RT-MAE 2006-06

ON THE ESTIMATION OF THE INTENSITY OF POINT PROCESSES VIA WAVELETS

by

José Carlos Simon de Miranda
and
Pedro Alberto Morettin

Palavras-Chave: Intensity, non-internally correlated point processes, Poisson processes, sure inference, threshold, wavelets.
Classificação AMS: 62M12, 60G10.

- Abril de 2006 -

On the Estimation of the Intensity of Point Processes via Wavelets

José Carlos Simon de Miranda and Pedro A. Morettin

University of São Paulo

Abstract

In this article we consider the problem of estimating the intensity of a non-homogeneous point process on the real line. The approach used is via wavelet expansions. Estimators of the intensity are proposed and their properties are studied, including the case of threshold versions. Properties for the non-homogeneous Poisson process follow as special cases. An application is given for the series of daily Dow Jones indices. Extensions to more general settings are also indicated.

keywords. Intensity, non-internally correlated point processes, point processes, Poisson processes, sure inference, threshold, wavelets.

1. Introduction

In this article we consider the problem of estimating the intensity of a point process $\{N(t), t \in \mathbb{R}\}$ denoted by $p_N(t)$. This topic has been discussed in several works, and we mention Brillinger (1975, 1978), Snyder (1975), Rathbun and Cressie (1994) and Helmers and Zitikis (1999).

Contrary to the approach adopted in several works, we do not assume that the intensity $p_N(t)$ is within a family of parametric models $p_N(t;\theta)$, so that the only issue there is the estimation of the unknown parameter vector θ. The approach that will be used in this work is via wavelet expansions, as in Donoho et al. (1996). Wavelets provide a way of estimating intensities for non-homogeneous point processes, due to their ability to smooth with a variable bandwidth. We will focus on processes on the real line, but extensions to higher dimensions and more general spaces are possible. Other papers on wavelets and point processes are Brillinger (1997), Timmermann and Nowak (1997), Kolaczyk (1999a, 1999b) and Besbea et al. (2002).

In order to maintain the general character of the work, we develop an analysis of inference which allows us to obtain substitutes to the confidence intervals and bands for parameters and functions, respectively, without relying on the distributions (exact or asymptotic) of the respective estimators. Since the intervals and bands obtained through this analysis are extremely cautions we decided to call it sure inference analysis. We define the concept of inferential sequence, which is central to this analysis.

To solve the problem of estimation of the intensity of a point process we adopt the following approach. We expand the restriction of the intensity function to the interval where we know the points of a trajectory of the underlying process in a wavelet series. We then propose unbiased estimators for the coefficients of this expansion as well as estimators for the variance of each estimator. We also obtain an inferential sequence for the wavelet coefficients for non-internally correlated point processes and Poisson process as a particular case. From the estimators of the coefficients we obtain an unbiased estimator for the intensity function. The propositions of interest are proved for special classes of point processes, satisfying Assumptions A and B, to be introduced below.

The plan of the article is as follows. In Section 2 we provide some background on point processes and wavelets. In Section 3 we define the classes of point process satisfying Assumptions A and B. This section also establishes the sure inference analysis. In section 4 we propose a way of estimating the intensity function and derive some properties of the estimators, specializing for the case of a non-internally correlated point processes. Threshold estimators are studied in Section 5 and an application is given in Section 6. We close the work with some further considerations in Section 7.

2. BACKGROUND

In this section we provide some background material on point processes on the real line and on wavelets which will be used in the sequel.

2.1. Notation.

We first introduce some notation that will be necessary. We will work with Lebesgue measurable functions, $h : \mathbb{R}^m \to \mathbb{R}$ which are bounded in bounded intervals of $\mathbb{R}^m$ or, equivalently, which are integrable in the sense of Lebesgue and bounded on bounded intervals of $\mathbb{R}^m$. Let us call this class of functions $\mathcal{L}^m$. Denote by $\overline{\mathcal{L}}^m$ the class of functions which are Lebesgue integrable over bounded intervals of $\mathbb{R}^m$.

We will use the notation $|a,b|$, $a = (a_1, \ldots, a_m)$, $b = (b_1, \ldots, b_m)$ to represent any of the 4^m possible intervals of $\mathbb{R}^m$ which can be written in the form $\prod_{i=1}^m |a_i, b_i|$, where $|a_i, b_i|$ represents one of the intervals (a_i, b_i), $(a_i, b_i]$, $[a_i, b_i)$ or $[a_i, b_i]$ of the real line. We also use the notation χ_C for the characteristic function (or indicator) of a set C ($\chi_C(x) = 1 \leftrightarrow x \in C \ \wedge \ \chi_C(x) = 0 \leftrightarrow x \notin C$). Lebesgue measure on $\mathbb{R}^m$ will be indicated simply by ℓ independently of the dimension m. If it is necessary to emphasize the dimension we will write ℓ_m. The σ-algebra of Lebesgue measurable sets in $\mathbb{R}^m$ is denoted by $\Lambda_{\mathbb{R}^m}$. $\mathcal{B}_{\mathbb{R}^m}$ is used for the σ-algebra of Borel sets. Functions that differ over a subset of zero measure of their common domain or of common extensions of their domain are, naturally, when necessary, considered identical.

2.2. Point Processes.

We denote by $N(A)$ the number of events of a certain sort that occur in $A \subset \mathbb{R}$. If $A = (\alpha, \beta]$, we write $N(\alpha, \beta]$ instead of $N((\alpha, \beta])$. We also denote by N the integer valued function defined by the equalities $N(t) = N(0, t]$, if $t > 0, N(0) = 0$ and $N(t) = -N(t, 0]$ if $t < 0$. Clearly $N(\alpha, \beta] = N(\beta) - N(\alpha)$. Let $\{\cdots, \tau_{-2} \leq \tau_{-1} \leq \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots\}$ denote the times at which the events occur. Then $N(t) = n$, if and only if $\tau_{n-1} \leq t < \tau_n$.

Provided probabilities of the form

$$P(N(\alpha_1, \beta_1] = n_1, \ldots, N(\alpha_k, \beta_k] = n_k)$$

are defined and consistent, for all $k \in \mathbb{N}^* = \{1, 2, \ldots\}$, and all $n_1, \ldots, n_k$ non-negative integers, we can define an appropriate probability space $(\Omega, \mathcal{A}, P)$, such that there exists a measurable mapping from this space into $(\mathbb{R}^Z, \mathcal{B}_{\mathbb{R}^Z})$, defining then a stochastic point process that will also be called N. See Cramér and Leadbetter (1967) and Daley and Vere-Jones (1988) for details and alternative definitions.

One important point process is the (non-homogeneous) Poisson process, for which we are given a non-decreasing, right-continuous function $\Lambda(t)$, such that whenever $(\alpha_i, \beta_i] \cap (\alpha_j, \beta_j] = \phi$, for all

$i \neq j$,

$$P(N(\alpha_1, \beta_1] = n_1, \ldots, N(\alpha_k, \beta_k] = n_k) = \prod_{j=1}^{k} \left(\frac{[\Lambda(\beta_j) - \Lambda(\alpha_j)]^{n_j}}{n_j!} \exp\{\Lambda(\alpha_j) - \Lambda(\beta_j)\} \right).$$

As a consequence of this formula, the random variables $N(\alpha_j, \beta_j]$ form a completely independent set, or equivalently, events in disjoint intervals are independent. An important special case is when $\Lambda(t) = \lambda t$, λ being the mean intensity of the process.

Another important point process is the doubly stochastic point process, when we start with a realization $\Lambda(t)$ of a process, assumed to be stationary, non-decreasing, continuous from the right, and then generate a Poisson process with intensity $\Lambda(t)$.

Define $dN(t) = N(t + dt) - N(t)$. A basic assumption is that there exist boundedly finite measures M_k such that

$$E\{dN(t_1) \cdots dN(t_k)\} = M_k(dt_1, \ldots, dt_k),$$

that is, $E(\prod_{i=1}^{k} N) = M_k$.

We will be often dealing with integrals of the form

$$\int \varphi(t) dN(t) = \sum_j \varphi(\tau_j).$$

Suppose that $\varphi_i, 1 \leq i \leq k$, are (essentially) bounded measurable functions, with compact support. Then,

$$E\{\int \varphi_1(t_1) dN(t_1) \cdots \int \varphi_k(t_k) dN(t_k)\} = \int \varphi_1(t_1) \cdots \varphi_k(t_k) dM_k(t_1, \ldots, t_k).$$

In particular, we have the following theorem. (See Daley and Vere-Jones, 1988).

Theorem 2.1. (Campbell's Theorem) *Let N such that $EN(A) < \infty$ for all bounded set A that belongs to $\mathcal{B}_{\mathbb{R}}$. Then, for all bounded measurable function φ, with compact support, we have*

$$E\{\int \varphi(t) dN(t)\} = \int \varphi(t) E dN(t).$$

2.3. Intensity and Product Density. Suppose that there exists a positive real number δ and a constant $K_\delta > 0$ such that for all intervals $\Delta \subset \mathbb{R}$ with length $|\Delta| < \delta$, all integers $n > 1$ and all $t \in \mathbb{R}$, not only the relation

$$P\{N(\Delta) = n\} \leq K_\delta |\Delta|^n \tag{1}$$

holds, but also the limit

$$\lim_{|\Delta| \to 0, t \in \Delta} \frac{1}{|\Delta|} P\{N(\Delta) = 1\} = p_N(t) \tag{2}$$

exists uniformly in t. Inequality (1) implies that

$$P\{N(\Delta) > 1\} \leq K_\delta (\sum_{j \geq 2} |\Delta|^j) = O(|\Delta|^2).$$

Notice that if inequality (1) were valid for $n = 1$ then we would have $P\{N(\Delta) = 1\}/|\Delta| \leq K_\delta$ and hence, if it would exist, $p_N(t)$ would be a bounded function on $\mathbb{R}$. Notice also that (2) implies that $\forall x \in \mathbb{R}$, $P\{N(\{x\}) = 1\} = 0$, otherwise there would exist $t \in \mathbb{R}$ for which the limit $p_N(t)$ would be infinite.

Due to uniformity, relation (2) is equivalent to

$$P\{N(\Delta) = 1\} = p_N(t)|\Delta| + o_{t,\Delta}(|\Delta|),$$

for an infinitesimal $o_{t,\Delta}(z)$ with the following properties:
$\forall \varepsilon > 0$, $\exists \delta > 0$, $\forall t \in \mathbb{R}$, $\forall \Delta \subset \mathbb{R}, t \in \Delta$, $(0 < |\Delta| < \delta) \to |o_{t,\Delta}(|\Delta|)| \le \frac{\varepsilon}{2}|\Delta|$ and $o_{t,\Delta}(0) = 0$, that is,
$\forall \varepsilon > 0$, $\exists \delta > 0$, $(0 < z < \delta) \to \sup_{\substack{t \in \mathbb{R}, \Delta \subset \mathbb{R} \\ t \in \Delta, |\Delta| = z}} |o_{t,\Delta}(z)| \le \frac{\varepsilon}{2} z < \varepsilon z$ and $o_{t,\Delta}(0) = 0$.

The quantity $\sup_{\substack{t \in \mathbb{R}, \Delta \subset \mathbb{R} \\ t \in \Delta, |\Delta| = z}} |o_{t,\Delta}(z)| = o(z)$ is a non-negative infinitesimal independent of t and Δ. In this sense, we also write $|o_{t,\Delta}(|\Delta|)| \le o(|\Delta|)$.

For the easy of notation, we will write o_t instead of $o_{t,\Delta}$.

We say that $p_N(t)$ is the **intensity** of occurrence of events at time t.

Suppose now that there exists a positive real number δ and a constant $k_{\delta,m}$ such that for all intervals $\Delta_1, \ldots, \Delta_m$ of the real line with lengths $0 < |\Delta_i| < \delta$, $1 \le i \le m$, all integers $n_i \ge 1$ and all vectors $(t_1, \ldots, t_n) \in \mathbb{R}^m$ with $t_i \ne t_j$ for $i \ne j$, $1 \le i \le m$, $1 \le j \le m$, both properties below are valid:

(3) $$\text{if } (n_1, \ldots, n_m) \ne (1, \ldots, 1) \text{ then } P\{N(\Delta_i) = n_i, 1 \le i \le m\} \le k_{\delta,m} \prod_{i=1}^{m} |\Delta_i|^{n_i}$$

and for $\Delta = (|\Delta_1|, \ldots, |\Delta_m|) \in (\mathbb{R}_+^*)^m$, $t_i \in \Delta_i$, $1 \le i \le m$, there exists the limit

(4) $$\lim_{\Delta \to 0} \frac{1}{\prod_{i=1}^{m} |\Delta_i|} P\{N(\Delta_i) = 1, 1 \le i \le m\} = p_m(t_1, \ldots, t_m),$$

uniformly in $t = (t_1, \ldots, t_m)$.

