big(\sum_{j=1}^{n} T_1(x_j), \ldots, \sum_{j=1}^{n} T_l(x_j)\Big) \tag{2.6}$$

where $T_1, \ldots, T_l$ are the components of T. Moreover, for any ϑ_0 and ϑ_1 the test given by (2.3) takes the form

$$\xi = \langle p_{\vartheta_1} - p_{\vartheta_0}, \zeta_T\rangle > c. \tag{2.7}$$

Let us look at some examples. In the two first examples, $l = 1$, $\Theta \subseteq R$ and the function $\vartheta \mapsto p_\vartheta$ is increasing. Therefore, by (2.5) or by (2.7), any test of the form: accept the alternative if $\zeta_T(\mu) > c^*$, is uniformly optimal on the level $P_{\vartheta_0}\{\zeta_T > c^*\}$ for the null hypothesis $\vartheta \leqslant \vartheta_0$ against the alternative $\vartheta > \vartheta_0$. Similar statements are true regarding two-sided hypotheses.

1. *Constant intensity density*. Here, $\Theta =]0, +\infty[$, $p_\vartheta = \log\vartheta$, $T \equiv 1$, thus $h_\vartheta = \vartheta$ and $\varrho_\vartheta = \vartheta\sigma$. Therefore, the total number of points observed, that is,

$$\zeta_1(\mu) = \mu(X) = \#\operatorname{carr}\mu$$

is the sufficient statistics in question. The law of ζ_1 is available in usable form for any ϑ since it is the Poisson law with parameter $\vartheta\sigma(X)$.

2. *Exponential trend*. Here, X is a bounded interval of the real line R, σ the Lebesgue measure in X, $\Theta = R$, $p_\vartheta = \vartheta$ and $T(x) = x$, thus $h_\vartheta(x) = \exp(\vartheta x)$. The sufficient statistics ζ_T becomes in this case

$$\zeta_T(\varepsilon_{x_1} + \ldots + \varepsilon_{x_n}) = x_1 + \ldots + x_n. \tag{2.8}$$

Already in this simple case, however, the law of ζ_T is not readily available; we will come back later to its asymptotic form for $\sigma(X) \to \infty$.

3. Within the following model which includes 1 and 2 we can approximate uniformly any density h in a fixed bounded interval X: $T_m(x) = x^{m-1}$, $m = 1, \ldots, l$; thus

$$h_\vartheta(x) = \exp\Big(\sum_{m=1}^{l} p_\vartheta^{(m)} x^{m-1}\Big). \tag{2.9}$$

4. If the phenomenon under consideration is periodic with a known period, models of the following type are sometimes useful:

$$h_\vartheta(x) = \exp(p_\vartheta^{(1)} + p_\vartheta^{(2)}\sin(ax) + p_\vartheta^{(3)}\cos(ax)) \tag{2.10}$$

where a is known.

As remarked before, we can apply the usual statistical theory of exponential families (2.5) with ζ_T given by (2.6) to these examples, and in particular treat hypotheses concerning some of the $p_\vartheta^{(m)}$'s where the others are considered as nuisance parameters. For some particular cases see [9].

Regarding the estimation of p_ϑ it is reasonable, on account of the Rao–Blackwell theorem, to look for unbiased estimates, but their construction is not easy for general p_ϑ and T. Let us write down the maximum likelihood equations obtained by differentiating (2.5) with respect to the components $p_\vartheta^{(m)}$ of p_ϑ and setting the derivatives equal to 0. If T_m is ϱ_ϑ-integrable, we can differentiate

$$\varrho_\vartheta(X) = \sigma(h_\vartheta) = \sigma\big(\exp(\langle p_\vartheta, T\rangle)\big)$$

under the integral (σ-) sign with respect to $p_\vartheta^{(m)}$ (see [21], pp. 52–53) and we obtain

$$\varrho_\vartheta(T_m) = \zeta_{T_m}. \tag{2.11}$$

Now by definition of ϱ_ϑ we have $P_\vartheta(\zeta_{T_m}) = \varrho_\vartheta(T_m)$ for every ϑ, that is, ζ_{T_m} is an unbiased estimator of $\varrho_\vartheta(T_m)$, in fact the best unbiased estimator with respect to any convex continuous loss function. If we assume that $\hat\vartheta$: $\mathscr{M}^e \to R^l$ is a maximum likelihood estimator of ϑ, hence a solution of equations (2.11), then ζ_{T_m} takes the form

$$\zeta_{T_m} = \varrho_{\hat\vartheta}(T_m), \quad m = 1, \ldots, l. \tag{2.12}$$

We are therefore tempted to introduce the $\varrho_\vartheta(T_m)$, $m = 1, \ldots, l$, as new "natural" parameters instead of $p_\vartheta^{(1)}, \ldots, p_\vartheta^{(l)}$.

Note that if every T_m^2 is ϱ_ϑ-integrable for all p_ϑ in some neighbourhood of the origin of R^l, then we have

$$\frac{\partial\varrho_\vartheta(T_m)}{\partial p_\vartheta^{(k)}} = \sigma\big(T_m T_k \exp(\langle p_\vartheta, T\rangle)\big).$$

If, moreover, $T_1, \ldots, T_l$ are linearly independent in the space $\mathscr{L}_2(\sigma)$, then $\det(\sigma(T_m T_k)_{m,k=1,\ldots,l}) \neq 0$ and the desired parameter transformation is at least possible in a neighbourhood of $p_\vartheta = 0$. In many examples it can be done directly and globally.

In the first example above, $\varrho_\vartheta(T)$ is linear in ϑ, hence we get an unbiased estimator of ϑ. We have in fact $l = 1$, $T \equiv 1$, $\varrho_\vartheta(T) = \vartheta\sigma(X)$, hence the solution of (2.11) is

$$\hat{\vartheta} = \frac{\zeta_1}{\sigma(X)}. \tag{2.13}$$

In the second example, we have

$$\varrho_\vartheta(T) = \int_X x\exp(\vartheta x)\,dx. \tag{2.14}$$

Assume that $X \subseteq R_+$. Then the function $g(\vartheta) = \varrho_\vartheta(T)$ is strictly increasing and analytic, $\lim\limits_{\vartheta\to-\infty} g(\vartheta) = 0$, $\lim\limits_{\vartheta\to+\infty} g(\vartheta) = +\infty$; hence we can introduce $g(\vartheta)$ for all ϑ as a new parameter with the range $]0, +\infty[$. The statistics (2.8) is the best unbiased estimator of $g(\vartheta)$ and by (2.12) the maximum likelihood estimator of ϑ is $\hat{\vartheta} = g^{-1} \circ \zeta_T$; in this sense it appears well justified.

We are now going to study the behaviour of the statistics in question if we take an observation in a large domain. We are usually faced with the following situation: we have a fixed measure σ such that $\sigma(X) = +\infty$, and we are looking at Poisson processes P whose intensity measure ϱ is absolutely continuous with respect to σ with a density h. Since in practice we can observe the realization of such a process only in bounded domains, we consider a set $K \in \mathscr{B}_0$ and we apply the preceding theory to the restriction P^K of P to K. Then we investigate what happens if K becomes large in a certain way. Let us recall that, as usually in the statistics of stochastic processes, we dispose of only one realization from which to draw inferences; making K large corresponds to "taking many observations" in classical statistics.

To make the concept of the restriction P^K of any random measure P precise, denote for any $\mu \in \mathscr{M}_+$ by μ^K the measure defined by $\mu^K(f) = \mu(f 1_K)$ where $f \in \mathscr{K}_0$ or $\mu^K(A) = \mu(A \cap K)$ where $A \in \mathscr{B}_0$; thus μ^K is again an element of $\mathscr{M}_+ = \mathscr{M}_+(X)$ and not of $\mathscr{M}_+(K)$. Intuitively speaking, P^K is obtained from P by replacing every realization μ by μ^K, that is, P^K is the image of P under the measurable map $\mu \mapsto \mu^K$ of $\mathscr{M}_+$ into $\mathscr{M}_+$. Therefore, the law of the random variable ζ_f on the probability space $(\mathscr{M}_+, \mathscr{B}(\mathscr{M}_+), P^K)$ is the same as that of ζ_f^K on the original probability space $(\mathscr{M}_+, \mathscr{B}(\mathscr{M}_+), P)$ where, of course, $\zeta_f^K(\mu) = \mu^K(f)$. The same is true for joint distributions. In the case of a Poisson or more generally a Cox process we have $(P_\varrho)^K = P_{\varrho^K}$, $(P_W)^K = P_{W^K}$ and we simply write P_ϱ^K and P_W^K instead.

