

## ON ESTIMATION OF POISSON INTENSITY FUNCTIONS

ROELOF HELMERS<sup>1</sup> AND RIČARDAS ZITIKIS<sup>2</sup>

<sup>1</sup>*Centre for Mathematics and Computer Science, Kruislaan 413,  
1098 SJ Amsterdam, The Netherlands*

<sup>2</sup>*Laboratory for Research in Statistics and Probability, Carleton University,  
1125 Colonel By Drive, Ottawa, Canada K1S 5B6*

(Received June 26, 1996; revised October 13, 1997)

**Abstract.** Under the presence of only one realization, we consider a computationally simple algorithm for estimating the intensity function of a Poisson process with exponential quadratic and cyclic of fixed frequency trends. We argue that the algorithm can successfully be used to estimate any Poisson intensity function provided that it has a parametric form.

*Key words and phrases:* Poisson process, point process, linear Poisson process, modulated Poisson process.

### 1. Introduction

Let  $X$  be a Poisson point process on the Borel measurable space  $(\mathbf{R}^d, \mathbf{B}^d)$ ,  $d \in \mathbf{N}$ , with unknown locally integrable intensity function

$$(1.1) \quad \lambda \in L_{loc}^1(\mathbf{R}^d),$$

that is,  $\lambda$  is an integrable function on every compact subset of  $\mathbf{R}^d$ . For mathematical foundations of point processes theory we refer, for example, to Daley and Vere-Jones (1972, 1988), Ripley (1976), Kallenberg (1983), Kingman (1993), Reiss (1993).

Motivated by Helmers (1995), we assume throughout that only one realization of the process  $X$  is available. Furthermore, we assume that  $X$  is observed only in a bounded Borel set (called window)  $W \subset \mathbf{R}^d$  that depends on a parameter  $N$  and expands when  $N \rightarrow \infty$ . We also assume throughout that

$$(1.2) \quad \int_{\mathbf{R}^d} \lambda(x) dx (= \mathbf{E}X(\mathbf{R}^d)) = \infty$$

(cf., for example, Subsection 6.3 on pp. 136–137 of Rathbun and Cressie (1994) for a discussion concerning the necessity of assumption (1.2) in this context).

Using the Maximum Likelihood Estimation (MLE) method, Krickeberg (1982), Kutoyants (1984), Rathbun and Cressie (1994), among others, made profound contributions in estimating the intensity function  $\lambda$  under the presence of only one realization of  $X$ .

In this paper we consider an alternative route for constructing estimators for  $\lambda$  under the presence of only one realization of  $X$ . To illustrate the main idea of our investigation, as well as to justify it, we use the following example.

Let  $X_0$  be a Poisson process on  $[0, \infty)$  with the intensity function

$$\lambda_0(s) := \exp\{\alpha + \beta s + \gamma s^2 + K_1 \sin(\omega_0 s) + K_2 \cos(\omega_0 s)\}, \quad s \geq 0,$$

where  $\alpha, \beta, \gamma, K_1, K_2$  are unknown parameters, and  $\omega_0$  is a known 'frequency'. This model is of importance in diverse fields of applied mathematics such as, for example, geophysics, optics, reliability, biostatistics, meteorology (cf., for example, Lewis (1970, 1972), Cox (1972), Cox and Lewis (1978), Cox and Isham (1980), Vere-Jones and Ozaki (1982), Diggle (1983), Karr (1986), Ogata and Katsura (1986), Cressie (1991), Ripley (1991), Zheng and Vere-Jones (1994), Vere-Jones (1995), as well as references therein).

In particular, Lewis (1972) notes on pp. 42–43 in Section 5.4 therein that researchers encounter mathematical difficulties when constructing the MLE estimator for  $\lambda_0(s)$ . Indeed, as far as we are aware of, there are no explicit mathematical expressions for the MLE-based estimators of the parameters  $\alpha, \beta, \gamma, K_1$  and  $K_2$ , whence for the estimator of  $\lambda_0(s)$  either. We recognize, of course, that in practical situations this fact may not be a real problem since, using the MLE method, the parameters can nevertheless be estimated numerically. However, when investigating more complex intensity functions (cf., for example, Cox (1972)), it becomes difficult to obtain even numerical estimators because of, for example, stability problems in solutions of the MLE-based algebraic equations. When intensity functions are defined on  $\mathbf{R}^d$  for  $d \geq 2$ , then the MLE method becomes even more difficult to implement (cf., for example, Helmers (1995)). On the other hand, we also note that when constructing confidence intervals for  $\lambda_0(s)$ , or even confidence bands for  $\lambda_0$  over a certain region of  $\mathbf{R}^d$ , numerical estimators may not be satisfactory from the probabilistic point of view, and thus explicit empirical estimators may be required in such situations.

When facing all these difficulties with the MLE method, one may naturally ask whether it is really necessary to estimate unknown parameters (say,  $\alpha, \beta, \gamma, K_1$  and  $K_2$  of  $\lambda_0(s)$ ) when the original goal is to estimate the intensity function (say,  $\lambda_0(s)$ ) and not the parameters themselves. (The latter problem, of course, implies the first one but is definitely much more difficult to tackle.) This reasoning has motivated us to have a closer look at the following idea:

Let us choose a countable number of distinct points  $s_1, \dots, s_N, \dots \in \mathbf{R}^d$  such that

$$(1.3) \quad s_1, \dots, s_N \in \text{the interior of the window } W,$$

and construct a sequence of functions  $f_N : [0, \infty)^N \rightarrow [0, \infty)$  such that

$$(1.4) \quad f_N(\lambda(s_1), \dots, \lambda(s_N)) \rightarrow \lambda(s), \quad N \rightarrow \infty.$$