Observe that for $m = 1$ the symbol Δ has two different meanings, the interval and the length, but this will be of no harm.

The above limit measures the intensity of the joint occurrence of events in the distinct instants $t_1, \ldots, t_m$. We might call it the joint intensity. Since under the relations (3) and (4) it is also valid that $\lim_{\Delta \to 0} \frac{1}{\prod_{i=1}^{m} |\Delta_i|} E\{\prod_{i=1}^{m} N(\Delta_i)\} = p_m(t_1, \ldots, t_m)$, p_m is called **product density of order** m. Relation (4) implies that

$$P\{N(\Delta_i) = 1, 1 \le i \le m\} = p_m(t_1, \ldots, t_m) \prod_{i=1}^{m} |\Delta_i| + o_{t, \prod_{i=1}^{m} \Delta_i}(\Delta)$$

for $o_{t, \prod_{i=1}^{m} \Delta_i}(\Delta)$ an infinitesimal such that

$$\sup_{\substack{t \in \mathbb{R}^m - \mathcal{E}^m, \prod_{i=1}^{m} \Delta_i \subset \mathbb{R}^m \\ t \in \prod_{i=1}^{m} \Delta_i, |\Delta_i| = z_i, 1 \le i \le m}} |o_{t, \prod_{i=1}^{m} \Delta_i}(z)| = o(z),$$

where $z = (z_1, \ldots, z_m) \in (\mathbb{R}_+^*)^m$, is another infinitesimal which is independent of $t \in \mathbb{R}^m - \mathcal{E}^m$ and $\prod_{i=1}^m \Delta_i \subset \mathbb{R}^m$, that satisfies $\frac{o(\Delta)}{\prod_{i=1}^m |\Delta_i|} \to 0$ when $\Delta \to 0$. Here we denote by $\mathcal{E}^m$ the set

$$\{(t_1, \ldots, t_m) \in \mathbb{R}^m | t_i = t_j \text{ for some } i \neq j\}.$$

Again, for the easy of notation, we write o_t instead of $o_{t, \prod_{i=1}^m \Delta_i}$.

We can also define cumulants for $N(t)$; and in particular, we define the limit covariance, for $u \neq v$, by

$$q_2(u,v) = \lim_{\Delta \to 0} \frac{\mathrm{Cov}\,(N,N)(\Delta_1 \times \Delta_2)}{|\Delta_1||\Delta_2|}.$$

Whenever $p_2(u,v)$, $p_1(u)$ and $p_2(v)$ exist, we write

$$\begin{aligned} q_2(u,v) &= \lim_{\Delta \to 0} \frac{\mathrm{Cov}\,(N,N)(\Delta_1 \times \Delta_2)}{|\Delta_1||\Delta_2|} \\ &= \lim_{\Delta \to 0} \frac{E(N(\Delta_1)N(\Delta_2))}{|\Delta_1||\Delta_2|} - \lim_{\Delta \to 0} \frac{E(N(\Delta_1))}{|\Delta_1|} \frac{E(N(\Delta_2))}{|\Delta_2|} \\ &= p_2(u,v) - p_1(u)p_2(v). \end{aligned}$$

2.4. **Point Processes and Infinitesimals.** In this section we will present some useful results.

Proposition 2.1. *Under conditions (1) and (2), we have*

$$\begin{aligned} P\{N(\Delta) = 1\} \leq E\{N(\Delta)\} \leq P\{N(\Delta) = 1\} + O(|\Delta|^2), \\ P\{N(\Delta) = 1\} - A \leq \mathrm{Var}\,\{N(\Delta)\} \leq P\{N(\Delta) = 1\} + B, \end{aligned}$$

where A and B are $O(|\Delta|^2)$ whenever $\sup_{t \in \Delta} p_N(t)$ is finite.

Therefore we can write

$$E\{N(\Delta)\} = p_N(t)|\Delta| + o_t(|\Delta|)$$

and

$$\mathrm{Var}\,\{N(\Delta)\} = p_N(t)|\Delta| + o_t(|\Delta|).$$

These $o_t = o_{t,\Delta}$ may depend on t and Δ but their absolute values are bounded by other o's which are independent of t.

Proposition 2.2. *Under the hypotheses (3) and (4) we have, for $m \geq 1$,*

$$\begin{aligned} &P\{N(\Delta_i) = 1,\ 1 \leq i \leq m\} \leq E\left\{\prod_{i=1}^m N(\Delta_i)\right\} \\ &\leq P\{N(\Delta_i) = 1,\ 1 \leq i \leq m\} + k_{\delta,m} \left\{\prod_{i=1}^m \left(\frac{1}{1-|\Delta_i|}\right)^2 - 1\right\} \prod_{i=1}^m |\Delta_i|. \end{aligned}$$

Theorem 2.2. *Let $\mathcal{E}^m$ as before , φ an $E\left(\prod_{i=1}^{m} \mathrm{d}N(t_i)\right)$-integrable function over $\mathbb{R}^m - \mathcal{E}^m$, p_m the m-th order product density and $p_1 = p_N$ the intensity function of a point process N that satisfies (3) and (4). Then, if $p_m \in \overline{\mathcal{L}}^m$, $m \geq 1$, we have*

$$\int_{\mathbb{R}^m - \mathcal{E}^m} \varphi E\left(\prod_{i=1}^{m} \mathrm{d}N(t_i)\right) = \int_{\mathbb{R}^m - \mathcal{E}^m} \varphi p_m \prod_{i=1}^{m} \mathrm{d}t_i \ .$$

We observe that this theorem shows that if the intensity function or the product density p_m is a.e.$[\ell]$ defined as an uniform limit and it is Lebesgue-integrable over limited $\mathbb{R}^m$-intervals, then it is also the Radon-Nikodym derivative of $E\left(\prod_{i=1}^{m} N\right)$ with respect to ℓ. Clearly, $\mathcal{E}^1 = \emptyset$ and $\mathcal{E}^2 = D = \{(x,x) \in \mathbb{R}^2 | x \in \mathbb{R}\}$ is the diagonal set of $\mathbb{R}^2$.

Proofs and further results can be found in de Miranda and Morettin (2003a) and de Miranda (2003b).

2.5. **Wavelets.** Wavelets are building block functions localized in time or space. They are obtained from a single function $\psi(t)$, called the mother wavelet, by translations and dilations. The mother wavelet $\psi(t)$ satisfies the conditions

$$\int_{-\infty}^{\infty} \psi(t)dt = 0, \tag{5}$$

$$\int_{-\infty}^{\infty} | \psi(t) | \, dt < \infty, \tag{6}$$

and may also satisfy

$$\int_{-\infty}^{\infty} \frac{| \hat{\psi}(\omega) |^2}{| \omega |} d\omega < \infty, \tag{7}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$, that is,

$$\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(t) e^{-i\omega t} dt.$$

Given a mother wavelet $\psi(t)$, for all real numbers $a, b (a \neq 0)$, we construct a wavelet by translation and dilation of $\psi(t)$,

$$\psi^{(a,b)}(t) = | a |^{-1/2} \psi\left(\frac{t-b}{a}\right),$$

where a represents the dilation parameter and b the translation parameter.

For some very special choices of ψ and a, b, the set $\{\psi^{(a,b)}\}$ constitute an orthonormal basis for $L^2(\mathbb{R})$. In particular, if we choose $a = 2^{-j}$, $b = k2^{-j}$, $j, k \in \mathbb{Z}$, then there exists ψ, such that

$$\psi_{k,j}(t) = \psi^{(a,b)}(t) = 2^{j/2} \psi(2^j t - k), \tag{8}$$

constitute an orthonormal basis for $L^2(I\!R)$.

There are many different forms of $\psi(t)$ all of which satisfy the conditions (5), (6) and (7). The oldest and simplest example of a function ψ for which the $\psi_{k,j}$ defined by (8) constitute an orthonormal basis for $L^2(I\!R)$ is the Haar function,

$$\psi^{(H)}(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

From (9), we have

$$\psi^{(H)}_{k,j}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \le t < 2^{-j}(k+\frac{1}{2}) \\ -2^{j/2}, & 2^{-j}(k+\frac{1}{2}) \le t < 2^{-j}(k+1) \\ 0, & \text{otherwise.} \end{cases}$$

One way to find a wavelet function is by the use the dilation equation

$$\phi(t) = \sqrt{2}\sum_k l_k\phi(2t-k),$$

where $\phi(t)$ is the so-called *scaling function* (or father wavelet), satisfying $\int_{-\infty}^{\infty}\phi(t)dt = 1$. Then the mother wavelet $\psi(t)$ is obtained from the father wavelet through

$$\psi(t) = \sqrt{2}\sum_k h_k\phi(2t-k),$$

with $h_k = (-1)^k l_{1-k}$, called the quadrature mirror filter relation, where the coefficients l_k and h_k are the low-pass and high-pass filter coefficients given by the formulas

$$l_k = \sqrt{2}\int_{-\infty}^{\infty}\phi(t)\phi(2t-k)dt$$

and

$$h_k = \sqrt{2}\int_{-\infty}^{\infty}\psi(t)\phi(2t-k)dt,$$

respectively.

For the Haar wavelet,

$$\phi^{(H)}(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{otherwise,} \end{cases}$$

hence,

$$l_k = \sqrt{2}\int\phi(t)\phi(2t-k)dt = \begin{cases} 1/\sqrt{2}, & k = 0,1 \\ 0, & \text{otherwise,} \end{cases}$$

and

$$h_0 = l_1 = 1/\sqrt{2}, \qquad h_1 = -l_0 = -1/\sqrt{2}.$$

Consequently,

$$\psi(t) = \sqrt{2}((1/\sqrt{2})\phi(2t) - (1/\sqrt{2})\phi(2t-1)),$$

and (9) is obtained.

Another way to construct wavelet bases is applying multi-resolution analysis. See Meyer (1992). Except for some special cases, there are no analytic formulas for computing wavelet functions.

An important result due to Daubechies guarantees, for all r, the existence of orthonormal bases for $L^2(\mathbb{R})$ of the form $2^{j/2}\psi_{(r)}(2^jx-k)$, $j,k \in \mathbb{Z}$, having the following properties: the support of $\psi_{(r)}$ is the interval $[0, 2r+1]$,

$$0 = \int \psi_{(r)}(x)\mathrm{d}x = \cdots = \int x^r\psi_{(r)}(x)\mathrm{d}x,$$

$\psi_{(r)}$ has $\lfloor \gamma r \rfloor$ continuous derivatives and the positive constant γ is approximately 1/5. The Haar basis is a special case where $r = 0$. In this work we assume that ϕ and ψ are (essentially) bounded with compact support. See Daubechies (1992).

We close this section with some comments on wavelet and Fourier analysis. The functions in a wavelet basis are indexed by two parameters, while in the Fourier basis we have only one parameter, the frequency. So a wavelet function is localized in time and they are good building block for signals which have non-smooth features and features which change over time. For these kinds of signals, Fourier transform coefficients are not well suited. Intuitively, scale can be thought as "inverse frequency", as shown by the following argument (Priestley, 1996). As j increases the scale factor 2^j also increases and there is a shrinking in time that shows that scale has been reduced. At the same time, the oscillations in the mother wavelet increase and exhibit a "high frequency" behavior. On the other hand, as j decreases and scale increases we obtain a "low frequency" behavior. The analysis to be presented below is in time-scale, requiring appropriate interpretations. See Morettin (1999) for details.

3. Assumptions and Sure Inference Analysis

3.1. Assumptions. We make now two assumptions in order to include a larger class of point processes. From now on we do not impose uniformity of the defining limit for the intensity given by equation (2).

Assumption B. A point process N satisfies Assumption B when not only its expectation measure is absolutely continuous in relation to Lebesgue measure, $EN \ll \ell$, that is, when there exists $\mathrm{d}EN/\mathrm{d}\ell \in \overline{\mathcal{L}}^1$, but also the following relation holds: $\forall t \in \mathbb{R}\ \forall \Delta \subset \mathbb{R}$, Δ interval, $t \in \Delta$, $EN(\Delta) = P\{N(\Delta) = 1\} + o_{t,\Delta}(|\Delta|)$.