Let now $P = P_\varrho$ be the Poisson process with a finite or infinite intensity measure ϱ and f a measurable and locally bounded function on X such that $\varrho(f^2) = \infty$.

By (1.13) we have

$$P\zeta_f^K = \varrho^K(f), \qquad \operatorname{var}_P \zeta_f^K = \varrho^K(f^2).$$

We form the standardized random variable

$$\frac{\zeta_f^K - \varrho^K(f)}{\sqrt{\varrho^K(f^2)}}, \tag{2.15}$$

and we want to find manageable criteria in order that its law converges weakly to the standard normal law for a given sequence (K_m) such that

$$\lim_{m\to\infty} \varrho^{K_m}(f^2) = \infty. \tag{2.16}$$

We can obtain a necessary and sufficient condition which is easy to apply by evaluating directly the characteristic function of the law of the variable (2.15). In fact, by (1.12) the logarithm of this characteristic function is

$$t \mapsto \varrho^K\left(\exp\left(it\frac{f}{\sqrt{\varrho^K(f^2)}}\right) - 1 - it\frac{f}{\sqrt{\varrho^K(f^2)}}\right).$$

By expanding this with the help of the exponential series, or simply by applying the classical definition of the cumulants and (1.13) we find that this is equal to $-t^2/2$ plus

$$\sum_{k=3}^{\infty} \frac{(it)^k}{k!} \frac{\varrho^K(f^k)}{\varrho^K(f^2)^{k/2}}; \tag{2.17}$$

the convergence of this series is obvious since f is bounded in K. Note that the coefficient of $(it)^k/k!$ is the kth cumulant of the variable (2.15).

Thus we see that (2.15) with $K = K_m$ is asymptotically normally distributed for $m \to \infty$ if and only if (2.17) with $K = K_m$ converges to 0 for every $t \in R$. Let us take up the examples above.

1. *Constant intensity density.* We are interested in the statistics ζ_1, that is, $f \equiv 1$, and we get the usual normal approximation of the Poisson law: in fact, $\varrho^K(f^k) = \varrho(K)$ for all k, hence the coefficients of (2.17) go to 0 for any sequence (K_m) with $\varrho(K_m) \to \infty$.

2. *Exponential trend.* We consider, slightly more generally, a function of the form $f(x) = x^\beta$, $\beta > 0$, and the sets $K_s = [0, s]$ where $s > 0$. Then, looking first at the case $\vartheta = 0$, we find

$$\varrho_0^{K_s}(f^k) = \frac{1}{k\beta+1} s^{k\beta+1};$$

hence

$$\frac{\varrho_0^{K_s}(f^k)}{\varrho_0^{K_s}(f^2)^{k/2}} = \frac{(2\beta+1)^{k/2}}{k\beta+1} s^{1-k/2} \to 0$$

for $s \to \infty$ if $k \geqslant 3$, thus (2.17) also goes to 0. Applying this with $\beta = 1$ we would therefore reject the nullhypothesis $\vartheta = 0$, that is, no trend, in favour of the alterative $\vartheta > 0$, that is, a positive trend, if, for the observation (1.2), we find that

$$\frac{\sum_{j=1}^{n} x_j - s^2/2}{\sqrt{s^3/3}} > c$$

where c is determined by the level of the test as in the case of a standard normal distribution. Applications can be found in [9], p. 49, and [25].

If $\vartheta > 0$, the convergence to 0 of $\varrho_\vartheta^{K_s}(f^k)/\varrho_\vartheta^{K_s}(f^2)^{k/2}$ with $k \geqslant 3$ is immediate, whereas for $\vartheta < 0$ condition (2.16), with $\varrho = \varrho_\vartheta$, is not satisfied for any β. It does hold, of course, for any $\vartheta \geqslant 0$.

To give an example where asymptotic normality is not present in spite of (2.16), consider again the Lebesgue measure ϱ_ϑ on R and $K_s = [0, s]$ as before, but $f(x) = \exp x$. Then

$$\lim_{s\to\infty} \frac{\varrho^{K_s}(f^k)}{(\varrho^{K_k}(f^2))^{k/2}} = \frac{2^{k/2}}{k} \quad \text{for} \quad k = 3, 4, \ldots;$$

hence the limit law still exists but is not normal.

Until now we have motivated our "asymptotic" study by the desire to find manageable approximations to the distributions of the relevant statistics. There is, however, also the more fundamental question about the possibility of a "perfect estimation" of a parameter on the basis of the observation of a single realization in the entire space X, or in other words, the distinguishability of different Poisson processes by such an observation. Let us look at this problem from a general point of view.

We shall use the intensity measures as natural indices. Thus we consider a family of Poisson processes $(P_\varrho)_{\varrho\in\Sigma}$ where $\Sigma \in \mathscr{B}(\mathscr{M}_+)$. By a parameter p of this family we mean a measurable map of Σ into a measurable space $(U, \mathscr{U})$. An estimate of p is a measurable map η of $\mathscr{M}^{\cdot\cdot}$ into $(U, \mathscr{U})$: if we observe the point measure μ as the realization of the process in question, then $\eta(\mu)$ is the estimated value of p on the basis of this observation.

It will be convenient to describe any estimate η by the associate partition of $\mathscr{M}^{\cdot\cdot}$ into the mutually disjoint sets

(2.18) $$\mathscr{M}_u = \{\mu \in \mathscr{M}^{\cdot\cdot}: \eta(\mu) = u\} = \eta^{-1}\{u\}, \quad u \in U.$$

Since

$$\bigcup\{\mathscr{M}_u: u \in V\} = \eta^{-1}(V) \quad \text{for every } V \subseteq U,$$

this partition is measurable, that is, $\bigcup\{\mathscr{M}_u: u \in V\} \in \mathscr{B}(\mathscr{M}^{\cdot\cdot})$ for every $V \in \mathscr{U}$. Conversely, given a measurable partition $(\mathscr{M}_u)_{u\in U}$ of $\mathscr{M}^{\cdot\cdot}$, the map η defined by $\eta(\mu) = u$ for $\mu \in \mathscr{M}_u$ is measurable and satisfies (2.18).

The map η is called a *perfect estimate of p* if, vaguely speaking, it gives almost surely the "true value" of the parameter, that is, $P_\varrho\{\eta = p(\varrho)\} = 1$ for every $\varrho \in \Sigma$, or by (2.18):

(2.19) $$P_\varrho(\mathscr{M}_{p(\varrho)}) = 1 \quad \text{for every } \varrho \in \Sigma.$$

Thus, P_ϱ is carried by $\mathscr{M}_{p(\varrho)}$. This implies, of course, that P_ϱ and $P_{\varrho'}$ are singular, that is, $P_\varrho \perp P_{\varrho'}$, whenever $p(\varrho) \neq p(\varrho')$. Conversely, assume that the family $(P_\varrho)_{\varrho\in\Sigma}$ has this property. It Σ is finite, it can be proved rapidly that then a perfect estimate of p exists [19]; in the case of an infinite set Σ this is still true under some additional assumptions of a technical nature. We will, however, not need this fact and exploit the relation between mutual singularity of laws and the existence of perfect estimates only by either exhibiting directly perfect estimates and deducing mutual singularity in the sense above, or by deducing the non-existence of perfect estimates from non-singularity.

First we recall a criterion for singularity or absolute continuity of two Poisson laws. Let $\varrho_0, \varrho_1 \in \mathscr{M}_+$ and take any measure $\sigma \in \mathscr{M}_+$ such that $\varrho_0, \varrho_1 \ll \sigma$. Set

$$h_0 = \frac{d\varrho_0}{d\sigma}, \quad h_1 = \frac{d\varrho_1}{d\sigma}, \quad P_0 = P_{\varrho_0}, \quad P_1 = P_{\varrho_1}.$$

THEOREM. *$P_0 \perp P_1$ if and only if* $\int_X (\sqrt{h_0} - \sqrt{h_1})^2 d\sigma = +\infty$ *and $P_0 \ll P_1$ if and only if $\varrho_0 \ll \varrho_1$* and $\int_X (\sqrt{h_0} - \sqrt{h_1})^2 d\sigma < +\infty$.