(We note that  $s_1, \dots, s_N, \dots \in \mathbf{R}^d$  are chosen points and not points of the point process  $X$ , and that the functions  $f_N$  can depend on points  $s, s_1, \dots, s_N$ .) Furthermore, let  $B_1(0) \subset \mathbf{R}^d$  denote the unit ball centred at 0, and let

$$(1.5) \quad A_n := (hB_1(0) + s_n) \cap W,$$

$n = 1, \dots, N$ , where  $h > 0$  is a certain parameter that may depend on  $n$  and/or  $N$  and converges to 0 when  $n$  and/or  $N$  converges to  $\infty$ . With these notations, and with ' $\approx$ ' standing for 'loosely speaking, asymptotically behaves like', we get

$$(1.6) \quad \lambda(s_n) \approx \frac{1}{\text{Vol}(A_n)} \int_{A_n} \lambda(x) dx = \frac{1}{\text{Vol}(A_n)} \mathbf{E}X(A_n) \approx \frac{1}{\text{Vol}(A_n)} X(A_n),$$

when  $h > 0$  converges to 0. Denote

$$\xi_n(h) := \frac{1}{\text{Vol}(A_n)} X(A_n).$$

Consequently, replacing  $\lambda(s_n)$  by  $\xi_n(h)$  in (1.4), we get the empirical estimator

$$(1.7) \quad \hat{\lambda}(s) := f_N(\xi_1(h_1), \dots, \xi_N(h_N))$$

that converges (in a certain sense and under certain assumptions) to  $\lambda(s)$  when  $h \rightarrow 0$  and  $N \rightarrow \infty$ . In this paper, however, for the sake of simplicity we restrict ourselves to the convergence in the mean of order 2 only, i.e. we shall aim at assumptions under which the statement

$$(1.8) \quad \mathbf{E}\{\hat{\lambda}(s) - \lambda(s)\}^2 \rightarrow 0$$

holds true for any fixed  $s$ , and when  $h \rightarrow 0$  and  $N \rightarrow \infty$ .

The intuition behind the idea given in (1.3)–(1.7) is fairly clear: Since the point process  $X$  always places only a finite number of points into any compact subset of  $\mathbf{R}^d$ , due to (1.1), in order to obtain a consistent estimator of  $\lambda(s)$  one must accumulate the necessary empirical information about  $\lambda(s)$  from different places of the window  $W$  where  $X$  is observed. (This implies, in particular, that the window  $W$  should necessarily depend on  $N$ , and  $\text{Vol}(W) \rightarrow \infty$  when  $N \rightarrow \infty$ ; on p. 480 of Cressie (1991) such a framework is called 'increasing domain asymptotics'.) At this point one naturally faces the necessity of making assumptions either on the point process  $X$  itself (like stationarity, ergodicity, etc.) or on the corresponding intensity function  $\lambda$  (like some knowledge of its shape, that is very clearly reflected in parametric models like that of  $\lambda_0(s)$  given above). Due to the practical problem we originally faced (cf. Helmers (1995)), we have found it more natural to start our work with assumptions on the intensity function  $\lambda$ , rather than making structural assumptions (like stationarity, ergodicity, etc.) on the point process  $X$ . And this our inclination is, of course, reflected in the idea (1.3)–(1.7) described above.

## 2. Algorithm

When investigating a number of Poisson intensity functions using the idea described in (1.3)–(1.7), we singled out several steps that were common in each example, and we shall now write them down. (An application of these steps in the case of Poisson process  $X_0$  will be discussed in detail in next Section 3.)

*Step 2.1.* Let  $N, \kappa \in \mathbf{N}$ , and let us construct:

- i) points  $s_1, s_2, \dots \in \mathbf{R}^d$ ;
- ii) bounded Borel sets  $W_0 := W_{0,N} \subset \mathbf{R}^d$  such that  $W_0$  expands when  $N \uparrow \infty$ ;
- iii) a function  $L : (0, \infty)^\kappa \times (\mathbf{R}^d)^{\kappa+1} \rightarrow [0, \infty)$ ;
- iv) subsequences  $(k_j(n), n \in \mathbf{N})$ ,  $j = 1, \dots, \kappa$ , of the sequence of natural numbers, such that

$$(2.1) \quad s_{k_j(n)} \in W_0$$

for all  $n = 1, \dots, N$  and  $j = 1, \dots, \kappa$ , and the following statement

$$(2.2) \quad \frac{1}{N} \sum_{n=1}^N L(\lambda(s_{k_1(n)}), \dots, \lambda(s_{k_\kappa(n)}), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s) \rightarrow \lambda(s)$$

holds true when  $N \rightarrow \infty$ .

In order to construct a function  $L$  required in (2.2), one naturally has to have some information about the shape of intensity function  $\lambda$ . Parametric models (like that described by  $\lambda_0$ ) are particularly convenient in such instances, but non-parametric intensity functions can also be successfully treated. For example, if we knew that a (non-parametric or parametric) intensity function  $\lambda$  were, say, purely cyclic with fixed frequency, then, having appropriately chosen points  $s_n$ , we would have  $\lambda(s_n) = \lambda(s)$  and thus  $N^{-1} \sum_{n=1}^N \lambda(s_n) = \lambda(s)$ . This immediately shows, for example, that  $L(x) = x$  is a good candidate for  $L$  required in (2.2). However, we have to note immediately that even a slightest discrepancy from the purely cyclic case may result in far less trivial Step 2.1.

*Step 2.2.* Let us construct  $\kappa$  sequences  $h_1(n) \downarrow 0, \dots, h_\kappa(n) \downarrow 0$ , when  $n \rightarrow \infty$ , of positive real numbers such that the statement

$$(2.3) \quad \frac{1}{N} \sum_{n=1}^N \{L(\mathbf{E}\xi_{k_1(n)}(h_1(n)), \dots, \mathbf{E}\xi_{k_\kappa(n)}(h_\kappa(n)), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s) - L(\lambda(s_{k_1(n)}), \dots, \lambda(s_{k_\kappa(n)}), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s)\} \rightarrow 0$$

holds true when  $N \rightarrow \infty$ .