We notice that for such processes there exists p_N, the defining limit of the intensity and $\mathrm{d}EN/\mathrm{d}\ell = p_N$ a.e.$[\ell]$. In fact, the following result holds.

Theorem 3.1. *Let N be a point process that satisfies Assumption B. Then the intensity defining limit p_N exists and $\mathrm{d}EN/\mathrm{d}\ell = p_N$ a.e.$[\ell]$.*

Proof For all $t \in \mathbb{R}$, we compute the defining limit $p_N(t)$:

$$p_N(t) = \lim_{\substack{t\in\Delta \\ |\Delta|\to 0}} \frac{P\{N(\Delta)=1\}}{|\Delta|} = \lim_{\substack{t\in\Delta \\ |\Delta|\to 0}} \frac{EN(\Delta) - o_t(|\Delta|)}{|\Delta|} = \lim_{\substack{t\in\Delta \\ |\Delta|\to 0}} \frac{EN(\Delta)}{|\Delta|}.$$

Let $f = \frac{\mathrm{d}EN}{\mathrm{d}\ell}$, $\varphi(x) = \int_c^x f(y)\mathrm{d}y$, $\Delta = |a,b|$, $a < b$, $h_1 = b - t$ and $h_2 = t - a$. Thus,

$$p_N(t) = \lim_{\substack{h_1\to 0 \\ h_2\to 0}} \frac{\varphi(t+h_1) - \varphi(t-h_2)}{h_1 + h_2}.$$

Now,

$$\begin{aligned}\frac{\varphi(t+h_1) - \varphi(t-h_2)}{h_1+h_2} &= \frac{\varphi(t+h_1) - \varphi(t)}{h_1}\frac{h_1}{h_1+h_2} + \frac{\varphi(t-h_2) - \varphi(t)}{-h_2}\frac{h_2}{h_1+h_2} \\ &= (f(t) + o_t(h_1))\frac{h_1}{h_1+h_2} + (f(t) + o_t(-h_2))\frac{h_2}{h_1+h_2} \\ &= f(t) + \left(o_t(h_1)\frac{h_1}{h_1+h_2} + o_t(-h_2)\frac{h_2}{h_1+h_2}\right),\end{aligned}$$

where, by Lebesgue differentiation theorem, o_t is an infinitesimal a.e.$[\ell]$ (this means that the set of t's such that o_t is not an infinitesimal has zero Lebesgue measure).

Since $0 \le \frac{h_1}{h_1+h_2} \le 1$ and $0 \le \frac{h_2}{h_1+h_2} \le 1$, we have

$$\lim_{\substack{h_1\to 0 \\ h_2\to 0}} \frac{\varphi(t+h_1) - \varphi(t-h_2)}{h_1+h_2} = f(t) + 0 \ \text{ a.e.}[\ell].$$

Thus, $p_N(t) = \frac{\mathrm{d}EN}{\mathrm{d}\ell}$ a.e.$[\ell]$. ■

Assumption A. A point process N satisfies Assumption A when it is under Assumption B and the equality

$$E(N \times N)(A \cap D) = EN\pi_1(A \cap D)$$

holds for all $A \in \Lambda_{\mathbb{R}^2}$, where D is the diagonal set of $\mathbb{R}^2$ and π_1 is the first canonical projection.

We observe that this condition is equivalent to say that the measure $E(N \times N)$ restricted to diagonal, $E(N \times N)|_D : \Lambda_D \to \mathbb{R}$, is the induced measure over the diagonal by the measure EN over the straight line through π_1, that is, $E(N \times N)|_D = EN\pi_1$.

Definition 3.1. *A point process is called non-internally correlated (NIC) if and only if for all A and B disjoint Lebesgue measurable sets we have $Cov(N(A), N(B)) = 0$.*

Clearly, Poisson processes are particular cases of NIC point processes.For Poisson processes, complete independence of the random variables $N(A_1), \dots, N(A_k)$, for all $k \in \mathbb{N}^*$, is assumed, where $A_1, \dots, A_k$ are disjoint measurable sets, while for a NIC point process we only need to assume zero covariance for all pairs of random variables $N(A_1), N(A_2)$.

For point processes satisfying Assumption B, we have the following proposition.

Proposition 3.1. *If N satisfies Assumption B then, for all $E\mathrm{d}N$-integrable function, φ, we have $\int \varphi \mathrm{d}EN = \int \varphi p_N \mathrm{d}t$.*

Proof Immediate, since $p_N = \mathrm{d}EN/\mathrm{d}\ell$ a.e.$[\ell]$. ■

For point processes satisfying Assumption A we have the following proposition.

Proposition 3.2. *If N satisfies Assumption A then, for all functions φ_1 integrable with respect to the covariance measure* $\mathrm{Cov}\,(N,N)$, *we have:*

$$\int \varphi_1 \mathrm{dCov}\,(N,N) = \int_{\mathbb{R}^2 - D} \varphi_1 \mathrm{dCov}\,(N,N) + \int_{\mathbb{R}} \varphi p_N \mathrm{d}t, \quad \varphi(t) = \varphi_1(t,t).$$

Proof It is enough to prove that $\int_D \varphi_1 \mathrm{dCov}\,(N,N) = \int_{\mathbb{R}} \varphi p_N \mathrm{d}t$.

$$\begin{aligned}
\int_D \varphi_1 \mathrm{dCov}\,(N,N) &= \int_D \varphi_1 \mathrm{d}(E(N \times N) - EN \times EN) \\
&= \int_D \varphi_1 \mathrm{d}E(N \times N) - \int_D \varphi_1 \mathrm{d}(EN \times EN) \\
&= \int_D \varphi\pi_1 \mathrm{d}E(N \times N) - \int_D \varphi_1 \frac{\mathrm{d}EN}{\mathrm{d}\ell} \otimes \frac{\mathrm{d}EN}{\mathrm{d}\ell} \mathrm{d}\ell \times \mathrm{d}\ell \\
&= \int_D \varphi\pi_1 \mathrm{d}(EN\pi_1) - 0 = \int_{\pi_1(D)} \varphi \mathrm{d}EN = \int_{\mathbb{R}} \varphi p_N \mathrm{d}t
\end{aligned}$$

since $\ell_2(D) = \ell \times \ell(D) = 0$. ■

We will also write

$$\int \varphi(t) p_N(t) \mathrm{d}t = \int \varphi(t) \mathrm{Var}\,(\mathrm{d}N(t))$$

where the right hand side means $\iint_D \varphi_1(u,v) \mathrm{Cov}\,(\mathrm{d}N(u), \mathrm{d}N(v))$, D being diagonal set of $\mathbb{R}^2$ and $\varphi(t) = \varphi_1(t,t)$.

The following Proposition is useful for the calculation of covariances of random variables associated to point process that are written as integrals.

Proposition 3.3. *Let X and Y be random variables defined by the stochastic integrals $X = \int_A f\mathrm{d}N$ and $Y = \int_B g\mathrm{d}N$, D diagonal set of $\mathbb{R}^2$, π_1 the first canonical projection and $A, B \in \Lambda_{\mathbb{R}}$ such that* $(\mathrm{supp}\, f \cap A) \times (\mathrm{supp}\, g \cap B)$ *is bounded. For N under Assumption A we have*

$$\mathrm{Cov}\,(X,Y) = \int_{(A\times B)-D} f \otimes g \;\; \mathrm{Cov}\,(\mathrm{d}N, \mathrm{d}N) + \int_{\pi_1((A\times B)\cap D)} f g p_N \mathrm{d}\ell.$$

If $\mathrm{Cov}\,(\mathrm{d}N, \mathrm{d}N) \ll \mathrm{d}\ell \times \mathrm{d}\ell$, *i.e., there exists* $q_2 \in \overline{\mathcal{L}}^2$, $\mathrm{dCov}\,(N,N) = q_2(u,v)\mathrm{d}u\mathrm{d}v$,

$$\mathrm{Cov}\,(X,Y) = \int_{(A\times B)-D} f(u)g(v)q_2(u,v)\mathrm{d}u\mathrm{d}v + \int_{\pi_1((A\times B)\cap D)} f(t)g(t)p_N(t)\mathrm{d}t.$$

If N is NIC then

$$\mathrm{Cov}\,(X,Y) = \int_{\pi_1((A\times B)\cap D)} f(t)g(t)p_N(t)\mathrm{d}t.$$

Proof Since

$$E(XY) = E\left(\iint_{A\times B} f(u)g(v)\mathrm{d}N(u)\mathrm{d}N(v)\right) = \iint_{A\times B} f(u)g(v)E(\mathrm{d}N(u)\mathrm{d}N(v))$$

and also

$$E(X)E(Y) = \int_A f(u)E\mathrm{d}N(u)\int_B g(v)E\mathrm{d}N(v) = \iint_{A\times B} f(u)g(v)E\mathrm{d}N(u)E\mathrm{d}N(v),$$

we have

$$\begin{aligned}\mathrm{Cov}\,(X,Y) &= \iint_{A\times B} f(u)g(u)\,[E(\mathrm{d}N(u)\mathrm{d}N(v)) - E\mathrm{d}N(u)E\mathrm{d}N(v)] \\ &= \iint_{A\times B} f(u)g(v)\mathrm{Cov}\,(\mathrm{d}N(u),\mathrm{d}N(v)).\end{aligned}$$

Thus by Proposition 3.2 it follows that

$$\begin{aligned}\mathrm{Cov}\,(X,Y) &= \iint_{A\times B-D} f(u)g(v)\mathrm{Cov}\,(\mathrm{d}N(u),\mathrm{d}N(v)) \\ &\quad + \int_{\pi_1((A\times B)\cap D)} f(t)g(t)\mathrm{Var}\,(\mathrm{d}N(t)) \\ &= \iint_{A\times B-D} f(u)g(v)q_2(u,v)\mathrm{d}u\mathrm{d}v + \int_{\pi_1((A\times B)\cap D)} f(t)g(t)p_N(t)\mathrm{d}t.\end{aligned}$$

If N is NIC, then $q_2(u,v) = 0$ and the proposition is established. ■

Observe that, since Poisson processes are special cases of NIC point processes, the third equality above is fulfilled for Poisson processes.

Assumption B and Assumption A are suitable for immediate generalization for point process on $\mathbb{R}^m$. We have proved (see de Miranda, 2003b) that for $m = 1$ these assumptions are equivalent.

3.2. **Sure Inference Analysis.** Let us assume that $X : \Omega \to \mathbb{R}$ is an unbiased estimator for x and that $\mathrm{Var}\,(X) = \sigma_1^2$. Suppose also that we have sequences of non-negative estimators and finite variances, respectively, $\hat{V}_n$ and V_n for all $n \geq 1$ such that $V_1 = \mathrm{Var}\,(X)$, $V_{n+1} = \mathrm{Var}\,\hat{V}_n$ and $E\hat{V}_n = V_n$. Then, by Chebychev's inequality we will have for $\lambda_1 > 0$, $P\{X(\omega) \in [x - \lambda_1\sigma_1, x + \lambda_1\sigma_1]\} \geq 1 - 1/\lambda_1^2$ and, equivalently, $P\{x \notin [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\} \leq 1/\lambda_1^2$. Let $\sigma_n = \sqrt{V_n}$ and $\hat{\sigma}_n = \sqrt{\hat{V}_n}$. We can write similarly $P\{V_n \notin [\hat{V}_n(\omega) - \lambda_{n+1}\sqrt{V_{n+1}}, \hat{V}_n(\omega) + \lambda_{n+1}\sqrt{V_{n+1}}]\} \leq 1/(\lambda_{n+1})^2$, for all $n \geq 1$. It may happen, as it is often in practice, that we do not know the value of σ_1 and use $\hat{\sigma}_1(\omega)$ and $X(\omega)$ to form confidence intervals for x when the distribution of $X : \Omega \to \mathbb{R}$ is known. In these situations, after some analysis, it could be concluded, "with probability p", that x belongs to the interval $[X(\omega) - \lambda_1\hat{\sigma}_1(\omega), X(\omega) + \lambda_1\hat{\sigma}_1(\omega)]$. We are interested in the situation where we do not know the distribution of X and we want to decrease the uncertainty due to the replacement of σ_1 by $\hat{\sigma}_1(\omega)$.