Proof: see [22].

If ϱ_0 and ϱ_1 are bounded, P_0 and P_1 are not singular because the measure 0 is realized both under P_0 and under P_1 with positive probability. Hence, for arbitrary $\varrho_0, \varrho_1 \in \mathscr{M}_+$, we cannot distinguish with certainty between P_0 and P_1 by observing only a single realization in a fixed compact set K. Intuitively, it is to be expected that perfect estimates, if they exist, can be obtained from estimates based on the observation in a compact set K by going to the limit $K \nearrow X$.

Let us first look at the case of a two-element family $\Sigma = \{\varrho_0, \varrho_1\}$ again. Given a fixed level, we may form the test (2.3) for every compact K with h_0 and h_1 replaced by their restrictions to K, or in other words with ξ replaced by $\xi^K(\mu) = \xi(\mu^K)$. Thus we accept the hypothesis $P = P_1$ when $\xi(\mu^K) > c_K$ with c_K having been chosen as a function of the level.

PROPOSITION [8]. *Let (K_n) be a sequence of compact sets such that $K_n \nearrow X$. Then the sequence of the afore-mentioned tests for $P = P_0$ against $P = P_1$ is consistent, that is,* $\lim_{n\to\infty} P_1\{\xi^{K_n} > c_{K_n}\} = 1$, *if and only if $P_0 \perp P_1$.*

The proof uses the customary martingale arguments.

In the majority of concrete situations, a perfect estimator is derived, as a limit for $K \nearrow X$, by applying some ergodic type or related limit theorem. We consider some examples.

1. *Constant intensity density.* Here, Σ is the set of all $\vartheta\sigma$ with a fixed σ and $\vartheta \in \Theta =]0, +\infty[$ where now, however, $\sigma(X) = \infty$. We consider ϑ itself as "the parameter", that is, $p(\vartheta\sigma) = \vartheta$, and we write P_ϑ for $P_{\vartheta\sigma}$. The estimator (2.13) based on K is

$$\hat{\vartheta}_K = \frac{\zeta_1^K}{\sigma(K)},$$

in practical terms

$$\hat{\vartheta}_K(\mu) = \frac{\#(K \cap \operatorname{carr}\mu)}{\sigma(K)}.$$

By (1.14), writing $\operatorname{var}_\vartheta$ for the variance with respect to P_ϑ, we have

$$\operatorname{var}_\vartheta(\hat{\vartheta}_K) = \frac{\vartheta}{\sigma(K)};$$

hence $\hat{\vartheta}_{K_n}$ converges in $L_2(P_\vartheta)$ to ϑ as $\sigma(K_n) \to \infty$. Given a finite or countable subset Θ' of Θ, we can then construct, by extracting successively subsequences and applying the diagonal principle, a subsequence $K'_m = K_{n_m}$ such that, for every $\vartheta \in \Theta'$, we have P_ϑ-almost surely $\lim\limits_{m\to\infty} \hat{\vartheta}_{K'_m} = \vartheta$. In particular, $P_{\vartheta_0} \perp P_{\vartheta_1}$ if $\vartheta_0 \neq \vartheta_1$ which, of course, also follows immediately from the theorem above.

If X has a richer structure, we can find sequences K_n such that

$$\lim_{n\to\infty} \hat{\vartheta}_{K_n} = \vartheta \quad P_\vartheta\text{-almost surely for every } \vartheta > 0. \tag{2.20}$$

By the general ergodic theorem for random measures, this is true, for example, in the following situation: $X = R^d$, σ is the d-dimensional Lebesgue measure λ, and (K_n) is a regular sequence, that is, $\lim\limits_{n\to\infty} \lambda(K_n) = \infty$ and

$$\liminf_{n\to\infty} \frac{\lambda(K_n)}{\lambda(K_n^b)} > 0$$

where K^b denotes the smallest closed ball with center 0 which contains K. In particular, K_n may itself be a ball of radius s_n which contains 0, or in the case $d = 1$ the interval $[0, s_n]$, where $s_n \to \infty$.

If (2.20) holds, a perfect estimator of p is given by

$$\eta(\mu) = \begin{cases} \lim\limits_{n\to\infty} \hat{\vartheta}_{K_n}(\mu) & \text{if this limit exists,} \\ 0 & \text{if not.} \end{cases}$$

2. *Exponential trend.* Here, $X = R_+$, Σ is the set of all ϱ_ϑ such that $\varrho_\vartheta \ll \lambda$, and $h_\vartheta = d\varrho_\vartheta/d\lambda$ is given by

$$h_\vartheta(x) = \exp(\vartheta x), \quad \vartheta \in \Theta \subseteq R.$$

As before, $p(\varrho_\vartheta) = \vartheta$ and $P_\vartheta = P_{\varrho_\vartheta}$.

If $\vartheta_0 < \vartheta_1$, then the theorem above implies at once:

$$P_{\vartheta_0} \perp P_{\vartheta_1} \quad \text{if and only if} \quad \vartheta_1 \geqslant 0,$$

$$P_{\vartheta_0} \sim P_{\vartheta_1} \quad \text{if and only if} \quad \vartheta_1 < 0.$$

Thus, a perfect estimation is impossible as soon as the set Θ of all allowed parameter values contains a negative number, that is, Σ contains an intensity measure with negative exponential trend. On the other hand, any two positive exponential trends can be distinguished almost surely on the basis of a single observation in all

of R_+. Note that we had already observed an analogous dichotomy when studying the asymptotic law of the statistics $\zeta_f^{K_s}$, with $f(x) = x^\beta$, $\beta > 0$, and $K_s = [0, s]$, for $s \to \infty$.

3. *Line processes.* Here, $X = R \times S$ where $S = S_1$ stands for the unit circle, thus X is the infinite cylinder. X may be thought of as a representation of the set of all oriented lines in the plane R^2, such a line x being represented by its signed distance p from the origin 0 which is positive when 0 lies on the left bank of x, and by the angle φ between the abscissa and x; thus $x = (p, \varphi)$ with $p \in R$ and $\varphi \in S$. Via this representation, the group of all translations in R^2 generates a group $\mathscr{G}$ acting in X. The most general translation invariant Poisson line process is then given by a Poisson process P_ϱ on X whose intensity measure has the form $\varrho = \lambda \otimes \varkappa$ where λ is the one-dimensional Lebesgue measure and $\varkappa \in \mathscr{M}_+(S)$. Let Σ be the set of all these measures ϱ; to simplify the notation we write $P_\varkappa$ for $P_{\lambda\otimes\varkappa}$. Every $P_\varkappa$ is ergodic for $\mathscr{G}$ (see [10]) which implies that $P_\varkappa \perp P_{\varkappa'}$ for $\varkappa \neq \varkappa'$; this follows as well immediately from the theorem above.

In the present example, the parameter we are interested in is the intensity measure itself. This amounts to $p(\lambda \otimes \varkappa) = \varkappa$. Perfect estimates can be constructed in the following natural way [10]. Let K_s, for $s > 0$, be the set of all lines that hit the disk of centre 0 and radius s, thus $K_s = [-s, s] \times S$. For any measure $\mu \in \mathscr{M}_+(X)$, denote by μ_s the projection to S of the restriction of μ to K_s, that is,

$$\mu_s(A) = \mu([-s, s] \times A) \quad \text{for every } A \in \mathscr{B}(S).$$

Then

$$\hat{\varkappa}_s(\mu) = \frac{\mu_s}{2s}, \quad \mu \in \mathscr{M}^{\cdot\cdot}(X),$$

may be regarded as an estimate of $\varkappa$. By (1.14) it is unbiased:

$$P_\varkappa \hat{\varkappa}_s(f) = \varkappa(f), \quad f \in \mathscr{K}(S),$$

with variance

$$\operatorname{var}_\varkappa \hat{\varkappa}_s(f) = \frac{\varkappa(f^2)}{2s}.$$

Hence $\hat{\varkappa}_s(f) \to \varkappa(f)$ in $\mathscr{L}_2(P_\varkappa)$ for $s \to \infty$, but the following statements can also easily be proved:

For every $f \in \mathscr{K}(S)$ *we have* $P_\varkappa$*-almost surely* $\lim\limits_{s\to\infty} \hat{\varkappa}_s(f) = \varkappa(f)$ (see [10]);

$P_\varkappa$*-almost surely we have* $\hat{\varkappa}_s \to \varkappa$ *weakly* (see [24], Chapt. II, Theorem 7.1);

$P_\varkappa$*-almost surely we have*

$$\lim_{s\to\infty} \sup_{I \in \mathscr{J}} |\hat{\varkappa}_s(I) - \varkappa(I)| = 0,$$

where $\mathscr{J}$ *is the class of all segments of* S (see [20]).