In order to give some immediate insight into Step 2.2, we note that if  $s_p$  is a Lebesgue point of the (intensity) function  $\lambda$ , then

$$(2.4) \quad \mathbf{E}\xi_p(h) \rightarrow \lambda(s_p)$$

when  $h \downarrow 0$ . (Let us also note in passing that, since  $\lambda \in L^1_{loc}(\mathbf{R}^d)$  by assumption, the set of all Lebesgue points of  $\lambda$  is dense in  $\mathbf{R}^d$ .) In view of (2.4), statement (2.3) is a kind of continuity assumption on  $L$ , though continuity (in the usual sense) of  $L$  does not in general imply (2.3). Nevertheless, when investigating different examples we have not found Step 2.2 to be difficult.

*Step 2.3.* Let us construct a function  $L^* : [0, \infty)^\kappa \times (\mathbf{R}^d)^{\kappa+1} \rightarrow [0, \infty)$  such that the statement

$$(2.5) \quad \frac{1}{N} \sum_{n=1}^N \{ \mathbf{E}L^*(\xi_{k_1(n)}(h_1(n)), \dots, \xi_{k_\kappa(n)}(h_\kappa(n)), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s) \\ - L(\mathbf{E}\xi_{k_1(n)}(h_1(n)), \dots, \mathbf{E}\xi_{k_\kappa(n)}(h_\kappa(n)), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s) \} \rightarrow 0$$

holds true when  $N \rightarrow \infty$ .

If we deleted the asterisk  $*$  from (2.5), then, loosely speaking, statement (2.5) would mean the commutative relation  $\mathbf{E}L = L\mathbf{E}$ . In general, however, we cannot use  $L$  instead of  $L^*$  because of the fact that the random variable  $\xi_{k_j(n)}(h_j(n))$  can take value zero, whereas the function  $L$  may not be defined on the set  $[0, \infty)^\kappa \times (\mathbf{R}^d)^{\kappa+1}$  (cf. part iii) of Step 2.1). We believe, however, that in most cases it will suffice to take  $L^* := L^+$ , where

$$L^+(x_1, \dots, x_\kappa, y_1, \dots, y_\kappa, z) \\ := \begin{cases} L(x_1, \dots, x_\kappa, y_1, \dots, y_\kappa, z) & \text{if } x_1 > 0, \dots, x_\kappa > 0, \\ 0, & \text{otherwise.} \end{cases}$$

With the function  $L^*$  introduced in Step 2.3 we are now able to define the (general) estimator  $\hat{\lambda}(s)$  of  $\lambda(s)$  as follows:

$$(2.6) \quad \hat{\lambda}(s) := \frac{1}{N} \sum_{n=1}^N L^*(\xi_{k_1(n)}(h_1(n)), \dots, \xi_{k_\kappa(n)}(h_\kappa(n)), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s).$$

In view of Steps 2.1–2.3, it is trivial to see that if the Poisson point process  $X$  is observed in the window

$$(2.7) \quad W := \{x \in \mathbf{R}^d : \|W_0 - x\| \leq h_1(N) \vee \dots \vee h_\kappa(N)\},$$

and

$$(2.8) \quad \text{Var } \hat{\lambda}(s) \rightarrow 0,$$

then the statement

$$(2.9) \quad E\{\hat{\lambda}(s) - \lambda(s)\}^2 \rightarrow 0$$

holds true when  $N \rightarrow \infty$ . We have thus arrived at a desired result (cf. (1.3)–(1.8)).

Let us note in passing that when investigating optimality questions related to the estimator  $\hat{\lambda}(s)$  of (2.6), one may find it necessary to use the following more general considerations

$$\begin{aligned} \lambda(s_n) &\approx \left\{ \int_W K\left(\frac{x-s_n}{h}\right) dx \right\}^{-1} \int_W K\left(\frac{x-s_n}{h}\right) \lambda(x) dx \\ &= \left\{ \int_W K\left(\frac{x-s_n}{h}\right) dx \right\}^{-1} \int_W K\left(\frac{x-s_n}{h}\right) (EX)(dx) \\ &\approx \left\{ \int_W K\left(\frac{x-s_n}{h}\right) dx \right\}^{-1} \int_W K\left(\frac{x-s_n}{h}\right) X(dx) \\ &=: \rho_n(h) \end{aligned}$$

instead of those given in (1.6), where  $K : \mathbf{R}^d \rightarrow \mathbf{R}$  is a certain function. It is easy to see that when  $K$  is the indicator of the unit ball  $B_1(0)$ , then  $\rho_n(h) = \xi_n(h)$ . Without going into a deeper and more sophisticated discussion related to this generalization, at this stage of investigation we only refer to the monographs of Devroye and Györfi (1985) and Rosenblatt (1991), as well as to references therein, for necessary probabilistic and statistical details concerning the subject.

The verification of assumption (2.8) can sometimes be rather involved. However, if the following assumption is satisfied, then the task becomes easier.

ASSUMPTION 2.1. Let the sets  $\{k_j(n) : n \in \mathbf{N}\}$ ,  $j = 1, \dots, \kappa$ , be pairwise disjoint, and let

$$\|s_{k_j(m)} - s_{k_\ell(n)}\| > h_j(m) + h_\ell(n)$$

for all  $j, \ell \in \{1, \dots, \kappa\}$  and  $m, n \in \{1, \dots, N\}$  such that  $k_j(m) \neq k_\ell(n)$ .

Indeed, under Assumption 2.1 we get that (2.8) holds true provided that

$$(2.10) \quad \frac{1}{N^2} \sum_{n=1}^N \text{Var} L^*(\xi_{k_1(n)}(h_1(n)), \dots, \xi_{k_\kappa(n)}(h_\kappa(n)), s_{k_1(n)}, \dots, s_{k_\kappa(n)}, s) \rightarrow 0,$$

when  $N \rightarrow \infty$ . Statement (2.10) is usually easy to prove. For example, this goal can essentially be achieved following the lines of the proof of statement (2.2).