Since $P\{\sigma_1 \notin [\sqrt{\max\{0, \hat{V}_1(\omega) - \lambda_2\sqrt{V_2}\}}, \sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\,]\} \leq 1/\lambda_2^2$, we have

$$\begin{aligned} P\left\{\sigma_1 \leq \sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\right\} &= 1 - P\left\{\sigma_1 > \sqrt{\hat{V}_1(\omega) + \sqrt{V_2}}\right\} \\ &\geq 1 - P\left\{\sigma_1 \notin [\sqrt{\max\{0, \hat{V}_1(\omega) - \lambda_2\sqrt{V_2}\}}, \sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\,]\right\} \\ &\geq 1 - \frac{1}{\lambda_2^2}. \end{aligned}$$

Let $L(\omega, \lambda_1, \lambda_2) = \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}$, $A(\omega, \lambda_1, \lambda_2) = X(\omega) - L(\omega, \lambda_1, \lambda_2)$ and $B(\omega, \lambda_1, \lambda_2) = X(\omega) + L(\omega, \lambda_1, \lambda_2)$.

Moreover, let

$$\begin{aligned} \Omega^+ &= \{\omega \in \Omega | \sigma_1 \leq \sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\}, \\ \Omega^0 &= \{\omega \in \Omega | x \in [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\}, \\ \Omega^1 &= \{\omega \in \Omega | x \in [X(\omega) - L(\omega, \lambda_1, \lambda_2), X(\omega) + L(\omega, \lambda_1, \lambda_2)]\}. \end{aligned}$$

We then have $P(\Omega^0) \geq \left(1 - \frac{1}{\lambda_1^2}\right)$ and $P(\Omega^+) \geq \left(1 - \frac{1}{\lambda_2^2}\right)$.

Therefore, since $L(\omega, \lambda_1, \lambda_2) \geq \lambda_1\sigma_1$ when $\sigma_1 \leq \sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}$, it follows that $(\Omega^+ \cap \Omega^1) \supset (\Omega^+ \cap \Omega^0)$ and we can write

$$\begin{aligned} &P\{x \in [A(\omega, \lambda_1, \lambda_2), B(\omega, \lambda_1, \lambda_2)]\} = P(\Omega^1) \\ &\geq P(\Omega^1 \cap \Omega^+) \geq P(\Omega^0 \cap \Omega^+) \\ &\geq P(\Omega^0) + P(\Omega^+) - 1 \geq 1 - \frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}. \end{aligned}$$

The inequality above allows us to obtain conclusions such as: with at least probability $\left(1 - \frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right)$, x belongs to the interval $[X(\omega) - \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\,, X(\omega) + \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{V_2}}\,]$.

This interval can be replaced in practice by

$$\left[X(\omega) - \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{\hat{V}_2(\omega)}}\,, X(\omega) + \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{\hat{V}_2(\omega)}}\right]$$

and this replacement brings some uncertainty. This uncertainty is the reason why we will use " " and say that "with at least probability p", x belongs to the later interval above. We can continue the process of analyzing the worst case and get probabilities of the form $1 - \sum_{i=1}^{n} \frac{1}{\lambda_i^2}$ for intervals of the form$[X(\omega) - L_m(\omega, \lambda_1, \ldots, \lambda_m), X(\omega) + L_m(\omega, \lambda_1, \ldots, \lambda_m)]$ with

$$L_m(\omega, \lambda_1, \ldots, \lambda_m) = \lambda_1\sqrt{\hat{V}_1(\omega) + \lambda_2\sqrt{\cdots + \lambda_{m-1}\sqrt{\hat{V}_{m-1}(\omega) + \lambda_m\sqrt{V_m}}}}.$$

Definition 3.2. *The triple $(X, (V_n)_{n\in\mathbb{N}^*}, (\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a random variable $X : \Omega \to \mathbb{R}$, a sequence of positive numbers $(V_n)_{n\in\mathbb{N}^*}$ and a sequence of random variables $(\hat{V}_n : \Omega \to \mathbb{R})_{n\in\mathbb{N}^*}$, is an* **inferential sequence** *for $x \in \mathbb{R}$ if and only if the following are valid:*

(i) $EX = x$, $V_1 = \text{Var}(X)$,
(ii) $\forall n \in \mathbb{N}^*$ $V_{n+1} = \text{Var}(\hat{V}_n)$,
(iii) $\forall n \in \mathbb{N}^*$ $E\hat{V}_n = V_n$,
(vi) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\Omega) \subset \mathbb{R}_+$.

We will use the notation $(X,\ V_n,\ \hat{V}_n)$ to represent an inferential sequence and, occasionally, we will simply say that the sequences V_n and $\hat{V}_n$ form an inferential sequence for x. Observe that this definition implies the fact that all random variables, that is, X and $\hat{V}_n$, $n \geq 1$, have finite means and variances, which is a necessary condition to apply Chebychev's inequality for each of them.

Theorem 3.2. (On the inferential sequence of random variables.) *Let $(X,\ V_n,\ \hat{V}_n)$ be an inferential sequence for $x \in \mathbb{R}$ and $\sigma_n = \sqrt{V_n}$, for all $n \in \mathbb{N}^*$. If*

$$L_m(\omega, \lambda_1, \ldots, \lambda_m) = \lambda_1 \sqrt{\hat{V}_1(\omega) + \lambda_2 \sqrt{\cdots + \lambda_{m-1} \sqrt{\hat{V}_{m-1}(\omega) + \lambda_m \sqrt{V_m}}}},$$

$\lambda_i \in \mathbb{R}_+^*$ for $1 \leq i \leq m$, $m \in \mathbb{N}^*$, then

$$P\{x \in [X(\omega) - L_m(\omega, \lambda_1, \ldots, \lambda_m), X(\omega) + L_m(\omega, \lambda_1, \ldots, \lambda_m)]\} \geq 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}.$$

Proof By induction. If $m = 1$, then

$$P\{x \in [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\} = 1 - P\{x \notin [X(\omega) - \lambda_1\sigma_1, X(\omega) + \lambda_1\sigma_1]\} \geq 1 - \frac{1}{\lambda_1^2},$$

by Chebychev's inequality.

For the facility of notation, let $A_k(\omega) = X(\omega) - L_k(\omega, \lambda_1, \ldots, \lambda_k)$ and $B_k(\omega) = X(\omega) + L_k(\omega, \lambda_1, \ldots, \lambda_k)$.

Assuming that the statement is valid for $m-1$, $m \geq 2$, we have

$$\begin{aligned} P\{x \in [A_m(\omega), B_m(\omega)]\} &\geq P\{x \in [A_m(\omega), B_m(\omega)] \ \wedge\ V_{m-1} \leq \hat{V}_{m-1}(\omega) + \lambda_m \sqrt{V_m}\} \\ &\geq P\{x \in [A_{m-1}(\omega), B_{m-1}(\omega)] \ \wedge\ V_{m-1} \leq \hat{V}_{m-1}(\omega) + \lambda_m \sqrt{V_m}\} \end{aligned}$$

since $[A_{m-1}(\omega), B_{m-1}(\omega)] \subset [A_m(\omega), B_m(\omega)]$ when $V_{m-1} \leq \hat{V}_{m-1}(\omega) + \lambda_m\sqrt{V_m}$.

Thus, $P\{x \in [A_m(\omega), B_m(\omega)]\}$

$$\begin{aligned} &\geq P\{x \in [A_{m-1}(\omega), B_{m-1}(\omega)]\} + P\{V_{m-1} \leq \hat{V}_{m-1}(\omega) + \lambda_m \sqrt{V_m}\} - 1 \\ &\geq \left(1 - \sum_{i=1}^{m-1} \frac{1}{\lambda_i^2}\right) + \left(1 - \frac{1}{\lambda_m^2}\right) - 1 \ = \ 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}, \end{aligned}$$

since $P\{V_{m-1} \leq \hat{V}_{m-1}(\omega) + \lambda_m\sqrt{V_m}\} \ \geq$

$$1 - P\{V_{m-1} \notin [\hat{V}_{m-1}(\omega) - \lambda_m\sqrt{V_m}, \hat{V}_{m-1}(\omega) + \lambda_m\sqrt{V_m}\,]\,\} \geq 1 - \frac{1}{\lambda_m^2}.$$

In this way the statement is also valid for m and the induction is completed. ■

If we substitute $\hat{V}_m(\omega)$ by V_m some uncertainty will be introduced in our analysis. Of course, this kind of uncertainty can be eliminated if we know the value of σ_J, (V_J), for some J or some upper bound for σ_J.

We remark that with the use of this analysis we can obtain confidence intervals with at "least probability p" for x, whatever the distribution of X is. Moreover, this analysis is more conservative than the one made assuming that the distribution of X is known.

As random variables are estimators of real numbers, stochastic processes can be understood as estimators of functions.

If $X : \Omega \times \mathbb{R} \to \mathbb{R}$ is a stochastic process which is an unbiased estimator for the function $x : \mathbb{R} \to \mathbb{R}$, that is, such that EX and x are equal ($E(X(\omega, t)) = x(t)$ for all $t \in \mathbb{R}$), and we have sequences of non-negative estimators (stochastic process in this case) $\hat{V}_n : \Omega \times \mathbb{R} \to \mathbb{R}$, and variance functions, $V_n : \mathbb{R} \to \mathbb{R}$ for $n \geq 1$, such that

$$V_1 = \mathrm{Var}\,(X) : \begin{array}{l} \mathbb{R} \to \mathbb{R}\,, \\ t \to \mathrm{Var}\,(X(t) : \Omega \to \mathbb{R}) \end{array} \qquad V_{n+1} = \mathrm{Var}\,(\hat{V}_n) \text{ and } E\hat{V}_n = V_n,$$

We can develop a sure inference analysis to obtain "at least probability p" confidence bands in a completely similar way to that presented above for random variables.

From now on, I is simply an arbitrary set.

Definition 3.3. *The triple $(X, (V_n)_{n\in\mathbb{N}^*}, (\hat{V}_n)_{n\in\mathbb{N}^*})$ formed by a stochastic process $X : \Omega \times I \to \mathbb{R}$, a sequence of functions $(V_n : I \to \mathbb{R})_{n\in\mathbb{N}^*}$ and a sequence of stochastic processes $(\hat{V}_n : \Omega \times I \to \mathbb{R})_{n\in\mathbb{N}^*}$ is said to be an* **inferential sequence** *for $x : I \to \mathbb{R}$ if and only if:*

(i) $EX = x$, $V_1 = Var(X)$,
(ii) $\forall n \in \mathbb{N}^*$ $V_{n+1} = Var(\hat{V}_n)$,
(iii) $\forall n \in \mathbb{N}^*$ $E\hat{V}_n = V_n$,
(vi) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\Omega \times I) \subset \mathbb{R}_+$.

Theorem 3.3. (On the inferential sequence of stochastic processes.) *Let $(X, V_n, \hat{V}_n)$ be an inferential sequence for $x : I \to \mathbb{R}$. Let for all $m \in \mathbb{N}^*$, $L_m : \Omega \times I \times (\mathbb{R}_+^*)^m \to \mathbb{R}_+$ be given by*

$$L_m(\omega, t, \lambda_1, \ldots, \lambda_m) = \lambda_1 \sqrt{\hat{V}_1(\omega, t) + \lambda_2 \sqrt{\ldots + \lambda_{m-1} \sqrt{\hat{V}_{m-1}(\omega, t) + \lambda_m \sqrt{V_m(t)}}}},$$

then, for all $t \in I$ and all $m \in \mathbb{N}^*$, we have

$$P\{x(t) \in [X(\omega, t) - L_m(\omega, t, \lambda_1, \ldots, \lambda_m), X(\omega, t) + L_m(\omega, t, \lambda_1, \ldots, \lambda_m)]\} \geq 1 - \sum_{i=1}^{m} \frac{1}{\lambda_i^2}.$$

Proof It is sufficient to see that, for each fixed t, we have that

$$(X(t), (V_n(t))_{n\in\mathbb{N}^*}, (\hat{V}_n(t))_{n\in\mathbb{N}^*})$$

is an inferential sequence for $x(t)$. ∎

See de Miranda (2003c) for a detailed presentation of this subject.

4. Estimation of the Intensity

Let N be a point process over the measurable space $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, with unknown intensity function p_N.