In particular,

$$\hat{\varkappa} = \begin{cases} \text{weak limit of } \hat{\varkappa}_s & \text{if it exists,} \\ 0 & \text{if not,} \end{cases} \tag{2.21}$$

is a perfect estimate of $\varkappa$.

To conclude this chapter, let us mention a different interpretation of the idea of "observing the process first in a bounded domain and then going to the limit". Instead of fixing such a domain in advance, we may determine it as a function of the observation. In particular, we may do this in such a way that the number of points of the realization which we take into account is fixed in advance. We will confine ourselves to two examples.

1. *Constant intensity density on the positive half-line.* Consider on $R_+ = [0, +\infty[$ the family Σ of all $\varrho = \vartheta\lambda$ where λ is again the Lebesgue measure and $\vartheta > 0$. Let n be a fixed natural number. For any $\mu \in \mathscr{M}^{\cdot}(X)$, arrange the points of $\operatorname{carr}\mu$ in increasing order:

$$\operatorname{carr}\mu = \{x_1(\mu), x_2(\mu), \ldots\}, \quad 0 \leqslant x_1(\mu) < x_2(\mu) \ldots$$

Then $n^{-1}x_n$ converges P_ϑ-almost surely to ϑ^{-1}; hence

$$\eta = \begin{cases} \lim\limits_{n\to\infty} n^{-1}x_n & \text{if this limit exists,} \\ 0 & \text{if not} \end{cases}$$

is a perfect estimate of ϑ^{-1}. See [9], [18], [19] for a more detailed study of the estimates $n^{-1}x_n$.

3. *Line processes.* In the context described above, suppose that $\varkappa \neq 0$. Then $P_\varkappa$-almost all $\mu \in \mathscr{M}^{\cdot\cdot}(X)$ are carried by a sequence of lines $x_n(\mu) = (p_n(\mu), \varphi_n(\mu))$ whose distances from 0 are all different, and we can number them in such a way that $0 \leqslant |p_1(\mu)| < |p_2(\mu)| < \ldots$ This is a "measurable numbering", that is, the p_n's and φ_n's are random variables in R and S, respectively. Moreover, the random set $\{p_1, p_2, \ldots\}$ behaves like the realization of a Poisson process on R with intensity measure $\varkappa(S)\lambda$, the random angles $\varphi_1, \varphi_2, \ldots$ are independent and identically distributed according to the law $\varkappa^0 = \varkappa/\varkappa(S)$, and $\{p_1, p_2, \ldots\}$ and $\{\varphi_1, \varphi_2, \ldots\}$ are independent of each other. Having observed $x_1, \ldots, x_n$ where n is fixed in advance, we estimate $\varkappa(S)^{-1}$ as in the preceding example by p_n/n, and $\varkappa^0$ is estimated by the "empirical law"

$$\hat{\varkappa}_n^0 = \frac{1}{n}(\varepsilon_{\varphi_1} + \ldots + \varepsilon_{\varphi_n}).$$

For $n \to \infty$, these estimates together give us a perfect estimate of $\varkappa$. In particular, the assertions made above on the various kinds of convergence of $\hat{\varkappa}_s$ to $\varkappa$ hold as well for $\hat{\varkappa}_n^0$ and $\varkappa^0$, respectively. For proofs and further details see [10], [20].

3. Filterring of the random intensity of Cox processes

Let $\mathscr{W}$ be a family of random measures in X. For every $W \in \mathscr{W}$ we can build the Cox process P_W. It is natural to consider the problem of statistical inference about W, based on a realization of P_W. For example, in the case of the basic space $X = R^k$ and translation invariant laws P_W, the ergodicity of P_W with respect to the translation group can be proved for every W of a fairly large class $\mathscr{W}$ by using standard arguments about correlation measures ([18], [19]). Then a perfect estimate of W will be possible within the family $(P_W)_{W\in\mathscr{W}}$.

However, in most applications of Cox processes we are interested in a quite different type of statistical inference which is not aimed at W. Recall that a typical realization μ of P_W is obtained by first producing a realization ϱ of W and then a realization μ of the Poisson process P_ϱ. Now, usually we can observe μ but not ϱ; in other words, the hidden realization manifests itself in the observable point measure μ, and we are interested in inference about ϱ based on μ. Thus we are faced with what is called a *filterring problem*: inference about the realization of one process from the observation of the simultaneous realization of another process. Of course, these two processes must be "jointly distributed", that is, defined on the same probability space.

Examples of problems of this kind abound [23]. Apart from the one that gave rise to the concept of a Cox process where ϱ represents the strength of a thread running through a loom and μ the sequence of instances where this thread breaks, let us only mention the following one. A substance is injected into the veins of a guinea-pig. Its concentration in the animal's blood and in the course of time is subject to chance, hence it may be regarded as the realization ϱ of a random measure W. Thus ϱ is the object of our interest but it cannot be observed. We can, however, "mark" the original substance with a radioactive substance, and the coordinates in space and time of the emissions represent the points of the corresponding realization μ of the Cox process P_W. Quite often μ is observable, hence we have the problem of finding ϱ from μ.

From a purely mathematical point of view, too, it is often more sensible to look at the filterring problem and not at that of estimating W or a parameter of W. In fact, there are many situations where every $W \in \mathscr{W}$ is carried by the same set $\Sigma \in \mathscr{B}(\mathscr{M}_+(X))$ which has the property that the P_ϱ with $\varrho \in \Sigma$ are mutually singular laws. Then it can easily be shown that, for $W, W' \in \mathscr{W}$, we have $P_W \perp P_{W'}$ if and only if $W \perp W'$ (see [19]). Hence a perfect estimate of W is only possible if the laws $W \in \mathscr{W}$ are mutually singular, and this is rarely so. Speaking more intuitively, when the observed realization μ falls into the carrier of P_ϱ, it is in principle possible to determine ϱ but impossible to say anything about the behaviour of W outside the set $\{\varrho\}$ unless we have already a lot of a priori knowledge about W.

Three important classes of examples of such a set Σ were given in the preceding chapter. In the first example, any Cox process P_W such that W is carried by Σ, is a mixed Poisson process with respect to σ. If, moreover, $X = R^k$ and $\sigma = \lambda$ is the k-dimensional Lebesgue measure, then P_W is ergodic for the group of translations if and only if it is actually a Poisson process; thus in general P_W is very much nonergodic.

In the third example, we have mixtures of translation invariant Poisson line processes. Again, any of them is ergodic under the group $\mathscr{G}$ if and only if it is a Poisson process. Let us look at the simplest parameter, namely the density of the intensity measure with respect to the surface measure on $X = R\times S$. The surface measure is $\lambda_X = \lambda_R\otimes\lambda_S$ where λ_R and λ_S denote the Lebesgue measure on R and S, respectively. The intensity measure of P_W, being invariant under $\mathscr{G}$, has the form

$$\nu_{P_W} = c(W)\,\lambda_X \quad \text{with} \quad c(W)\geqslant 0.$$

In order to simplify the notations, we do not regard W as a probability law in Σ but in $\mathscr{M}_+(S)$, by the identification $\varkappa \leftrightarrow \lambda\otimes\varkappa$. Then by (1.16):

$$c(W) = \frac{1}{2\pi}\int\limits_{\mathscr{M}_+(S)} \varkappa(S)\,W(d\varkappa);$$

this is the density in question. By the reasoning above, we can expect a perfect estimation of the parameter c only in the case where, for every $W\in\mathscr{W}$, the function defined on $\mathscr{M}_+(S)$ by $\varkappa\mapsto\varkappa(S)$ is W-almost surely constant, that is, almost every ergodic component $P_\varkappa$ of P_W has the same intensity density $(2\pi)^{-1}\varkappa(S)$. In fact, in this case the obvious estimator $(2\pi)^{-1}\hat{\varkappa}(S)$, with $\hat{\varkappa}$ given by (2.21), is perfect for c within $\mathscr{W}$ (see [10], [17]).