As an illustration of the algorithm described above, we now formulate following Proposition 2.1 concerning the Poisson process  $X_0$ .

PROPOSITION 2.1. Let  $X_0$  be Poisson point process on  $[0, \infty)$  with the intensity function  $\lambda_0$ , where  $\omega_0 > 0$  is a fixed and apriori known frequency, and  $\alpha$ ,

$\beta, \gamma, K_1$  and  $K_2$  are unknown parameters from an apriori fixed compact interval of  $\mathbf{R}$ . Furthermore, let

$$(2.11) \quad s_n := s + 2\pi n/\omega_0,$$

and

$$(2.12) \quad \begin{cases} k_1(n) := 16n^4 - 8, \\ k_2(n) := 4n^2 - 2, \\ k_3(n) := 8n^3 - 4, \\ k_4(n) := 2n - 1; \end{cases}$$

$$(2.13) \quad \begin{cases} h_1(n) := n^{-5}, \\ h_2(n) := n^{-3}, \\ h_3(n) := n^{-4}, \\ h_4(n) := n^{-2}. \end{cases}$$

Assume that the process  $X$  is observed in the window

$$(2.14) \quad W := s + [-h_1(N) \vee \dots \vee h_4(N), 2\pi(16N^4 - 8)/\omega_0 + h_1(N) \vee \dots \vee h_4(N)].$$

If the estimator  $\hat{\lambda}_0(s)$  is  $\hat{\lambda}(s)$  of (2.6) with  $\kappa = 4, L^* = L^+$ , and

$$(2.15) \quad L(x_1, \dots, x_4, y_1, \dots, y_4, z) := \{x_1^{-y_2/y_1^2} x_2\}^{(z-y_4)/y_2} x_3^{(z^2-y_4^2)/y_3^2} x_4,$$

then the statement

$$(2.16) \quad E\{\hat{\lambda}_0(s) - \lambda_0(s)\}^2 \rightarrow 0$$

holds true when  $N \rightarrow \infty$ .

### 3. Proof of Proposition 2.1

3.0 On the intuition behind the construction of  $L$  in (2.15) (cf. also Section 4 below)

A possible way towards (2.15) starts with the identity

$$(3.1) \quad \Lambda_m := \exp\{-\beta(s_m - s) - \gamma(s_m^2 - s^2)\} \lambda_0(s_m) = \lambda_0(s).$$

Since the parameters  $\beta$  and  $\gamma$  are unknown, the right-hand side of (3.1) cannot be considered as a definition of the function  $L$  required in Step 2.1, though it satisfies the relation

$$\frac{1}{N} \sum_{m=1}^N \Lambda_m = \lambda_0(s).$$

Therefore, we shall now 'eliminate' the unknown parameters  $\beta$  and  $\gamma$  from the expression

$$\exp\{-\beta(s_m - s) - \gamma(s_m^2 - s^2)\}$$

in (3.1), and this will ultimately lead us to a desired function  $L$ . Let us first 'eliminate'  $\gamma$ . We have that

$$(3.2) \quad \begin{aligned} \lambda_0(s_n)^{s^2/s_n^2} &= \exp\{[\alpha + \beta s_n + K_1 \sin(\omega_0 s_n) + K_2 \cos(\omega_0 s_n)]s^2/s_n^2\} \\ &\quad \cdot \exp\{\gamma s^2\} \\ &\rightarrow \exp\{\gamma s^2\}, \quad n \rightarrow \infty. \end{aligned}$$

If we now use (3.2) in (3.1), we shall get the following approximation 'without'  $\gamma$ :

$$(3.3) \quad \exp\{-\beta(s_m - s)\} \lambda_0(s_n)^{(s^2 - s_m^2)/s_n^2} \lambda_0(s_m) \approx \lambda_0(s)$$

when  $n$  is large. Let us now 'eliminate'  $\beta$  from (3.3). Before doing it we note that if  $\gamma$  in the definition of  $\lambda_0(s)$  were equal to 0, then the parameter  $\beta$  would be easily eliminated as follows:

$$(3.4) \quad \begin{aligned} &\lambda_0(s_\ell)^{(s - s_m)/s_\ell} \\ &= \exp\left\{[\alpha + \gamma |_{\gamma=0} s_\ell^2 + K_1 \sin(\omega_0 s_\ell) + K_2 \cos(\omega_0 s_\ell)] \frac{s - s_m}{s_\ell}\right\} \\ &\rightarrow \exp\{-\beta(s_m - s)\}, \quad \ell \rightarrow \infty. \end{aligned}$$

However, we do not know whether  $\gamma$  in  $\lambda_0$  is zero or not. Therefore, the elimination of  $\beta$  becomes more involved. We do it as follows: Using the idea given in (3.2), we first eliminate  $\gamma$  from  $\lambda_0(s)$ . Then, having no explicit  $\gamma$  like in (3.3) (which actually amounts to the case  $\gamma = 0$ ), we use the idea given in (3.4) to eliminate the unknown  $\beta$ . In more detail, after these two steps we get

$$(3.5) \quad \{\lambda_0(s_k)^{-s_\ell^2/s_k^2} \lambda_0(s_\ell)\}^{(s - s_m)/s_\ell} \approx \exp\{-\beta(s_m - s)\}$$

when both  $k$  and  $\ell$  are large. Now, using (3.5) in (3.3), we get the approximation

$$(3.6) \quad \{\lambda_0(s_k)^{-s_\ell^2/s_k^2} \lambda_0(s_\ell)\}^{(s - s_m)/s_\ell} \lambda_0(s_n)^{(s^2 - s_m^2)/s_n^2} \lambda_0(s_m) \approx \lambda_0(s)$$

when  $k$ ,  $\ell$  and  $n$  are large. We may now consider the right-hand side of (3.6) as a definition of the function  $L$  needed in (2.2) of Step 2.1. However, we have not yet found out how to choose the subsequences

$$\begin{aligned} k_1(n) &:= k, \\ k_2(n) &:= \ell, \\ k_3(n) &:= n, \\ k_4(n) &:= m \end{aligned}$$

in order to have statement (2.2) satisfied. This will be done in our next Subsection 3.1 below.  $\square$