Let $\{\psi_{i,j} : i, j \in \mathbb{Z}\}$ be an orthonormal wavelet basis of the form $\psi_{i,j}(t) = 2^{j/2}\psi(2^j t - i)$ or $\psi_{i,j}(t) = 2^{j/2}\psi(2^j t - iT)$ for some mother wavelet ψ obtained, if necessary by the composition of a standard wavelet with an affine transformation, such that its support is $[0, T]$. Let ϕ be the father wavelet corresponding to ψ.

Similarly, let $\{\phi_{k,\ell i}, \psi_{i,j} : i, k \in \mathbb{Z},\ j \geq \ell i, j,\ \ell i \in \mathbb{Z}\}$ be an orthonormal wavelet basis that contains all the scales beyond some fixed integer ℓi.

It is extremely pleasant to adopt the following notation. Let ${}_d\mathbb{Z} = \{z \in \mathbb{Z} : z \geq d\}, d \in \mathbb{Z} \cup \{-\infty\}$ and $Ze(\ell i) = \mathbb{Z} \cup (\mathbb{Z} \times_{\ell i} \mathbb{Z})$ if $\ell i \in \mathbb{Z}$. If $\ell i = -\infty$, then $Ze(\ell i) = \mathbb{Z}^2$.

Let us use Greek letters for indexes in $Ze(\ell i)$ and we shall write $\psi_\eta = \phi_{\eta,\ell i}$ if and only if $\eta \in \mathbb{Z}$ and $\psi_\eta = \psi_{i,j}$ if and only if $\eta = (i, j) \in \mathbb{Z}^2$.

Thus, the wavelet expansions

$$f(t) = \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} \delta_{ij}\psi_{i,j}(t)$$

and

$$f(t) = \sum_{k\in\mathbb{Z}} \gamma_k\phi_{k,\ell i}(t) + \sum_{i\in\mathbb{Z}} \sum_{j\in_{\ell i}\mathbb{Z}} \delta_{ij}\psi_{i,j}(t)$$

will be simply written

$$f = \sum_{\eta\in Ze(\ell i)} \alpha_\eta\psi_\eta,$$

with the coefficients α_η given by

$$\begin{aligned} \int_{-\infty}^{\infty} f\psi_\eta dt &= \int_{R} (\sum_{\xi} \alpha_\xi\psi_\xi)\psi_\eta dt = \sum_{\xi} \int_{R} \alpha_\xi\psi_\xi\psi_\eta dt \\ &= \sum_{\xi} \alpha_\xi < \psi_\xi, \psi_\eta >= \alpha_\eta. \end{aligned}$$

Our aim is to obtain the restriction of p_N to $[0, T]$ based on the points of a trajectory of the process that are contained in this interval. Define

$$p = \begin{cases} p_N & \text{if } t \in [0, T], \\ 0 & \text{otherwise.} \end{cases}$$

From now on we assume that $p \in L^2[0, T]$. Therefore for the wavelet expansion of p we have

(10)
$$p = \sum_{\eta} \beta_\eta\psi_\eta,$$

with

(11)
$$\beta_\eta = \int_{R} p\psi_\eta dt = \int_0^T p\psi_\eta dt.$$

The main purpose is to estimate p through the expansion (10) and for this we need to estimate the wavelet coefficients β_η given by (11).

We set $q = \mathrm{dCov}\,(N,N)/\mathrm{d}\ell$ if $\mathrm{Cov}\,(N,N) \ll \ell_2$. If we do not have $\mathrm{Cov}\,(N,N) \ll \ell_2$, we may replace $q_2(u,v)dudv$ by $\mathrm{dCov}\,(N,N)$ in the statements of the theorems and propositions that follow.

Although for point process on the real line Assumptions A and B are equivalent, we have made explicit in the theorems that follow, what assumption, in its defining form, was necessary to derive the claimed conclusions. This is so done because these theorems are suitable for direct generalization for $\mathbb{R}^m$ and, for point process on $\mathbb{R}^m$, this distinction between assumptions may be necessary. In this way we have made our choice to state the theorems in a form that they can be directly generalized.

4.1. Estimation of the Wavelet Coefficients. We propose the following estimator of β_η:

$$\hat{\beta}_\eta = \int_0^T \psi_\eta dN(t).$$

The main properties of this estimator are given in the following theorem.

Theorem 4.1. *If N satisfies Assumption B, then*

(i) *the estimator $\hat{\beta}_\eta$ is unbiased.*
If N satisfies Assumption A then

(ii) *for all η and ξ,*

$$Cov(\hat{\beta}_\eta, \hat{\beta}_\xi) = \int\int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)dudv + \int_0^T \psi_\eta(u)\psi_\xi(u)p(u)du, \tag{12}$$

where $C = [0,T]^2 - \{(x,x) \in \mathbb{R}^2 : 0 \le x \le T\}$.

(iii) *In particular,*

$$Var(\hat{\beta}_\eta) = \int\int_C \psi_\eta(u)\psi_\eta(v)q_2(u,v)dudv + \int_0^T \psi_\eta^2(u)p(u)du. \tag{13}$$

Proof. (i) Since

$$E(\hat{\beta}_\eta) = E\int_0^T \psi_\eta dN(t) = \int_0^T \psi_\eta p_N(t)dt = \int_0^T \psi_\eta p dt = \beta_\eta,$$

$\hat{\beta}_\eta$ is unbiased.

(ii) Apply proposition 3.3 for $X = \hat{\beta}_\eta$, $Y = \hat{\beta}_\xi$ and $A = B = [0,T]$.
(iii) Immediate from (ii). ∎

Assume that N is a NIC point process. In this case $q_2(u,v) = 0$ and (12) becomes

$$\mathrm{Cov}(\hat{\beta}_\eta, \hat{\beta}_\xi) = \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt = E\int_0^T \psi_\eta(t)\psi_\xi(t)dN(t) \tag{14}$$

and (13) reduces to

$$\text{(15)} \qquad \mathrm{Var}(\hat{\beta}_\eta) = \int_0^T \psi_\eta^2(t)p(t)dt = \int_0^T \psi_\eta^2(t)E\{dN(t)\} = E\int_0^T \psi_\eta^2(t)dN(t).$$

This leads us to propose the following expressions as estimators of (14) and (15),

$$\widehat{\mathrm{Cov}}(\hat{\beta}_\eta, \hat{\beta}_\xi) = \int_0^T \psi_\eta(t)\psi_\xi(t)dN(t)$$

and

$$\text{(16)} \qquad \widehat{\mathrm{Var}}(\hat{\beta}_\eta) = \int_0^T \psi_\eta^2(t)dN(t),$$

respectively, which are obviously unbiased.

Let us use the following notation for a sequence of estimators and variances:

$$V_{\xi,0} = \beta_\xi,\ \hat{V}_{\xi,0} = \hat{\beta}_\xi,\ V_{\xi,n+1} = \mathrm{Var}(\hat{V}_{\xi,n}),\ n \geq 0.$$

By direct substitution of (10) into (15) we obtain

$$\mathrm{Var}(\hat{\beta}_\xi) = \int_0^T \psi_\xi^2(t)\sum_\eta \beta_\eta\psi_\eta(t)dt = \sum_\eta \beta_\eta \int_0^T \psi_\xi^2(t)\psi_\eta(t)dt.$$

Defining

$$K_{\xi,2}^\eta = \int_0^T \psi_\xi^2(t)\psi_\eta(t)dt,$$

we have that (15) can be written as

$$\text{(17)} \qquad V_{\xi,1} = \mathrm{Var}(\hat{\beta}_\xi) = \sum_\eta \beta_\eta K_{\xi,2}^\eta.$$

Now, let us compute the variance of the estimator (16):

$$V_{\xi,2} \quad = \quad \mathrm{Var}(\hat{V}_{\xi,1}) = \mathrm{Var}\left(\int_0^T \psi_\xi^2(t)dN(t)\right).$$

Thus, by Proposition 3.3 with $f = g = \psi_\xi^2$ we have

$$V_{\xi,2} = \int\int_C \psi_\xi^2(u)\psi_\xi^2(v)q_2(u,v)dudv + \int_0^T \psi_\xi^4(u)p(u)du.$$

Since $q_2(u,v) = 0$ for a NIC point process, we have

$$V_{\xi,2} = \int_0^T \psi_\xi^4(t)p(t)dt = \int_0^T \psi_\xi^4(t)EdN(t) = E\int_0^T \psi_\xi^4(t)dN(t).$$

As before, define

$$\hat{V}_{\xi,2} = \int_0^T \psi_\xi^4(t)dN(t),$$

which is an unbiased estimator of $V_{\xi,2}$. If we write

$$K^{\eta}_{\xi,4} = \int_0^T \psi_\xi^4(t)\psi_\eta(t)dt,$$

we have, similarly to $V_{\xi,1}$,

$$V_{\xi,2} = \sum_\eta \beta_\eta K^{\eta}_{\xi,4}.$$

Defining

$$K^{\eta}_{\xi,m} = \int_0^T \psi_\xi^m(t)\psi_\eta(t)dt,$$

$$K_{\xi,m} = \int_0^T \psi_\xi^m(t)dN(t),$$

we get the following result.

Theorem 4.2. *If N is a NIC point process, satisfying Assumption A, then*

$$V_{\xi,n} = \sum_\eta \beta_\eta K^{\eta}_{\xi,2^n} \quad n \geq 1, \quad \hat{V}_{\xi,n} = K_{\xi,2^n},\ n \geq 0,$$

are sequences such that $\hat{V}_{\xi,n}$ is an unbiased estimator of $V_{\xi,n}$, $V_{\xi,n+1} = Var(\hat{V}_{\xi,n})$ and $\hat{V}_{\xi,0} = \hat{\beta}_\xi$.

Proof First let us prove that $\hat{V}_{\xi,0} = \hat{\beta}_\xi$ and that the estimators are unbiased. For $n = 0$, since

$$\hat{V}_{\xi,0} = K_{\xi,1} = \int_0^T \psi_\xi(t)dN(t) = \hat{\beta}_\xi$$

we have that

$$E(\hat{V}_{\xi,0}) = E\int_0^T \psi_\xi(t)dN(t) = \int_0^T \psi_\xi(t)p(t)dt =$$

$$\int_{\mathbf{R}} \psi_\xi(t)p(t)dt = \int_{\mathbf{R}} \psi_\xi(t)\sum_\eta \beta_\eta\psi_\eta(t)dt = \sum_\eta \beta_\eta \int_{\mathbf{R}} \psi_\eta\psi_\xi dt =$$

$$\sum_\eta \beta_\eta < \psi_\xi, \psi_\eta > = \beta_\xi = V_{\xi,0}.$$

For $n \geq 1$ we obtain

$$E(\hat{V}_{\xi,n}) = E(K_{\xi,2^n}) = E\int_0^T \psi_\xi^{2^n}(t)dN(t) =$$

$$\int_0^T \psi_\xi^{2^n}(t)p(t)dt = \int_0^T \psi_\xi^{2^n}(t)\sum_\eta \beta_\eta\psi_\eta(t)dt =$$

$$\sum_\eta \beta_\eta \int_0^T \psi_\xi^{2^n}(t)\psi_\eta(t)dt = \sum_\eta \beta_\eta K^{\eta}_{\xi,2^n} = V_{\xi,n}.$$

We turn now to the sequence of variances. For $n = 0$ we have from $\hat{V}_{\xi,0} = \hat{\beta}_\xi$ and (17) that

$$V_{\xi,1} = \sum_\eta \beta_\eta K^\eta_{\xi,2} = \mathrm{Var}(\hat{\beta}_\xi) = \mathrm{Var}(\hat{V}_{\xi,0}).$$

For $n \geq 1$, using proposition 3.3 with $f = g = \psi_\xi^{2^n}$, we write

$$\mathrm{Var}\left(\hat{V}_{\xi,n}\right) = \mathrm{Var}\left(K_{\xi,2^n}\right) = \mathrm{Var}\left(\int_0^T \psi_\xi^{2^n}(t)dN(t)\right)$$

$$= \int\int_C \psi_\xi^{2^n}(u)\psi_\xi^{2^n}(v)q_2(u,v)dudv + \int_0^T \psi_\xi^{2^{n+1}}(t)p(t)dt.$$

Since $q_2(u,v) = 0$, we obtain

$$\mathrm{Var}(\hat{V}_{\xi,n}) = \int_0^T \psi_\xi^{2^{n+1}}(t)p(t)dt = \int_0^T \psi_\xi^{2^{n+1}}(t)\sum_\eta \beta_\eta\psi_\eta(t)dt =$$

$$\sum_\eta \beta_\eta \int_0^T \psi_\xi^{2^{n+1}}(t)\psi_\eta(t)dt = \sum_\eta \beta_\eta K^\eta_{\xi,2^{n+1}} = V_{\xi,n+1}.$$

■

We remark that for all n and ξ, $V_{\xi,n+1}$ is finite, due to the essentially boundedness of ψ_ξ as well as compactness of its support.