Let us come back to the general situation. Since we are going to deal with several "jointly distributed" processes, we start with a basic probability space $(\Omega, \mathscr{A}, P)$ and three random elements on it. Firstly, we have a map $\tilde{\varrho}$ of Ω into $\mathscr{M}_+ = \mathscr{M}_+(X)$ which is measurable with respect to $\mathscr{A}$ and $\mathscr{B}(\mathscr{M}_+)$. Hence its law W is a random measure in the sense defined in the first chapter, but we will call $\tilde{\varrho}$ as well a random measure.

Secondly, we take a measurable map $\tilde{\mu}$ of Ω into $\mathscr{M}^{\cdot\cdot} = \mathscr{M}^{\cdot\cdot}(X)$ such that the joint distribution P^W of $\tilde{\varrho}$ and $\tilde{\mu}$ is that of a random measure with law W and the Cox process built on it. This means that for every $\mathscr{B}(\mathscr{M}_+)\otimes\mathscr{B}(\mathscr{M}^{\cdot\cdot})$-measurable and P^W-integrable or positive function g on $\mathscr{M}_+\times\mathscr{M}^{\cdot\cdot}$ we have

$$P^W(g) = \int\limits_{\mathscr{M}_+}\int\limits_{\mathscr{M}^{\cdot\cdot}} g(\varrho,\mu)P_\varrho(d\mu)\,W(d\varrho) \tag{3.1}$$

where

$$P^W(g) = \int\limits_{\mathscr{M}_+\times\mathscr{M}^{\cdot\cdot}} g\,dP^W = \int\limits_{\Omega} g(\tilde{\varrho}(\omega),\tilde{\mu}(\omega))P(d\omega).$$

Thirdly, there is a random element u on Ω with values in some measurable space, and we want to estimate $u(\omega)$ on the basis of a complete or partial observation of $\tilde{\mu}(\omega)$. Quite often u will be a function, or "parameter", of $\tilde{\varrho}$ alone. For example, $u(\omega)$ might be $\tilde{\varrho}(\omega)$ itself, or $\tilde{\varrho}((\omega))(f)$ with a particular function $f\in\mathscr{H}_0$, or the value of the density $d\tilde{\varrho}(\omega)/d\sigma$ of $\tilde{\varrho}(\omega)$ with respect to a fixed measure σ in a fixed point x_0, etc.

For a given law W in $\mathscr{M}_+$ the construction of a probability space $(\Omega, \mathscr{A}, P)$ and random measures $\tilde{\varrho}$ and $\tilde{\mu}$ with the required properties is of course very easy: is suffices to take $\Omega = \mathscr{M}_+\times\mathscr{M}^{\cdot\cdot}$, $\mathscr{A} = \mathscr{B}(\mathscr{M}_+)\otimes\mathscr{B}(\mathscr{M}^{\cdot\cdot})$, $P = P^W$ where P^W is defined by (3.1), and finally $\tilde{\varrho}(\varrho,\mu) = \varrho$, $\tilde{\mu}(\varrho,\mu) = \mu$. The advantage of this "standard space" is that it leads to slightly simpler notations, and we will stick to it for this reason.

Suppose we are given a family $\mathscr{W}$ of laws in $\mathscr{M}_+$. In the particular case where u is a function of ϱ only, our filterring problem amounts to a problem of statistical inference about the indices ϱ of the family of Poisson processes $(P_\varrho)_{\varrho\in\mathscr{M}_+}$. The difference with the situation treated in the second chapter lies in the fact that there is also given, and known, the family $\mathscr{W}$. In accordance with definition (1.15) of a Cox process we may interpret every W as an a priori distribution in the set of all P_ϱ's. The smaller $\mathscr{W}$ is, the more specific a priori information we have. If $\mathscr{W}$ consists of a single law W, the a priori distribution is completely known, we are in the pure Bayesian situation, and we can apply Bayesian methods. If $\mathscr{W}$ is somewhat larger but not all of $\mathscr{P}(\mathscr{M}_+)$, we have some information about the a priori law, and finally if $\mathscr{W}$ contains all laws W in $\mathscr{M}_+$ we know nothing beforehand. The last case is precisely the one treated in Chapter 2, and the two others will be taken up now.

As usual, the quality of our decisions will be measured by a loss function. Let Δ be the decision space, endowed with some reasonable sigma-algebra. Then the loss function L is a real-valued function on $\mathscr{M}_+\times\mathscr{M}^{\cdot\cdot}\times\Delta$, and $L(\varrho,\mu,\delta)$ represents the loss, or the cost, when $\tilde{\varrho}$ and $\tilde{\mu}$ take the values ϱ and μ, respectively, and we take the action δ. Since our decision is to be based on the observation of the point measure μ, we mean by a strategy D a measurable map from $\mathscr{M}^{\cdot\cdot}$ into Δ. For fixed ϱ, the risk which we run with D is the average loss

$$K(\varrho, D) = \int\limits_{\mathscr{M}^{\cdot\cdot}} L(\varrho,\mu,D(\mu))P_\varrho(d\mu),$$

and the Bayes risk for an a priori law $W\in\mathscr{W}$ is, by (3.1), equal to

$$K(W, D) = \int\limits_{\mathscr{M}_+} K(\varrho, D)\,W(d\varrho) = P^W L(\tilde{\varrho},\tilde{\mu}, D\circ\tilde{\mu}), \tag{3.2}$$

assuming, of course, that all these expectations exist. We are looking for strategies that make the $K(W, D)$'s with $W\in\mathscr{W}$ as small as possible in an appropriate sense.

In most applications, the loss will not depend on μ. In particular, if we are interested in inference about a parameter of ϱ, that is, if $u(\varrho,\mu) = u(\varrho)$ does not depend on μ, the loss $L(\varrho,\mu,\delta) = L(\varrho,\delta)$ will usually be some kind of measure of the deviation between the true value $u(\varrho)$ and the estimated value δ. Abusing our notations a little bit more, we will then regard $\tilde{\varrho}$ as a random element defined on $\mathscr{M}_+$ as well as on Ω, namely as the identical map $\tilde{\varrho}(\varrho) = \varrho$.

As explained in the preceding chapter, we will have to found a decision on the observation of μ in a compact subset K of X. Therefore, as long as we are looking at a fixed domain and are not interested in the asymptotic behaviour of our procedures for $K\nearrow X$, we will assume that X itself is compact.

Consider first the purely Bayesian case where we have a single well-known W. Suppose that W-almost all ϱ have a density $h_\varrho = d\varrho/d\sigma$ with respect to a fixed $\sigma\in\mathscr{M}_+$. By (2.1), if we have observed the point measure (1.2), the likelihood function is, up to a factor which does not depend on ϱ:

$$\varrho \mapsto \exp(-\varrho(X)) \prod_{j=1}^{n} h_\varrho(x_j).$$

Therefore the a posteriori law of $\tilde{\varrho}$ given the observation of μ is the law in $\mathcal{M}_+$ defined by

$$W_\mu(d\varrho) = \frac{\exp(-\varrho(X)) \prod_{j=1}^{n} h_\varrho(x_j)\, W(d\varrho)}{\int_{\mathcal{M}_+} \exp(-\varrho'(X)) \prod_{j=1}^{n} h_{\varrho'}(x_j)\, W(d\varrho')}. \tag{3.3}$$

Now suppose that L is a function of ϱ and δ only. The a posteriori loss caused by a decision δ is then given by

$$W_\mu L(\tilde{\varrho}, \delta) = \int_{\mathcal{M}_+} L(\varrho, \delta)\, W_\mu(d\varrho),$$

and the usual Bayes theory [11] shows that, in order to have a strategy D which minimizes $K(W, D)$ it suffices to choose $D(\mu)$ for every $\mu \in \mathcal{M}^{\cdot\cdot}$ so as to minimize $\delta \mapsto W_\mu L(\tilde{\varrho}, \delta)$ for $\delta = D(\mu)$, that is,

$$W_\mu L(\tilde{\varrho}, D(\mu)) \leqslant W_\mu L(\tilde{\varrho}, \delta) \quad \text{for every } \delta \in \Delta.$$