3.1 Verification of statement (2.2)

Let

$$(3.7) \quad L_n := L(\lambda(s_{k_1(n)}), \dots, \lambda(s_{k_4(n)}), s_{k_1(n)}, \dots, s_{k_4(n)}, s)$$

for notational simplicity. Then the desired statement (2.2) follows from

$$\frac{1}{N} \sum_{n=1}^N (\exp\{\log L_n - \log \lambda(s)\} - 1) \rightarrow 0,$$

which, in turn, is a consequence of the statement

$$(3.8) \quad \frac{1}{N} \sum_{n=1}^N |\log L_n - \log \lambda(s)| \exp\{|\log L_n - \log \lambda(s)|\} \rightarrow 0,$$

when  $N \rightarrow \infty$ . One can easily check that, with the notations  $z := s\omega_0/(2\pi)$ ,  $c_0 := 2\pi/\omega_0$ , and  $C(s) : K_1 \sin(\omega_0 s) + K_2 \cos(\omega_0 s) \equiv$  the cyclic part of the intensity function  $\lambda_0(s)$ , we have the following representation:

$$(3.9) \quad \begin{aligned} \log L_n = & -\frac{k_4(n)}{z + k_2(n)} \{\alpha + c_0\beta(z + k_2(n)) + c_0^2\gamma(z + k_2(n))^2 + C(s)\} \\ & + \frac{k_4(n)z + k_4(n)k_2(n)}{z + k_1(n)} \\ & \cdot \{\alpha + c_0\beta(z + k_1(n)) + c_0^2\gamma(z + k_1(n))^2 + C(s)\} \\ & - \frac{k_4^2(n)}{(z + k_3(n))^2} \{\alpha + c_0\beta(z + k_3(n)) + c_0^2\gamma(z + k_3(n))^2 + C(s)\} \\ & - \frac{2zk_4(n)}{(z + k_3(n))^2} \{\alpha + c_0\beta(z + k_3(n)) + c_0^2\gamma(z + k_3(n))^2 + C(s)\} \\ & + \log \lambda(s) + c_0\beta k_4(n) + 2c_0^2\gamma z k_4(n) + c_0^2\gamma k_4^2(n). \end{aligned}$$

Under the assumptions of Proposition 2.1 on the parameters, representation (3.9) immediately implies the following estimate of the distance between  $\log L_n$  and  $\log \lambda(s)$ :

$$(3.10) \quad |\log L_n - \log \lambda_0(s)| \leq c\epsilon(n) := c \max \left\{ \frac{k_4(n)}{k_2(n)}, \frac{k_4(n)k_2(n)}{k_1(n)}, \frac{k_4^2(n)}{k_3(n)} \right\},$$

where  $c := c(\omega_0, \alpha, \beta, \gamma, K_1, K_2)$  is a constant. We have chosen the sequences  $k_1(n), \dots, k_4(n)$  in such a way that  $\epsilon(n) \leq 1$  holds true. The latter fact implies, in particular, that (3.8) follows from the statement

$$(3.11) \quad \frac{1}{N} \sum_{n=1}^N \epsilon(n) \rightarrow 0, \quad N \rightarrow \infty,$$

the validity of which will soon be assured by a special choice of the sequences  $k_1(n), \dots, k_4(n)$ . Namely, let, for a moment, the sequence  $k_4(n)$  be arbitrary,  $k_2(n) \approx k_4(n)A(n)$  with some  $A(n) \uparrow \infty$  when  $n \rightarrow \infty$ ,  $k_1(n) \approx k_4(n)k_2(n)A(n) \approx (k_4(n)A(n))^2$ , and  $k_3(n) \approx k_4^2(n)A(n)$ . Taking now, for example,  $k_4(n) \approx n$ , we get the following table:

$$(3.12) \quad \begin{cases} k_1(n) \approx n^2 A(n)^2, \\ k_2(n) \approx n A(n), \\ k_3(n) \approx n^2 A(n), \\ k_4(n) \approx n. \end{cases}$$

We now see that statement (3.11) holds true provided that

$$(3.13) \quad \frac{1}{N} \sum_{n=1}^N \frac{1}{A(n)} \rightarrow 0$$

when  $N \rightarrow \infty$ . Taking  $A(n) \approx n$ , we see that statement (3.13) holds true, and so does (3.11). Tables (3.12) and (2.12), as it is easy to see, are asymptotically equivalent. We have chosen (2.12) due to its simplicity and, mainly, to have Assumption 2.1 automatically satisfied (this problem will be discussed in Subsection 3.4 below). It is also important to notice from the above calculations that there is room for more accurate choices of sequences  $k_1(n), \dots, k_4(n)$  and points  $s_{k_j(n)}$ , especially if some additional information about the parameters  $\alpha, \beta, \gamma, K_1, K_2$  becomes available. This implies, in particular, that the window

$$W_0 := s + [0, 2\pi(16N^4 - 8)/\omega_0]$$

—that covers all the points  $s_{k_j(n)}$ ,  $n = 1, \dots, N$ ,  $j = 1, \dots, 4$ , defined by (2.11) and (2.12)—can be made smaller. Consequently, the window