Theorem 4.3. (Inferential sequence for the wavelet coefficients.) *Under the condition of Theorem 4.2, we have: for all $\xi \in Ze(\ell i)$ the sequences $\hat{V}_{\xi,n}$ and $V_{\xi,n}$, $n \geq 0$ form an inferential sequence of random variables for β_ξ.*

Proof Given the preceding theorem, it is enough to prove that $\forall \xi \in Ze(\ell i)$ $\forall n \in \mathbb{N}^*$ $\hat{V}_{\xi,n}(\Omega) \subset \mathbb{R}_+$. Since $\forall \omega \in \Omega$

$$\hat{V}_{\xi,n}(\omega) = \left(\int_0^T \psi_\xi^{2^n}(t)dN(t)\right)(\omega) = \int_0^T \psi_\xi^{2^n}(t)dN_\omega(t) \geq 0\ ,$$

the theorem follows. ■

Therefore, in the case of a NIC point process N, the estimators for β_ξ and the respective and successive variances are easy to compute, being all of the form $\int_0^T \psi_\xi^{2^n}(t)dN(t)$, and for a particular trajectory with m points in the interval $[0,T]$, at times $\tau_0, \tau_1, \dots, \tau_{m-1}$, this expression reduces to $\sum_{i=0}^{m-1} \psi_\xi^{2^n}(\tau_i)$.

In order to obtain the successive variances, for simulation purposes in an actual problem, it may be be necessary to know the values $K^\eta_{\xi,2^n}$, which depend on the particular wavelet family used.

For the Haar family, the following result is valid. The proof will not be given and it is available from the authors. See de Miranda (2003a).

Proposition 4.1. *For the Haar wavelets re-scaled to the interval* $[0,T]$ *with* $\|\psi_{(0,0)}\| = 1$, *the* $K^{\eta}_{\xi,2^n}$ *are written as:*

(i) *If* $\xi, \eta \in \mathbb{Z} \times_{\ell i} \mathbb{Z}$, $\xi = (x_1, y_1)$, $\eta = (x_2, y_2)$ *then:*
For $n > 0$ *we have*
(a) *If* $y_1 < 0$ *and* $y_2 < 0$, *then*

$$K^{\eta}_{\xi,2^n} = \begin{cases} 2^{(2^{n-1}y_1 + y_2/2)} T^{-(2^{n-1}-1/2)} & \text{when } x_1 = x_2 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

(b) *If* $y_1 \geq 0$ *or* $y_2 \geq 0$, *then*

$$K^{\eta}_{\xi,2^n} = \begin{cases} 2^{((2^{n-1}-1)y_1 + y_2/2)} T^{-(2^{n-1}-1/2)} & \text{whenever } A \text{ or } B \text{ are valid,} \\ -2^{((2^{n-1}-1)y_1 + y_2/2)} T^{(2^{n-1}-1/2)} & \text{whenever } C \text{ is valid,} \\ 0 & \text{otherwise,} \end{cases}$$

where A *is* $(y_1 \geq 0 > y_2 \wedge 0 \leq x_1 \leq 2^{y_1} - 1)$,
B *is* $(y_1 > y_2 \geq 0 \wedge 0 \leq x_2 \leq 2^{y_2} - 1 \wedge 2^{y_1 - y_2} x_2 \leq x_1 < 2^{y_1 - y_2} + 2^{y_1 - y_2 - 1})$
and C *is* $(y_1 > y_2 \geq 0 \wedge 0 \leq x_2 \leq 2^{y_2} - 1 \wedge 2^{y_1 - y_2} x_2 + 2^{y_1 - y_2 - 1} \leq x_1 < 2^{y_1 - y_2}(x_2 + 1))$.
For $n = 0$ *we have*
(a) *If* $y_1 \geq 0$ *or* $y_2 \geq 0$ *then*

$$K^{\eta}_{\xi,1} = \begin{cases} \delta_{\xi,\eta} & \text{if } 0 \leq x_1 \leq 2^{y_1} - 1, \\ 0 & \text{otherwise.} \end{cases}$$

(b) *If* $y_1 < 0$ *and* $y_2 < 0$, *then*

$$K^{\eta}_{\xi,1} = \begin{cases} 2^{(y_1 + y_2)/2} & \text{if } x_1 = x_2 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

(ii) *If* $\xi, \eta \in \mathbb{Z}$, *then*
(a) *For* $\ell i \geq 0$,

$$K^{\eta}_{\xi,2^n} = \begin{cases} \delta_{\xi,\eta} \left(\dfrac{2^{\ell i}}{T} \right)^{\left(\frac{2^n - 1}{2}\right)} & \text{when } 0 \leq \eta \leq 2^n - 1, \\ 0 & \text{othwerwise.} \end{cases}$$

(b) *For* $\ell i < 0$,

$$K^{\eta}_{\xi,2^n} = \delta_{\xi,\eta} \delta_{\eta,0} \left(\frac{2^{\ell i}}{T} \right)^{\frac{2^n + 1}{2}} T.$$

(iii) *If* $\xi = x_1 \in \mathbb{Z}$ *and* $\eta = (x_2, y_2) \in \mathbb{Z} \times_{\ell i} \mathbb{Z}$, *then*

$$K^{\eta}_{\xi,2^n} = \begin{cases} 2^{((2^{n-1}\ell i) + y_2/2)} T^{-(2^n - 1/2)} & \text{when } \ell i \leq y_2 < 0 \text{ and } x_1 = 0, \\ 0 & \text{otherwise.} \end{cases}$$

(iv) *If* $\xi = (x_1, y_1) \in \mathbb{Z} \times_{\ell i} \mathbb{Z}$ *and* $\eta = x_2 \in \mathbb{Z}$, *then*
(a) *For* $n \geq 1$ *we have*

$$K^{\eta}_{\xi,2^n} = \begin{cases} 2^{(2^{n-1}y_1+\ell i/2)}T^{-(2^{n-1}-1/2)} & \text{for } y_1 < 0 \text{ and } x_1 = x_2 = 0, \\ 2^{((2^{n-1}-1)y_1+\ell i/2)}T^{-(2^{n-1}-1/2)} & \text{for } (y_1 \geq 0, 0 \leq x_1 \leq 2^{y_1}-1, \\ & \text{and } x_2 = 0) \text{ or } (\ell i \geq 0, 0 \leq \eta \leq 2^{\ell i}-1 \\ & \text{and } 2^{y_1-\ell i}\eta \leq x_1 < 2^{y_1-\ell i}(\eta+1)), \\ 0 & \text{otherwise.} \end{cases}$$

(b) *For* $n = 0$, *we have*

$$K^{\eta}_{\xi,2^n} = \begin{cases} \delta_{\eta,0}\delta_{x_1,0}2^{(y_1+\ell i)/2} & \text{if } y_1 < 0, \\ 0 & \text{otherwise.} \end{cases}$$

4.2. **Estimation of the Intensity Function.** We are now in position to estimate the intensity function p through a synthesis procedure using the estimates of the wavelet coefficients.

Theorem 4.4. *Let* $\hat{p} = \sum_{\eta \in Ze(\ell i)} \hat{\beta}_\eta \psi_\eta$.

If N satisfies Assumption B , then

(i) *the function* $\hat{p}$ *is an unbiased estimator for the intensity function* p.
If N satisfies Assumption A, then
(ii) *the variance of* $\hat{p}$ *is given by*

$$Var(\hat{p}) = \sum_{\eta,\xi} \left(\int\int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)dudv + \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt \right) \psi_\eta\psi_\xi.$$

If N satisfies Assumption A and it is a NIC point process, then

(iii)
$$\text{Var}\,(\hat{p}) = \sum_{\eta,\xi} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt \right) \psi_\eta\psi_\xi,$$

(iv) *and an unbiased estimator for* $Var(\hat{p})$ *is*

$$\widehat{Var}(\hat{p}) = \sum_{\eta,\xi} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)dN(t) \right) \psi_\eta\psi_\xi.$$

Proof (i) Since E is a continuous linear functional,

$$E(\hat{p}) = E(\sum_\eta \hat{\beta}_\eta\psi_\eta) = \sum_\eta \beta_\eta\psi_\eta = p.$$

(ii) $\mathrm{Var}(\hat{p}) = E(\sum_\eta(\hat{\beta}_\eta - \beta_\eta)\psi_\eta)^2 = E\left(\sum_\xi \sum_\eta(\hat{\beta}_\eta - \beta_\eta)(\hat{\beta}_\xi - \beta_\xi)\psi_\eta\psi_\xi\right) =$

$= \sum_\xi \sum_\eta \mathrm{Cov}(\hat{\beta}_\eta, \hat{\beta}_\xi)\psi_\eta\psi_\xi.$

Therefore, in the general case,

$$\mathrm{Var}\,(\hat{p}) = \sum_\eta \sum_\xi \left(\int\int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)dudv + \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt \right) \psi_\eta\psi_\xi.$$

(iii) For a NIC point process, since $q_2(u,v) = 0$, the above expression reduces to the sum of the second term inside the parentheses.

(iv) Immediate, since $p(t)dt = EdN(t)$. ■

Inferential sequences for p can be obtained using the result of the following theorem.

Theorem 4.5. (Inferential Sequence for the Intensity.) *Let $\eta = (\eta_1, \ldots, \eta_{2^n}) \in (Ze(\ell i))^{2^n}$ be an element of the cartesian product of $Ze(\ell i)$ by itself 2^n times, and N a NIC point process that satisfies Assumption A. Let*

$$V_n(\hat{p}) = \sum_{\eta \in (Ze(\ell i))^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} p dt \right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}$$

and

$$\hat{V}_n(\hat{p}) = \sum_{\eta \in (Ze(\ell i))^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} dN \right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}, \text{ for all } n \geq 1.$$

Then $V_n(\hat{p})$ and $\hat{V}_n(\hat{p})$ are sequences of variances and estimators, respectively, such that:

(i) $E(\hat{p}) = p$, $V_1(\hat{p}) = \mathrm{Var}\,(\hat{p})$.
(ii) $\forall n \in \mathbb{N}^*$ $V_{n+1}(\hat{p}) = Var(\hat{V}_n(\hat{p}))$.
(iii) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\hat{p})$ *is an unbiased estimator for* $V_n(\hat{p})$.
(iv) $\forall n \in \mathbb{N}^*$ $\hat{V}_n(\hat{p})(\Omega \times [0,T]) \subset \mathbb{R}_+$

That is, $(\hat{p},\ V_n(\hat{p}),\ \hat{V}_n(\hat{p})\ )$ is an inferential sequence of stochastic processes for the intensity p.