Let us look at a real-valued parameter $u(\varrho)$. We will then naturally take $\Delta = R$. It is also very well known that for the loss function

$$L(\varrho, \delta) = \big(u(\varrho) - \delta\big)^2 \tag{3.4}$$

the a posteriori loss is minimal for the a posteriori expectation, that is,

$$\hat{u} = D(\mu) = W_\mu(u) \tag{3.5}$$

gives the best strategy provided that $u \in \mathcal{L}_2(W_\mu)$ for P_W-almost all μ. For the loss function $L(\varrho, \delta) = |u(\varrho) - \delta|$ if u has a continuous cumulative distribution function with respect to the law W_μ, its a posteriori median minimizes the a posteriori loss, that is, the strategy $\hat{u} = D(\mu)$ which satisfies

$$W_\mu\{u \leqslant \hat{u}\} = \tfrac{1}{2}.$$

As an example, let us take a mixed Poisson process (1.18), so W is described by a cumulative distribution F on R_+. Thus we are confronted with the same problem as in Example 1 of the second chapter, namely that of estimating ϱ itself, that is, the parameter $u(\vartheta\sigma) = \vartheta$. We assume that F is a Γ-distribution, partly because it is mathematically expedient, partly because many distributions observed can well be fitted to such a law:

$$dF(\vartheta) = \frac{\gamma^\beta \vartheta^{\beta-1} \exp(-\gamma\vartheta)}{\Gamma(\beta)}\, d\vartheta \quad \text{where} \quad \gamma, \beta > 0. \tag{3.6}$$

By (3.3) and (3.6):

$$F_\mu(d\vartheta) = C\vartheta^{n+\beta-1} \exp\big(-\vartheta(\gamma + \sigma(X))\big)$$

with some constant C which does not depend on ϑ, hence the a posteriori expectation (3.5) equals

$$\hat{u} = \frac{n+\beta}{\gamma + \sigma(X)}. \tag{3.7}$$

Recall that the uniformly best unbiased estimator for ϑ in the family $(P_{\vartheta\sigma})_{\vartheta<0}$ had been given by (2.13), namely,

$$\hat{\vartheta} = \frac{n}{\sigma(X)}.$$

The two estimators coincide, of course, asymptotically if $\sigma(X) \to \infty$ in the way described in Chapter 2, because we will then also have $n \to \infty$ almost surely.

Cases like the preceding one where (3.3) and (3.5) can be calculated explicitly, are rather rare, however. It is often easier to find a best strategy among the linear ones, in a sense to be defined, using $\mathcal{L}_2$-space methods. From now on we will employ exclusively the loss function (3.4) where u is a given real-valued parameter of ϱ.

We will need a few notations. Since we are now dealing with two random measures instead of one, we have to use two different notations to replace the previous ζ_f for $f \in \mathcal{H}$ as defined after the formula (1.3): for $\omega = (\varrho, \mu)$ we set

$$\tilde{\varrho}_f(\omega) = \varrho(f), \quad \tilde{\mu}_f(\omega) = \mu(f).$$

Consider the real Hilbert space $\mathcal{L}_2(P^W)$. We assume that W is a second order random measure. By the remark before (1.16) this is equivalent to either one of the following two statements: every $\tilde{\varrho}_f$ with $f \in \mathcal{H}$ belongs to $\mathcal{L}_2(P^W)$; every $\tilde{\mu}_f$ with $f \in \mathcal{H}$ belongs to $\mathcal{L}_2(P^W)$. Of course, it suffices to require this with $f \equiv 1$.

Let ν_W be the intensity measure of W. Then by (1.16):

$$\nu_W(f) = P^W \tilde{\mu}_f = P^W \tilde{\varrho}_f, \quad f \in \mathcal{H},$$

and

$$\begin{aligned} P^W(\tilde{\varrho}_f \tilde{\varrho}_g) &= \nu_W^{(2)}(f \otimes g), \\ P^W(\tilde{\mu}_f \tilde{\mu}_g) &= \nu_{P_W}^{(2)}(f \otimes g) = \nu_W^{(2)}(f \otimes g) + \nu_W(fg) \end{aligned} \tag{3.8}$$

for $f, g \in \mathcal{H}$; the respective covariances are obtained by replacing $\nu^{(2)}$ by γ.

For any subset $\mathcal{M} \subseteq \mathcal{L}_2(P^W)$ denote by $\mathcal{L}^W(\mathcal{M})$ the closed linear subspace of $\mathcal{L}_2(P^W)$ spanned by $\mathcal{M}$, and by $\mathrm{Proj}^W(v|\mathcal{M})$ the orthogonal projection of an element $v \in \mathcal{L}_2(P^W)$ onto $\mathcal{L}^W(\mathcal{M})$. By a *linear strategy* we mean an element of $\mathcal{L}^W(1, \{\tilde{\mu}_f : f \in \mathcal{H}\})$. This space is, of course, already spanned by the functions 1 and $\tilde{\mu}_A - \nu_W(A)$ with $A \in \mathcal{B}$. Note that a linear strategy is, in fact, a strategy, that is a function of $\tilde{\mu}$ only.

Suppose now that $u \in \mathcal{L}_2(W)$. By (3.1) this amounts to $u \circ \tilde{\varrho} \in \mathcal{L}_2(P^W)$, and we will use the letter u to denote the function $u \circ \tilde{\varrho}$ also. On account of (3.2) and (3.4), for any strategy D such that $D \circ \tilde{\mu} \in \mathcal{L}_2(P^W)$, the risk $K(W, D)$ equals the squared distance of u and $D \circ \tilde{\mu}$ in $\mathcal{L}_2(P^W)$. Therefore the best linear estimate of the value of u is given by

$$\hat{u} = \mathrm{Proj}^W\big(u \,|\, 1, \{\tilde{\mu}_A - \nu_W(A);\ A \in \mathcal{B}\}\big). \tag{3.9}$$

Our problem is to compute this, and in particular to investigate its dependence on W.

Since the elements $\tilde{\mu}_A - \nu_W(A)$ have the projection 0 on the space of all constants, and the projection of u on this space is $\bar{u} = W(u)$, we can write (3.9) in the form

$$\hat{u} = \bar{u} + \mathrm{Proj}^W\big(u-\bar{u}\,|\,\{\tilde{\mu}_A - \nu_W(A);\ A \in \mathscr{B}\}\big). \tag{3.10}$$

Next we make use of the following proposition which follows easily from (3.8) and the completeness of $\mathscr{L}_2(\nu_W)$.

PROPOSITION ([12], p. 116). *Let $v \in \mathscr{L}_2(P^W)$. Then $v \in \mathscr{L}^W\{\tilde{\mu}_A - \nu_W(A);\ A \in \mathscr{B}\}$ if and only if there exists a $\mathscr{B}$-measurable function f on X such that $\tilde{\mu}_f \in \mathscr{L}_2(P^W)$ and $v = \tilde{\mu}_f - \nu_W(f)$. This function f is unique* mod ν_W.

Essentially, the proposition says that the set of all $\tilde{\mu}_f - \nu_W(f)$ such that $\tilde{\mu}_f \in \mathscr{L}_2(P^W)$, is a closed subspace of $\mathscr{L}_2(P^W)$. Note that we cannot require f to be bounded, that is, in $\mathscr{H}$, but $\tilde{\mu}_f(\omega) = \mu(f)$ exists for P^W-almost all $\omega = (\varrho, \mu)$ because X is compact and therefore $\mu \in \mathscr{M}^e$ almost surely. By taking suitable monotone limits in (3.8) we see that the condition $\tilde{\mu}_f \in \mathscr{L}_2(P^W)$ is equivalent to "$\tilde{\varrho}_f \in \mathscr{L}_2(P^W)$ and $f \in \mathscr{L}_2(\nu_W)$", and (3.8) still holds for any two f and g such that $\tilde{\mu}_f, \tilde{\mu}_g \in \mathscr{L}_2(P^W)$.