$$W = \{x \in \mathbf{R} : |W_0 - x| \leq h_1(N) \vee \dots \vee h_4(N)\}$$

defined in (2.14) will in turn become smaller as well.  $\square$

### 3.2 Verification of statement (2.3)

In order to check (2.3), we firstly note that

$$(3.14) \quad \mathbf{E}\xi_{k_j(n)}(h_j(n)) = \lambda_0(s_{k_j(n)})\{1 + \theta h_j(n)s_{k_j(n)}\},$$

where  $|\theta| \leq c$  and  $c < \infty$  is a constant that depends only on the (apriori known) frequency  $\omega_0$  and the boundary of the set of parameters  $\alpha, \beta, \gamma, K_1$  and  $K_2$ . Having (3.14), we now easily see that statement (2.3) becomes a consequence of

$$(3.15) \quad \frac{1}{N} \sum_{n=1}^N \{h_j(n)s_{k_j(n)}\}L_n \rightarrow 0, \quad N \rightarrow \infty,$$

where the definition of  $L_n$  is given in (3.7). Due to the already verified statement  $N^{-1} \sum_{n=1}^N L_n \rightarrow \lambda(s)$  when  $N \rightarrow \infty$  (cf. Subsection 3.1), we see that (3.15) will follow if we choose  $h_j(n)$  so that

$$(3.16) \quad h_j(n)s_{k_j(n)} \rightarrow 0, \quad n \rightarrow \infty.$$

Trying not to over complicate Proposition 2.1 with at this stage unnecessary details, we have chosen  $h_j(n)$  as in (2.13). This choice implies, of course, that  $h_j(n)s_{k_j(n)} = O(n^{-1})$ —a much stronger result than we need in (3.16). We therefore see that there is ample of space to reduce the size of  $h_j(n)$ , which will clearly result in a smaller window  $W_0$ , and thus  $W$  as well.  $\square$

### 3.3 Verification of statement (2.5)

Let us show that statement (2.5) holds true with  $L^* = L^+$ . To this end we firstly note that when  $n \geq N_0$  for some (large) number  $N_0 \in \mathbf{N}$  depending only on the (fixed) frequency  $\omega_0$ , then the random variables  $\xi_{k_1(n)}(h_1(n)), \dots, \xi_{k_4(n)}(h_4(n))$  are independent. Having thus defined  $N_0$ , we decompose the sum  $\sum_{n=1}^N$  of (2.5) into the following two ones:  $\sum_{n=1}^{N_0-1}$  and  $\sum_{n=N_0}^N$ . It is easy to check that (2.5) with  $\sum_{n=1}^{N_0-1}$  instead of  $\sum_{n=1}^N$  holds true. Consequently, Step 2.3 will be completed if we prove (2.5) with  $\sum_{n=N_0}^N$  instead of  $\sum_{n=1}^N$ . In order to do that we need the following easy-to-prove result: If  $\eta$  is a Poisson random variable with  $E\eta = p \geq 1$ , and  $\alpha \in [-1/4, 1/4]$ , then there exists  $\theta, |\theta| \leq 10$ , such that

$$(3.17) \quad E\mathbf{1}\{\eta > 0\}\eta^\alpha = p^\alpha(1 + \theta p^{-1/4}).$$

(Let us note in passing that the range of  $\alpha$  in (3.17) can be increased, the exponent  $-1/4$  can also be made sharper. We do not, however, need such improvements at this stage of investigation.) To continue the proof of Proposition 2.1, we denote

$$(3.18) \quad \eta := X(h_j(n)[-1, 1] + s_{k_j(n)}).$$

Since  $\eta = 2h_j(n)\xi_{k_j(n)}(h_j(n))$ , equality (3.14) immediately implies

$$(3.19) \quad E\eta \rightarrow \infty, \quad n \rightarrow \infty,$$

provided that

$$(3.20) \quad h_j(n) \exp\{ck_j(n)\} \rightarrow \infty, \quad n \rightarrow \infty,$$

where  $c \in (0, \infty)$  is a fixed constant. (If in (3.16) we were definitely able to choose  $h_j(n) = 0$ , for example, then we now see from (3.20) that too small  $h_j(n)$ 's cannot be used. This observation is not surprising since we have to somehow accumulate the empirical information required for the estimation of  $\lambda_0(s)$ .) But statement (3.20) holds true with  $k_j(n)$  and  $h_j(n)$  defined in (2.12) and (2.13), respectively. Hence,  $E\eta \geq 1$  by (3.19), and we may thus use (3.17) with  $\eta$  defined in (3.18) to get the following statement

$$(3.21) \quad E\mathbf{1}\{\xi_{k_j(n)}(h_j(n)) > 0\}\xi_{k_j(n)}(h_j(n))^\alpha \\ = \{E\xi_{k_j(n)}(h_j(n))\}^\alpha \{1 + \theta\{2h_j(n)E\xi_{k_j(n)}(h_j(n))\}^{-1/4}\}$$

that holds true for any  $\alpha \in [-1/4, 1/4]$  and some  $|\theta| \leq 10$ . Consequently, statement (3.21) holds true with  $\alpha = \alpha_i, i = 1, \dots, 4$ , where

$$\begin{aligned} \alpha_1 &:= -(s - s_{k_4(n)})s_{k_2(n)}/s_{k_1(n)}^2, \\ \alpha_2 &:= (s - s_{k_4(n)})/s_{k_2(n)}, \\ \alpha_3 &:= (s^2 - s_{k_4(n)}^2)/s_{k_3(n)}^2, \\ \alpha_4 &:= 1. \end{aligned}$$

Indeed, all the three quantities  $\alpha_1, \alpha_2$  and  $\alpha_3$  converge to 0 when  $n \rightarrow \infty$ , and we therefore have  $|\alpha_1|, |\alpha_2|, |\alpha_3| \leq 1/4$  when  $n \geq \text{const}$ . In the case  $\alpha_4 = 1$ , statement (3.21) is trivially true. The completion of Step 2.3 becomes very similar to that of Step 2.2, and we therefore omit further details.  $\square$