Proof (i) Immediate.
(ii) Since E is a linear continuous functional, we have

$$\mathrm{Var}(\hat{V}_n(\hat{p})) = \mathrm{Var}\left(\sum_{\eta \in (Ze(\ell i))^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} dN \right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} \right) =$$

$$\sum_{\eta,\xi \in (Ze(\ell i))^{2^n}} \mathrm{Cov}\left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} dN, \int_0^T \prod_{m=1}^{2^n} \psi_{\xi_m} dN \right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} \prod_{m=1}^{2^n} \psi_{\xi_m}.$$

Using Proposition 3.3 we have

$$\mathrm{Var}(\hat{V}_n(\hat{p})) = \sum_{\eta,\xi\in(Ze(\ell i))^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} \prod_{m=1}^{2^n} \psi_{\xi_m} p dt\right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} \prod_{m=1}^{2^n} \psi_{\xi_m} =$$

$$= \sum_{\mu\in(Ze(\ell i))^{2^{n+1}}} \left(\int_0^T \prod_{\ell=1}^{2^{n+1}} \psi_{\mu_\ell} p dt\right) \prod_{\ell=1}^{2^{n+1}} \psi_{\mu_\ell} = V_{n+1}(\hat{p}).$$

(iii) Equality $E\hat{V}_n(\hat{p}) = V_n(\hat{p})$ follows from the linearity and continuity of E, Campbell's theorem and Proposition 3.1.
(iv) Since $\forall n \in \mathbb{N}^*$, $\forall \omega \in \Omega$, $\forall t \in [0,T]$,

$$\hat{V}_n(\hat{p})(\omega,t) = \sum_{\eta\in(Ze(\ell i))^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} dN_\omega\right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}(t) =$$

$$= \sum_{\eta\in(Ze(\ell i))^{2^n}} \int_0^T \left(\prod_{\ell=1}^{2^n} \psi_{\eta_\ell}\right) \left(\prod_{j=1}^{2^n} \psi_{\eta_j}(t)\right) dN_\omega = \sum_{\eta\in(Ze(\ell i))^{2^n}} \int_0^T \left(\prod_{\ell=1}^{2^n} \psi_{\eta_\ell}\psi_{\eta_\ell}(t)\right) dN_\omega =$$

$$= \int_0^T \sum_{\eta\in(Ze(\ell i))^{2^n}} \left(\prod_{\ell=1}^{2^n} \psi_{\eta_\ell}\psi_{\eta_\ell}(t)\right) dN_\omega =$$

$$= \int_0^T \sum_{\eta,\xi\in(Ze(\ell i))^{2^{n-1}}} \left(\prod_{\ell=1}^{2^{n-1}} \psi_{\eta_\ell}\psi_{\eta_\ell}(t)\right) \left(\prod_{m=1}^{2^{n-1}} \psi_{\xi_m}\psi_{\xi_m}(t)\right) dN_\omega =$$

$$= \int_0^T \sum_{\eta\in(Ze(\ell i))^{2^{n-1}}} \left(\prod_{\ell=1}^{2^{n-1}} \psi_{\eta_\ell}(t)\psi_{\eta_\ell}\right) \sum_{\xi\in(Ze(\ell i))^{2^{n-1}}} \left(\prod_{m=1}^{2^{n-1}} \psi_{\xi_m}(t)\psi_{\xi_m}\right) dN_\omega =$$

$$= \int_0^T \left(\sum_{\eta\in(Ze(\ell i))^{2^{n-1}}} \prod_{\ell=1}^{2^{n-1}} \psi_{\eta_\ell}(t)\psi_{\eta_\ell}\right)^2 dN_\omega \geq 0\,,$$

the theorem is proved. ■

5. Estimation Under Thresholding

Let ${}_d\mathbb{Z}_e = \{z \in \mathbb{Z} | d \leq z \leq e\}, d, e \in \mathbb{Z} \cup \{-\infty\} \cup \{+\infty\}$; $Ze(\ell i)_J = \mathbb{Z} \cup (\mathbb{Z} \times_{\ell i} \mathbb{Z}_J)$ if $\ell i \in \mathbb{Z}$ and $Ze(\ell i)_J = \mathbb{Z} \times \{z \in \mathbb{Z} | z \leq J\}$ if $\ell i = -\infty$.
We shall use the notation

$$\hat{p}_J = \sum_{\eta\in Ze(\ell i)_J} \hat{\beta}_\eta \psi_\eta = \sum_{j\leq J} \hat{\beta}_\eta \psi_\eta, \tag{18}$$

for the estimated intensity function using wavelets ψ_η with scales up to the J-th order, $J \geq 0$, noticing that when η is an ordered pair it is represented by (i,j). Observe that if $J < \ell i$ then the intensity function is estimated only by wavelets $\phi_{\eta,\ell i}$ and if $\ell i = -\infty$ the expansion contains only re-scaled wavelets from the mother wavelet.

The notation

$$(19) \qquad \hat{p}^T_{J,\lambda} = \sum_{\eta \in Ze(\ell i)_J} T\left(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}\right)\hat{\beta}_\eta\psi_\eta = \sum_{j \leq J} T\left(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}\right)\hat{\beta}_\eta\psi_\eta$$

will be used for the estimated intensity formed from wavelet coefficients undergoing a tresholding procedure.

A threshold function $T : I\!R \times I\!R_+ \to I\!R$ is a measurable function such that $0 \leq T(x,y) \leq 1$ and, for each $y, T(x,y) = 1$ if $|x| \geq y, T(x,y)$ is non-decreasing over $[0,y]$ and non-increasing over $[-y,0]$.

We hope that the use of T as the extreme of the interval $[0,T]$ and as a threshold function $T(x,y)$ will bring no harm in understanding what follows.

Denote by $\mathrm{ess\,sup}_A f(\mathrm{ess\,inf}_A f)$ the essential supremum (infimum) of a function f defined on the set A. For the easy of notation we will write $\mathrm{ess\,sup}_\eta f$ instead of $\mathrm{ess\,sup}_{\mathrm{supp}\psi_\eta} f = \mathrm{ess\,sup}_{x \in \mathrm{supp}\psi_\eta} f(x)$, where supp ψ_η indicates the support of the wavelet ψ_η. If $f = \sum_{\eta \in Ze(\ell i)} \alpha_\eta\psi_\eta$, we shall write $f_J = \sum_{\eta \in Ze(\ell i)_J} \alpha_\eta\psi_\eta$ and $f^T_{J,\lambda} = \sum_{\eta \in Ze(\ell i)_J} T(\alpha_\eta, L(\lambda,\eta))\alpha_\eta\psi_\eta \quad L : \mathbb{R}_+ \times Ze(\ell i)_J \to \mathbb{R}_+$.

Definition 5.1. *We will say that the function f is essentially α-Hölderian in $A \subset I\!R$ if and only if there exist two constants, K and α, and a set $D \subset I\!R$, $\ell(D) = 0$, such that for all x and y in $A - D$ we have $|f(x) - f(y)| \leq K|x-y|^\alpha, \alpha > 0$.*

This definition can be extended immediately for a function $f : X \to Y$, where $(X, \mathcal{A}, \mu)$ is a measure space and $(X, d_1), (Y, d_2)$ are metric spaces.

Definition 5.2. *Define $\mathrm{ess\,lim}_{x \to y} f(x) = L$, when there exists a set $D \subset I\!R$, $\ell(D) = 0$, such that the limit, when $x \to y$, of the function $f|_{A-D}$ (restriction of f to $A - D$), is the real number L.*

Analogously, we can extend the concept of essential limit, define essentially differentiable functions, etc.

5.1. Convergence rate and bounds for the bias. The results that follow give upper bounds for the magnitude of the bias, measured in the L^2 norm, of the estimators (18) and (19), in the case of p being essentially α-Hölderian.

Theorem 5.1. *Let $\{\psi_\eta | \eta \in Ze(\ell i)\}$, $\ell i \leq 0$, $\ell i \in \mathbb{Z} \cup \{-\infty\}$ be an orthonormal wavelet basis such that supp $\psi_{(0,0)} = [0,T]$ and $\psi_{(0,0)}$ is essentially bounded.*

Let N be a point process satisfying Assumption B. Suppose that p, the intensity function of N restricted to $[0,T]$, is essentially α-Hölderian with constants K and $\alpha > 0$. Then,

$$(20) \qquad \|p - E(\hat{p}_J)\|^2 \leq \frac{K^2M^2|\mathrm{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1-2^{-2\alpha})}\left(\frac{1}{2^{2\alpha}}\right)^{J+1}\chi_{(0,1]}(\alpha),$$

for all $J \geq 0$, where $M = \max\{|\mathrm{ess\,inf}_{[0,T]}\psi_{(0,0)}|, \mathrm{ess\,sup}_{[0,T]}\psi_{(0,0)}\}$.

Proof For all p and $J \geq 0$ the following equality holds:

$$\|p - E(\hat{p}_J)\|^2 = \|\sum_{\eta} \beta_\eta \psi_\eta - E\sum_{j \leq J} \hat{\beta}_\eta \psi_\eta\|^2 =$$

$$\|\sum_{j \leq J}(\beta_\eta - E(\hat{\beta}_\eta))\psi_\eta + \sum_{\eta \in \mathbb{Z}^2, j > J} \beta_\eta \psi_\eta\|^2 = \|\sum_{\eta \in \mathbb{Z}^2, j > J} \beta_\eta \psi_\eta\|^2 = \sum_{\eta \in \mathbb{Z}^2, j > J} \beta_\eta^2.$$

Now, we will find for each η, lower and upper bounds for β_η. Let $\psi_\eta^+ = \max\{\psi_\eta, 0\}$, $\psi_\eta^- = \max\{-\psi_\eta, 0\}$ and supp $\psi_\eta = [a_\eta, b_\eta]$. Since p is non negative,

$$\beta_\eta = \int \psi_\eta p dt = \int \psi_\eta^+ p dt - \int \psi_\eta^- p dt$$

$$\leq \int \psi_\eta^+ \text{ess sup}_\eta p dt - \int \psi_\eta^- \text{ess inf}_\eta p dt =$$

$$= \int (\psi_\eta^+ - \psi_\eta^-) \text{ess inf}_\eta p dt + \int \psi_\eta^+ (\text{ess sup}_\eta p - \text{ess inf}_\eta p) dt =$$

$$= \text{ess inf}_\eta p \int \psi_\eta dt + (\text{ess sup}_\eta p - \text{ess inf}_\eta p) \int \psi_\eta^+ dt$$

$$\leq 0 + K(b_\eta - a_\eta)^\alpha (b_\eta - a_\eta) \text{ess sup}_\eta \psi_\eta^+ =$$

$$= K(b_\eta - a_\eta)^{\alpha+1} 2^{j/2} \text{ess sup}_{(0,0)} \psi_{(0,0)}.$$

Analogously,

$$\beta_\eta \geq \int \psi_\eta^+ \text{ess inf}_\eta p dt - \int \psi_\eta^- \text{ess sup}_\eta p dt$$

$$\geq 0 - K(b_\eta - a_\eta)^\alpha \int \psi_\eta^- dt \geq -K(b_\eta - a_\eta)^{\alpha+1} 2^{j/2} |\text{ess inf}_{(0,0)} \psi_{(0,0)}|.$$

Let $M = \max\{\text{ess sup}_{[0,T]} \psi_{(0,0)}, -\text{ess inf}_{[0,T]} \psi_{(0,0)}\}$. Then we can write

$$|\beta_\eta| \leq KM(b_\eta - a_\eta)^{\alpha+1} 2^{j/2} = KM\left(\frac{|\text{supp}\psi_{(0,0)}|}{2^j}\right)^{\alpha+1} 2^{j/2},$$

$$\beta_\eta^2 \leq K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)} 2^{-(2\alpha+1)j}.$$

Since the j-th scale has at most 2^j non null coefficients,

$$\sum_{\eta \in \mathbb{Z}^2, j > J} \beta_\eta^2 \leq \sum_{j > J} 2^j (K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)} 2^{-(2\alpha+1)j}) =$$

$$= K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)} \sum_{j > J} 2^{-2\alpha j} =$$

$$= K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)} \frac{(2^{-2\alpha})^{J+1}}{(1 - 2^{-2\alpha})}.$$

Now, if $\alpha > 1$ and $A \subset \mathbb{R}$ is a real line interval, every essentially α-Hölderian function is constant on $A-D$ since, being x and $y \in A-D$, $x<y$, we can write $|f(y)-f(x)| \leq \sum_{i=0}^{n-1} |f(x_{i+1})-f(x_i)| \leq \sum_{i=0}^{n-1} K(x_{i+1}-x_i)^\alpha \leq K(\max_{0\leq i\leq n-1}(x_{i+1}-x_i))^{\alpha-1} \sum_{i=0}^{n-1}(x_{i+1}-x_i) = K(\max_{0\leq i\leq n-1}(x_{i+1}-x_i))^{\alpha-1}(y-x)$, where $x = x_0 < x_1 < \ldots < x_n = y$ and for all i, $0 \leq i \leq n$, $x_i \in A-D$. Since $\forall \varepsilon > 0$ we can choose these points such that $\max_{0\leq i\leq n-1}(x_{i+1}-x_i) < \varepsilon$ (otherwise we would have an interval contained in D and this would not obey $\ell(D) = 0$) we have $\forall \varepsilon > 0$ $|f(y)-f(x)| < \varepsilon$ and hence $f(y) = f(x)$ for all x and y in $A-D$.

Since p is essentially constant, we have $\beta_\eta = \int \psi_\eta p dt = 0$ for all $J \geq 0$, hence

$$\|p - E(\hat{p}_J)\|^2 = \sum_{\eta\in\mathbb{Z}^2, j>J} \beta_\eta^2 = 0.$$

■

The preceding theorem guarantees at least an exponential decay with J for the bias of $\hat{p}_J$. The following two theorems show that in case of thresholding, the square of the bias is bounded by a sum of two parts. One corresponding to the exponencial decay with J and another corresponding to the threshold. In theorem 5.2 the expansion in made using wavelets re-scaled from a mother wavelet only. For this reason it was necessary to assume the existence of the essencial limit at zero for $\psi_{(0,0)}$. Theorem 5.3 assumes that the expansion is made using father and mother wavelets.