By applying the proposition to the strategy (3.10) we find a function f such that $\tilde{\mu}_f \in \mathscr{L}_2(P^W)$ and

$$\hat{u} = \bar{u} + \tilde{\mu}_f - \nu_W(f). \tag{3.11}$$

The random variable $\tilde{\mu}_f - \nu_W(f)$ is unique in $\mathscr{L}_2(P^W)$; hence f is unique modν_W, and it is characterized by the fact that $u - \bar{u} - (\tilde{\mu}_f - \nu_W(f))$ is orthogonal to all $\tilde{\mu}_A - \nu_W(A)$ with $A \in \mathscr{B}$, or equivalently, to all $\tilde{\mu}_g - \nu_W(g)$ with $g \in \mathscr{H}$. This amounts to

$$\mathrm{cov}_W(u, \tilde{\mu}_g) = \mathrm{cov}_W(\tilde{\mu}_f, \tilde{\mu}_g) \tag{3.12}$$

for all $g \in \mathscr{H}$. It suffices to have this for all $g = 1_A$ with $A \in \mathscr{B}$, but on the other hand, it will then also hold for all g such that $\tilde{\mu}_g \in \mathscr{L}_2(P^W)$.

In order to transform (3.12) into something which might allow us to compute f, we remark that for any $\mathscr{B}$-measurable function g on X such that $\tilde{\mu}_g \in \mathscr{L}_2(P^W)$ and any random variable $u \in \mathscr{L}_2(P^W)$ which depends on $\tilde{\varrho}$ only, we have

$$P^W(u\tilde{\mu}_g) = P^W(u\tilde{\varrho}_g). \tag{3.13}$$

If $g \in \mathscr{H}$, this follows immediately from the definition of a Cox process, the equation $P_\varrho(\tilde{\mu}_g) = \varrho(g) = \tilde{\varrho}_g(\varrho)$ and the usual "conditional" reasoning:

$$P^W(u\tilde{\mu}_g) = P^W\big(P^W(u\tilde{\mu}_g|\tilde{\varrho})\big) = P^W(u\tilde{\varrho}_g);$$

the general case is then obtained again by taking monotone limits. By applying (3.13) to centered random variables we get

$$\mathrm{cov}_W(u, \tilde{\mu}_g) = \mathrm{cov}_W(u, \tilde{\varrho}_g).$$

For fixed u, the functional

$$\alpha_{W,u}(g) = \mathrm{cov}_W(u, \tilde{\varrho}_g), \tag{3.14}$$

defined for all g with $\tilde{\mu}_g \in \mathscr{L}_2(P^W)$, is a signed measure when restricted to $\mathscr{H}$; so we are justified in writing $\alpha_{W,u}(A) = \alpha_{W,u}(1_A)$ for $A \in \mathscr{B}$. Condition (3.12) which characterizes the function f appearing in the representation (3.11) of the best estimator $\hat{u}$ can now be restated as

$$\alpha_{W,u}(g) = \mathrm{cov}_W(\tilde{\mu}_f, \tilde{\mu}_g), \tag{3.15}$$

which, by (3.8), is the same as

$$\alpha_{W,u}(g) = \gamma_W(f \otimes g) + \nu_W(fg). \tag{3.16}$$

This holds for all g with $\tilde{\mu}_g \in \mathscr{L}_2(P^W)$, but it suffices to require it with $g = 1_A$, $A \in \mathscr{B}$.

By Pythagorean theorem, the risk run with $\hat{u}$, namely $P^W\big((u-\hat{u})^2\big)$, equals $\mathrm{var}_W(u) - \mathrm{var}_W(\tilde{\mu}_f)$; hence by (3.15) with $g = f$:

$$P^W\big((u-\hat{u})^2\big) = \mathrm{var}_W(u) - \alpha_{W,u}(f). \tag{3.17}$$

In principle, (3.16) allows us to find f and thus $\hat{u}$ if W is known. In fact, we see that only certain aspects of W enter into (3.16), and therefore a certain partial knowledge suffices: it is enough to know the number $\bar{u} = W(u)$ and the measures $\alpha_{W,u}, \nu_W$ and γ_W. In this sense we are in the situation mentioned above where we have a class $\mathscr{W}$ which consists of many, but not all, a priori laws: $\mathscr{W}$ is the class of all W with given, and known, $W(u)$, $\alpha_{W,u}$, ν_W and γ_W. As to the function u whose value for an observed realization $\omega = (\varrho, \mu)$ we want to estimate, it enters only via $W(u)$ and $\alpha_{W,u}$.

Before treating an example let us write (3.16), with $g = 1_A$, in a more explicit form:

$$\int_A \alpha_{W,u}(dy) = \int_{X\times A} f(x)\gamma_W(dx\,dy) + \int_A f(y)\nu_W(dy). \tag{3.18}$$

If we want to estimate the value $\varrho(h)$ where h is a fixed function, that is, if $u = \tilde{\varrho}_h$, then (3.16) becomes

$$\gamma_W(h\otimes g) = \gamma_W(f\otimes g) + \nu_W(fg).$$

In practice, however, only the case where W-almost all ϱ have a density with respect to a fixed measure $\sigma \in \mathscr{M}_+$ can be dealt with in a more explicit fashion without appealing to disintegrations of measures. We will assume that W-almost all ϱ can be written in the form

$$\varrho = w(\varrho, \cdot)\sigma \tag{3.19}$$

where $x \mapsto w(\varrho, x)$ is σ-integrable, and $\varrho \mapsto w(\varrho, x)$ is in $\mathscr{L}_2(W)$, so we get effectively a second order random measure. Consider the expectation function

$$\beta(y) = Ww(\cdot, y) = \int_{\mathscr{M}_+} w(\varrho, y)\,W(d\varrho)$$

and the covariance function

$$\varphi(x, y) = \mathrm{cov}_W\big(w(\cdot, x), w(\cdot, y)\big).$$

Then $\nu_W = \beta\sigma$, that is,

$$\nu_W(g) = \int_X \beta(y) g(y)\,\sigma(dy)$$

for every $g \in \mathscr{L}_1(\nu_W)$, and

$$\gamma_W(f \otimes g) = \int_X\int_X f(x) g(y) \varphi(x, y)\,\sigma(dx)\,\sigma(dy)$$

for every g such that $\tilde{\mu}_g \in \mathscr{L}_2(P^W)$. Therefore, (3.18) amounts to

$$\operatorname{cov}_W(u, w(\cdot, y)) = \int_X f(x)\varphi(x, y)\,\sigma(dx) + f(y)\beta(y) \tag{3.20}$$

for σ-almost all $y \in X$.

We are now going to treat the case of a "perturbed" mixed Poisson process. This means that w will be of the type

$$w(\varrho, x) = v(\varrho) + \varepsilon(\varrho, x) \tag{3.21}$$

where $v \in \mathscr{L}_2(W)$ and where we have, for every $x \in X$, $\varepsilon(\cdot, x) \in \mathscr{L}_2(W)$, $W\varepsilon(\cdot, x) = 0$ and v and $\varepsilon(\cdot, x)$ are uncorrelated. In this case, β is a constant, namely $\beta = Wv$, and

$$\varphi(x, y) = \operatorname{var}_W v + \operatorname{cov}_W(\varepsilon(\cdot, x), \varepsilon(\cdot, y)). \tag{3.22}$$

We are interested in two parameters: first, $u = v$, that is, we want to filter out the realization of the unperturbed mixed Poisson process, and secondly, for fixed $z \in X$, the value of the density w in the point z, that is, $u_z(\varrho) = w(\varrho, z)$. In the first case the left-hand side of (3.20) equals $\operatorname{var}_W v$, thus (3.20) becomes

$$\operatorname{var}_W v = \int_X f(x)\varphi(x, y)\,\sigma(dx) + f(y)\,Wv. \tag{3.23}$$

In the second case the left-hand side of (3.20) is equal to $\varphi(y, z)$, thus (3.20) becomes, with a function f_z in the place of f,

$$\varphi(y, z) = \int_X f_z(x)\varphi(x, y)\,\sigma(dx) + f_z(y)\,Wv. \tag{3.24}$$

Moreover, $\bar{u} = \bar{u}_z = Wv$, $\nu_W = (Wv)\sigma$. These relations and (3.22)–(3.24) show that in order to find the estimates u and u_z given by (3.11) in terms of f and f_z, respectively, we have to know the numbers Wv, $\operatorname{var}_W v$ and the covariance function $\operatorname{cov}_W(\varepsilon(\cdot, x), \varepsilon(\cdot, y))$.