3.4 *Verification of Assumption 2.1*

Assumption 2.1 is satisfied due to the trivial bound

$$|s_{k_j(m)} - s_{k_\ell(n)}| \geq 2\pi/\omega_0 > 0,$$

that holds true provided that the points  $s_{k_j(m)}$  and  $s_{k_\ell(n)}$  do not coincide.  $\square$

3.5 *Verification of statement (2.10)*

In order to verify statement (2.10), we firstly note that (2.10) is a consequence of

$$(3.22) \quad \frac{1}{N^2} \sum_{n=1}^N \mathbf{E} L^{*2}(\xi_{k_1(n)}(h_1(n)), \dots, \xi_{k_4(n)}(h_4(n)), s_{k_1(n)}, \dots, s_{k_4(n)}, s) \rightarrow 0,$$

when  $N \rightarrow \infty$ . Following the lines of Subsection 3.3, one can easily show that statement (2.5) holds true with the functions  $L^{*2}$  and  $L^2$  instead of  $L^*$  and  $L$ , respectively. Hence, (3.22) becomes a consequence of the following statement

$$(3.23) \quad \frac{1}{N^2} \sum_{n=1}^N L^2(\mathbf{E}\xi_{k_1(n)}(h_1(n)), \dots, \mathbf{E}\xi_{k_4(n)}(h_4(n)), s_{k_1(n)}, \dots, s_{k_4(n)}, s) \rightarrow 0,$$

when  $N \rightarrow \infty$ . If we now follow the lines of Subsection 2.2, we shall see that statement (3.23) is a consequence of

$$(3.24) \quad \frac{1}{N^2} \sum_{n=1}^N L^2(\lambda(s_{k_1(n)}), \dots, \lambda(s_{k_4(n)}), s_{k_1(n)}, \dots, s_{k_4(n)}, s) \rightarrow 0,$$

when  $N \rightarrow \infty$ . But (3.24) holds true because statement (2.2), as it is easy to check, holds true with  $L^2$  and  $\lambda_0^2(s)$  instead of  $L$  and  $\lambda_0(s)$ , respectively, and because of the fact that  $\lambda_0^2(s)/N \rightarrow 0$  when  $N \rightarrow \infty$ . In this way, the proof of (3.22) is finished. This also completes the verification of statement (2.10). The proof of Proposition 2.1 is completed as well.  $\square$

4. Appendix: A discussion with Referee A

Let  $X_1$  be Poisson point process on  $[0, \infty)$  with the intensity function

$$\lambda_1(s) := \exp\{\alpha + \beta s\}, \quad s \geq 0,$$

where  $\alpha$  and  $\beta$  are unknown parameters from an apriori fixed compact interval of  $\mathbf{R}$ . Using the idea of Subsection 3.0, one *directly* arrives at the function

$$(4.1) \quad L(x_1, x_2, y_1, y_2, s) := x_1^{(s-y_2)/y_1} x_2$$

and then easily shows that Step 2.1 is valid, for example, with

$$s_n := s + n,$$

and

$$\begin{cases} k_1(n) := 4n^2 - 2 \\ k_2(n) := 2n - 1, \end{cases}$$

for  $n \in \mathbf{N}$ .

Referee A of the present paper, however, proposed that, instead of using the somewhat cumbersome idea of Subsection 3.0, one can arrive at the function of (4.1) in the following way: Let us solve the equations

$$\begin{cases} \log \lambda_1(s_1) = \alpha + \beta s_1, \\ \log \lambda_1(s_2) = \alpha + \beta s_2. \end{cases}$$

If we now put the just obtained solutions

$$\begin{cases} \alpha = \log \left\{ \frac{\lambda_1(s_2)^{s_1}}{\lambda_1(s_1)^{s_2}} \right\}^{1/(s_1-s_2)}, \\ \beta = \log \left\{ \frac{\lambda_1(s_1)}{\lambda_1(s_2)} \right\}^{1/(s_1-s_2)} \end{cases},$$

into the definition of  $\lambda_1(s)$ , we shall arrive at the equality

$$(4.2) \quad \left\{ \frac{\lambda_1(s_2)^{s_1}}{\lambda_1(s_1)^{s_2}} \right\}^{1/(s_1-s_2)} \left\{ \frac{\lambda_1(s_1)}{\lambda_1(s_2)} \right\}^{s/(s_1-s_2)} = \lambda_1(s)$$

which is, of course, equivalent to

$$(4.3) \quad \lambda_1(s_1)^{(s-s_2)/(s_1-s_2)} \lambda_1(s_2)^{(s_1-s)/(s_1-s_2)} = \lambda_1(s).$$

Furthermore, assuming

$$(4.4) \quad \begin{cases} s_1, s_2 \rightarrow \infty, \\ s_1/s_2 \rightarrow \infty, \\ s \text{ is fixed,} \end{cases}$$

we immediately arrive at the approximation

$$(4.5) \quad \lambda_1(s_1)^{(s-s_2)/s_1} \lambda_1(s_2) \approx \lambda_1(s).$$

If we now compare the left-hand side of (4.5) with the definition of  $L$  in (4.1), the desired insight into the derivation of (4.1) will be achieved.

Referee A also proposed the following general idea for a construction of function  $L$  required by Step 2.1: Fix several values of  $s$ , say  $s_1, \dots, s_k$ , if the unknown parameter  $\theta := (\theta_1, \dots, \theta_k)$  is  $k$ -dimensional, and solve the  $k$  equations

$$\lambda(\theta; s_i) \equiv \lambda(\theta_1, \dots, \theta_k; s_i) = \lambda(s_i), \quad i = 1, \dots, k.$$

Then substitute the obtained values of  $\theta$  into the initial formula for the intensity function  $\lambda(s) \equiv \lambda(\theta; s)$ . Having this function of  $s$ , one organizes observations in such a way which provides consistency, etc.