Let us look first at the case of an "unperturbed" mixed Poisson process where $\varepsilon \equiv 0$ and where the parameters u and u_z coincide with v for every z. Then, by (3.22), we have $\varphi(x, y) = \operatorname{var}_W v$ for all x and y; hence both (3.23) and (3.24) give

$$\operatorname{var}_W v = \sigma(f)\operatorname{var}_W v + f(y)\,Wv$$

for all y which has the constant solution

$$f \equiv \frac{\operatorname{var}_W v}{\sigma(X)\operatorname{var}_W v + Wv}.$$

Upon setting

$$\beta = \frac{(Wv)^2}{\operatorname{var}_W v}, \qquad \gamma = \frac{Wv}{\operatorname{var}_W v},$$

the best linear strategy (3.11) takes the form

$$\hat{u}(\mu) = \frac{\mu(X) + \beta}{\sigma(X) + \gamma}$$

which coincides with the Bayes strategy (3.7) for the a priori law (3.6).

Next we look at processes on the positive half-line R_+. In many practical situations, perturbations in regions far apart may be considered independent, and so it might be reasonable to study a discrete analogon to the white noise. Let c_n with $n = 0, 1, \ldots$ be a sequence of independent and identically distributed random variables with finite variance $b^2 = \operatorname{var}_W c_n$ and expectation 0, and set $\varepsilon(t) = c_{[t]}$ for $0 \leqslant t$ where $[t]$ stands as usual for the greatest integer less or equal to t. Then

$$\varphi(x, y) = \begin{cases} \operatorname{var}_W v + b^2 & \text{if} \quad [x] = [y], \\ \operatorname{var}_W v & \text{if} \quad [x] \neq [y]. \end{cases}$$

We use the abbreviations

$$\bar{v} = Wv, \qquad a^2 = \operatorname{var}_W v$$

and consider everything within a fixed interval $X = [0, s]$ where $s > 0$. For σ we take the Lebesgue measure in X. Then (3.23) still has a constant solution, namely

$$f \equiv \frac{a^2}{sa^2 + \bar{v} + b^2},$$

which gives the best linear strategy

$$\hat{u}(\mu) = \frac{\mu(X) + \bar{v}^2/a^2 + \bar{v}(b/a)^2}{s + \bar{v}/a^2 + (b/a)^2}.$$

Note that, for $b/a \to \infty$, this converges to $\bar{v}$: thus for large perturbations relative to a^2 we do not take the observation into account at all. For $s \to \infty$, we get, of course, almost surely the estimator (2.13) again.

A solution of (3.24) can be obtained in the following form: set

$$p = \frac{\bar{v}a^2}{(\bar{v} + b^2)(sa^2 + \bar{v} + b^2)}, \qquad q = \frac{b^2}{\bar{v} + b^2}.$$

Then

$$f_z(x) = \begin{cases} p & \text{if} \quad [x] \neq [z], \\ p + q & \text{if} \quad [x] = [z], \end{cases}$$

and

$$\hat{u}_z(\mu) = \frac{\bar{v}^2}{sa^2 + \bar{v} + b^2} + p\mu([0, s]) + q\mu([[z], [z+1][).$$

For $s \to \infty$, this estimator behaves almost surely like

$$\frac{\mu([0, s])}{s} \cdot \frac{\bar{v}}{\bar{v} + b^2} + \mu([[z], [z+1][)\frac{b^2}{\bar{v} + b^2}.$$

Thus the two estimators of the type (2.13) based on the observation in the entire interval $[0, 1]$ and in $[[z], [z+1][$, respectively, are combined in the proportion $\bar{v} : b^2$.

References

[0] O. O. Aalen, *Statistical inference for a family of counting processes*, Inst. Math. Statist., Univ. of Copenhagen, 1976.

[1] P. Brémaud, *La méthode des semi-martingales en filtrage lorsque l'observation est un processus ponctuel marqué*, Sémin. Probab. X, Univ. Strasbourg 1974–75, Lecture Notes Math. 511 (1976), pp. 3–21.

[2] —, *Estimation de l'état d'une file d'attente et du temps de panne d'une machine par la méthode des semi-martingales*, Advances Appl. Probab. 7 (1975), pp. 845–863.

[3] —, *Prédiction, filtrage et déctection pour une observation mixte: la méthode de la probabilité de référence*, to appear in: J. Appl. Math. Optim.

[4] — and J. Jacod, *Processus pontuels et martingales: présentation des résultats récents sur la modélisation et le filtrage*, to appear in: Advances Appl. Probab.

[5] D. Brillinger, *The spectral analysis of stationary interval functions*, pp. 483–513 in: Proc. Sixth Berkeley Symp. Math. Stat. Probab., vol. I. University of California Press, Berkeley 1972.

[6] —, *Statistical inference for stationary point processes*, pp. 55–99 in: *Stochastic processes and related topics*, vol. I. Ed. Madan Lal Puri. Academic Press, New York-London 1975.

[7] —, *The identification of point process systems*, Ann. Probab. 3 (1975), pp. 909–929.

[8] M. Brown, *Statistical analysis of non-homogeneous Poisson processes*, pp. 67–89 in: *Stochastic point processes*, Ed. P. A. W. Lewis, Wiley, New York-London 1972.

[9] D. R. Cox, and P. A. W. Lewis, *The statistical analysis of series of events*, Methuen, London 1966.

[10] A. Fellous, *Estimation dans les modèles de processus de Cox isotropes de droites et d'axes du plan*, Thèse 3e cycle, Université Paris V, 1976.

[11] Th. Ferguson, *Mathematical statistics: a decision theoretic approach*, Academic Press, New York-London 1967.

[12] J. Grandell, *Doubly stochastic Poisson processes*, Lecture Notes Math. 529, Springer, Berlin-Heidelberg-New York 1976.

[13] J. Jacod, *Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales*, Z. Wahrscheinlichkeitsth. verw. Gebiete 31 (1975), pp. 235–253.

[14] O. Kallenberg, Random measures, Akademie-Verlag Berlin, 1975.

[15] J. Kerstan, K. Matthes, and J. Mecke, *Unbegrenzt teilbare Punktprozesse*, Akademie-Verlag, Berlin 1974.

[16] K. Krickeberg, *The Cox process*, Sympos. Math. 9 (1972), pp. 151–167.

[17] —, *Statistical inference of hyperplane processes*, pp. 335–388 in: *Conference on probability theory and mathematical statistics*, Vilnius 1973.

[18] —, *Fundamentos del análisis estadístico de procesos puntuales*, CIENES, Santiago de Chile 1973.

[19] —, *Lectures on point processes* (in vietnamese), Institute of Mathematics of Vietnam, Hanoi 1976.

[20] —, *An alternative approach to Glivenko–Cantelli theorems*; in: *Empirical distributions and processes*, Lecture Notes Math. 566, Springer, Berlin-Heidelberg-New York 1976, pp. 57–67.

[21] E. L. Lehmann, *Testing statistical hypothesis*, Wiley, New York 1959.

[22] F. Liese, *Eine informationstheoretische Bedingung für die Äquivalenz unbegrenzt teilbarer Punktprozesse*, Math. Nachr. 70 (1975), pp. 183–196.

[23] O. Macchi, *The coincidence approach to stochastic point processes*, Advances Appl. Probab. 7 (1975), pp. 83–122.

[24] K. R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, New York-London 1967.

[25] F. Streit, *Zufälligkeitstests and ihre Anwendung in der Versicherungsmathematik*, Mitt. Verein schweiz. Versicherungsmath. 74 (1974), pp. 173–192.

[26] Mori Toshio, *Ergodicity and identifiability for random translations of arbitrary point processes*, J. Appl. Probab. 12 (1975), pp. 734–743.

[27] *Stochastic point processes*, Ed. P. A. W. Lewis, Wiley, New York-London 1972.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December, 18 1976