Let us start our reaction to the Referee A proposal by noting that our (heuristic) approach given in Subsection 3.0 leads us *directly* to (4.5) without first arriving at (4.2). In order to appreciate this fact, one is advised to apply the Referee A proposal to the intensity function  $\lambda_0$ —calculations will immediately become quite involved. Moreover, if we decide to go further and complicate the task by considering polynomial trends of higher than 2 order, then calculations may become even impossible to write down, and the desired estimator either.

Indeed, a real problem (in fact, the only problem as we can see) with the Referee A proposal is that it involves *exact* solutions of algebraic equations, which (just like when using the MLE method) are sometimes difficult to obtain. (We have explained at the beginning of our manuscript why numerical solutions may not be satisfactory.) Our approach of Subsection 3.0 is designed to avoid such problems, or at least to reduce them to a minimal level.

On the other hand, we ought to emphasize that if the Referee A proposal works (just like in the case of  $\lambda_1$  described above), then we *do* recommend to use it and not to follow the route described in Subsection 3.0. Indeed, in the case of  $\lambda_1$ , for example, we shall likely do better in many respects when estimating  $\lambda_1$  if we choose the left-hand side of (4.2), or (4.3), as the definition of  $L$  required in (2.2) of Step 2.1. The reason is that equality (4.2), or (4.3), works without the assumption  $s_1/s_2 \rightarrow \infty$  spelled out in (4.4), and one may therefore use points  $s_1$  and  $s_2$  (both  $\rightarrow \infty$ ) that are asymptotically equivalent, i.e.,  $s_1 = O(s_2)$  and  $s_2 = O(s_1)$ . This should result in a possibly smaller window  $W$  required to collect the necessary information about  $X$ .

Consequently, we see that there are some seemingly minor, though sometimes substantial, differences between the Referee A and our approaches for constructing the function  $L$  required in (2.2). Both approaches, therefore, complement each other—one occasionally working better than the other one and vice versa—and thus worth of pursuing for the sake of manifold probabilistic and statistical properties, optimality considerations, etc.

In order to conclude the discussion, we note that the Referee A proposal can always serve as a very good clue for making a decision in what direction one has to go in order to construct a function  $L$  required in Step 2.1.

## Acknowledgements

The main idea of this paper was worked out during a visit of the second author at the Centre for Mathematics and Computer Science (CWI), Amsterdam, and completed while visiting the Laboratory for Research in Statistics and Probability, Ottawa, as a Canada International Fellow. He sincerely thanks both institutions for hospitality. The visit of R.Z. to CWI in October, 1994, was funded by the Stimulans programme of the Netherlands Organization for Scientific Research (NWO).

Both authors are grateful to Professor David Vere-Jones for his encouragement and suggestions during the preparation of the paper. We are also very much indebted to the two anonymous referees for their constructive criticism, comments and remarks that resulted in two major revisions of the original manuscript. We thank them all, as well as Editor Shinto Eguchi, for their time and kind patience.

## REFERENCES

- Cox, D. R. (1972). The statistical analysis of dependencies in point processes, *Stochastic Point Processes* (ed. P. A. W. Lewis), 55–66, Wiley, New York.
- Cox, D. R. and Isham, V. (1980). *Point Processes*, Chapman and Hall, London.
- Cox, D. R. and Lewis, P. A. W. (1978). *The Statistical Analysis of Series of Events*, Chapman and Hall, London.
- Cressie, N. A. C. (1991). *Statistics for Spatial Data*, Wiley, New York.
- Daley, D. J. and Vere-Jones, D. (1972). A summary of the theory of point processes, *Stochastic Point Processes* (ed. P. A. W. Lewis), 299–383, Wiley, New York.
- Daley, D. J. and Vere-Jones, D. (1988). *Introduction to the Theory of Point Processes*, Springer, New York.
- Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The  $L_1$  View*, Wiley, New York.
- Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Processes*, Academic Press, London.
- Helmers, R. (1995). On estimating the intensity of oil-pollution in the North-Sea, Report BS-N9501, Centre for Mathematics and Computer Science, Amsterdam.
- Kallenberg, O. (1983). *Random Measures*, Academic Press, New York.
- Karr, A. F. (1986). *Point Processes and Their Statistical Inference*, Marcel Dekker, New York.
- Kingman, J. F. C. (1993). *Poisson Processes*, Clarendon Press, Oxford.
- Krickeberg, K. (1982). Processus ponctuels en statistique, *Lecture Notes in Math.* (ed. P. L. Hennequin), **929**, 205–313, Springer, New York.
- Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes*, Heldermann, Berlin.
- Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound Vibration*, **12**, 353–375.
- Lewis, P. A. W. (1972). Recent results in the statistical analysis of univariate point processes, *Stochastic Point Processes* (ed. P. A. W. Lewis), 1–54, Wiley, New York.
- Ogata, Y. and Katsura, K. (1986). Point-process models with linearly parametrized intensity for application to earthquake data, *J. Appl. Probab.*, **23A**, 291–310.
- Rathbun, S. L. and Cressie, N. (1994). Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes, *Advances in Applied Probability*, **26**, 122–154.
- Reiss, R.-D. (1993). *A Course on Point Processes*, Springer, New York.
- Ripley, B. D. (1976). Locally finite random sets: foundations for point process theory, *Ann. Probab.*, **4**, 983–994.
- Ripley, B. D. (1991). *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

- Rosenblatt, M. (1991). *Stochastic Curve Estimation*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 3, IMS, Hayward, and ASA, Alexandria.
- Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk, *International Journal of Forecasting*, **11**, 503-538.
- Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical estimation applied to earthquake data, *Ann. Inst. Statist. Math.*, **34**, 189-207.
- Zheng, X. and Vere-Jones, D. (1994). Further applications of the stochastic stress release model to historical earthquake data, *Technophysics*, **229**, 101-121.