Theorem 5.2. *Under the conditions of Theorem 5.1 and the additional assumption that N is a NIC point process satisfying Assumption A , that $Ze(\ell i) = \mathbb{Z}^2$ and there exists* ess $\lim_{t\to 0}\psi_{(0,0)}(t) = L$, *we have that for every threshold function T and $\lambda \geq 0$,*

$$\text{(21)} \qquad \begin{aligned} \|p - E(\hat{p}^T_{J,\lambda})\|^2 \;\leq\; & \frac{K^2M^2|\text{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1-2^{-2\alpha})}\left(\frac{1}{2^{2\alpha}}\right)^{J+1}\chi_{(0,1]}(\alpha) \\ & + \lambda^2(k_1 + 2^{(J+1)} - 1)\operatorname*{ess\,sup}_{[0,T]} p, \end{aligned}$$

for some constant $k_1 \in \mathbb{R}$.

Proof We have that

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 = \|\sum_\eta \beta_\eta\psi_\eta - E(\sum_{j\leq J} T(\hat{\beta}_\eta, \lambda\sqrt{\text{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta\psi_\eta)\|^2 =$$

$$= \|\sum_{j>J}\beta_\eta\psi_\eta + \sum_{j\leq J}(\beta_\eta - E(T(\hat{\beta}_\eta, \lambda\sqrt{\text{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta))\psi_\eta\|^2,$$

therefore

$$\text{(22)} \qquad \|p - E(\hat{p}^T_{J,\lambda})\|^2 = \sum_{j>J}\beta_\eta^2 + \sum_{j\leq J}(\beta_\eta - E(T(\hat{\beta}_\eta, \lambda\sqrt{\text{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta))^2.$$

Now,

$$\beta_\eta - E(T(\hat{\beta}_\eta, \lambda\sqrt{\text{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta) = E(\hat{\beta}_\eta - T(\hat{\beta}_\eta, \lambda\sqrt{\text{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta).$$

Let $v = \hat{\beta}_\eta - T(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta$. It follows that v is a random variable such that $v = 0$ if $|\hat{\beta}_\eta| \geq \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}$, $0 \leq v \leq \hat{\beta}_\eta$ if $0 \leq \hat{\beta}_\eta < \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}$ and $\hat{\beta}_\eta \leq v \leq 0$ if $-\lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)} < \hat{\beta}_\eta \leq 0$.

Separating the integral $\int v dP$ into two integrals over the intervals $(-\lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}, 0]$ and $[0, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})$, we get

$$E(v) \leq \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}P([0, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})).$$

Similarly,

$$E(v) \geq -\lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}P((-\lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}, 0]).$$

It follows that

$$|E(v)|^2 \leq \lambda^2 \mathrm{Var}\,(\hat{\beta}_\eta) \max\{(P((-\lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)}, 0]))^2, (P([0, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})))^2\}$$

and we obtain

$$\sum_{\eta, j \leq J} (\beta_\eta - E(T(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}\,(\hat{\beta}_\eta)})\hat{\beta}_\eta))^2 \leq \sum_{\eta, j \leq J} \lambda^2 \mathrm{Var}\,(\hat{\beta}_\eta). \tag{23}$$

Since for NIC point processes,

$$\mathrm{Var}\,(\hat{\beta}_\eta) = \int \psi_\eta^2 p(t)dt \leq \text{ess sup}_\eta p \int \psi_\eta^2 dt =$$

$$= \text{ess sup}_\eta p \leq \text{ess sup}_{[0,T]} p,$$

we can write, for $\eta \in \mathbb{Z}^2$ and non negative m,

$$\sum_{\{j=m\}} \mathrm{Var}\,(\hat{\beta}_\eta) \leq \sum_{\{j=m\}} \text{ess sup}_\eta p \leq 2^m \text{ess sup}_{[0,T]} p,$$

due to the existence of exactly 2^m wavelets with scale m and with supports that are not disjoint of $[0, T]$. In this way,

$$\sum_{0 \leq j \leq J} \mathrm{Var}(\hat{\beta}_\eta) \leq \text{ess sup}_{[0,T]} p (\sum_{m=0}^{J} 2^m) = (2^{J+1} - 1) \text{ess sup}_{[0,T]} p. \tag{24}$$

For negative j, we are only interested in those η of the form $(0, j)$. This is so due to the fact that, for $\eta = (i, j)$, $i \neq 0$ we have $\mathrm{supp}\psi_{(i,j)} \cap [0, T] = \phi$. Since

$$\mathrm{Var}(\hat{\beta}_\eta) = \int \psi_\eta^2 p dt = \int_0^T \psi_\eta^2 p dt,$$

we obtain

$$\mathrm{Var}(\hat{\beta}_\eta) \leq \text{ess sup}_{[0,T]} p \int_0^T \psi_\eta^2 dt = \text{ess sup}_{[0,T]} p \int_0^T \{2^{j/2}\psi_{(0,0)}(2^j t - iT)\}^2 dt.$$

Since $i = 0$,

$$\text{Var}(\hat{\beta}_\eta) \leq \text{ess sup}_{[0,T]} p \int_0^T 2^j \psi^2_{(0,0)}(2^j t)dt = \text{ess sup}_{[0,T]} p \int_0^{T2^j} \psi^2_{(0,0)}(x)dx. \quad (25)$$

From the existence of the essential limit $\text{ess}\lim_{x\to 0^+} \psi_{(0,0)}(x) = L$, we have that there exists $D \subset I\!R$, $\ell(D) = 0$, such that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in [0,T] - D$ we have the implication $0 < x < \delta \to \psi_{(0,0)}(x) \in (L-\varepsilon, L+\varepsilon)$. So, for all j with $2^j T < \delta$ we have

$$\text{Var}(\hat{\beta}_\eta) \leq \text{ess sup}_{[0,T]} p \int_0^{T2^j} \psi^2_{(0,0)}(x)dx$$

$$\leq (\text{ess sup}_{[0,T]} p) 2^j T \max\{(L-\varepsilon)^2, (L+\varepsilon)^2\}.$$

Let j_* be the greatest integer j such that $2^j T < \delta$. Then

$$\sum_{\eta=(0,j), j\leq j_*} \text{Var}(\hat{\beta}_\eta) \leq 2^{j_*+1} (\text{ess sup}_{[0,T]} p) T \max\{(L-\varepsilon)^2, (L+\varepsilon)^2\}. \quad (26)$$

The inequalities (25) and (26) imply that

$$\sum_{\eta, j<0} \text{Var}(\hat{\beta}_\eta) \leq (\text{ess sup}_{[0,T]} p) 2^{j_*+1} T \max\{(L-\varepsilon)^2, (L+\varepsilon)^2\}$$

$$+ (\text{ess sup}_{[0,T]} p) \sum_{j=j_*+1}^{-1} \int_0^{2^j T} \psi^2_{(0,0)}(x)dx.$$

Let

$$k_1 = 2^{j_*+1} T \max\{(L-\varepsilon)^2, (L+\varepsilon)^2\} + \sum_{j=j_*+1}^{-1} \int_0^{2^j T} \psi^2_{(0,0)}(x)dx.$$

Then using (24) we write

$$\sum_{\eta, j\leq J} \lambda^2 \text{Var}(\hat{\beta}_\eta) = \lambda^2 \left(\sum_{0\leq j\leq J} \text{Var}(\hat{\beta}_\eta) + \sum_{j<0} \text{Var}(\hat{\beta}_\eta) \right)$$

$$\leq \lambda^2 (k_1 + 2^{J+1} - 1) \text{ess sup}_{[0,T]} p.$$

Finally, from Theorem 5.1, (22) and (23) we deduce

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 \leq \frac{K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1 - 1/2^{2\alpha})} \left(\frac{1}{2^{2\alpha}}\right)^{J+1} \chi_{(0,1]}(\alpha)$$

$$+ \lambda^2 (k_1 + 2^{J+1} - 1) \text{ess sup}_{[0,T]} p.$$

■

Theorem 5.3. *Let $\{\psi_\eta | \eta \in Ze(\ell i)\}$, $\ell i \in \mathbb{Z}$, $\ell i \leq 0$ be an orthonormal wavelet basis such that $\text{supp}\psi_{(0,0)} = [0,T]$ and $\psi_{(0,0)}$ is essentially bounded. Let N be a NIC point point process under Assumption A with p essentially α-Hölderian with constants K and $\alpha > 0$. Then, for every thresholding function T and $\lambda \geq 0$ we have*

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 \leq \frac{K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1-2^{-2\alpha})} \left(\frac{1}{2^{2\alpha}}\right)^{J+1} \chi_{(0,1]}(\alpha)$$

$$+\lambda^2(k_1 + 2^{J+1} - 1)\text{ess sup}_{[0,T]}p,$$

for some constant $k_1 \in \mathbb{R}$.

Proof It is enough to notice that, in this case,

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 = \sum_{\eta\in Ze(\ell i)-Ze(\ell i)_J} \beta_\eta^2 + \sum_{\eta\in Ze(\ell i)_J} E(v)^2 \leq \sum_{\eta\in Ze(\ell i)-Ze(\ell i)_J} \beta_\eta^2 + \sum_{\eta\in Ze(\ell i)_J} \lambda^2 \text{Var}(\hat{\beta}_\eta).$$

From $\sum_{\eta\in\mathbb{Z}} \text{Var}(\hat{\beta}_{\eta,\ell i}) + \sum_{\ell i\leq j\leq -1} \text{Var}(\hat{\beta}_{(0,j)}) + \sum_{j=0}^{J} \text{Var}(\hat{\beta}_{(i,j)})$

$$\leq \int_0^T \phi^2_{(0,\ell i)} p\, dt + (\text{ess sup}_{[0,T]}p) \left(\sum_{j=\ell i}^{-1} \int_0^{2^j T} \psi^2_{(0,0)}(x)dx + (2^{J+1} - 1) \right)$$

and $\int_0^T \phi^2_{(0,\ell i)} p\, dt \leq \text{ess sup}_{[0,T]}p \int_0^T \phi^2_{(0,\ell i)} dt =$

$$= \text{ess sup}_{[0,T]}p \int_0^T (2^{\ell i/2})^2 \phi^2_{(0,0)}(2^{\ell i}t)dt = \text{ess sup}_{[0,T]}p \int_0^{2^{\ell i}T} \phi^2_{(0,0)}(x)dx\ ,$$

by Theorem 5.1 we establish the inequality with

$$k_1 = \int_0^{2^{\ell i}T} \phi^2_{(0,0)}(x)dx + \sum_{j=\ell i}^{-1} \int_0^{2^j T} \psi^2_{(0,0)}(x)dx.$$

■

For Haar wavelets both Theorem 5.2 and 5.3 reduce to the following proposition.

Proposition 5.1. *Under the hypothesis of Theorem 5.2 or Theorem 5.3, for the Haar wavelet family, we have*

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 \leq \frac{K^2 T^{2\alpha+1}}{(1-2^{-2\alpha})} \left(\frac{1}{2^{2\alpha}}\right)^{J+1} \chi_{(0,1]}(\alpha) + \lambda^2 2^{(J+1)} \operatorname*{ess\,sup}_{[0,T]} p. \tag{27}$$

Proof For the Haar wavelets, $M = T^{-1/2} = L$. So for all $\varepsilon > 0$,

$$\|p - E(\hat{p}^T_{J,\lambda})\|^2 \leq \frac{K^2 T^{-1} T^{2(\alpha+1)}}{(1-1/2^{2\alpha})} \left(\frac{1}{2^{2\alpha}}\right)^{J+1} \chi_{(0,1]}(\alpha)$$

$$+\lambda^2\left(2^{j_*+1}T\max\{(T^{-1/2}+\varepsilon)^2,(T^{-1/2}-\varepsilon)^2\}\right)\text{ess sup}_{[0,T]}p$$

$$+\lambda^2\left(\sum_{j=j_*+1}^{-1}2^jTT^{-1}+2^{J+1}-1\right)\text{ess sup}_{[0,T]}p$$

$$=\frac{K^2T^{2\alpha+1}}{(1-1/2^{2\alpha})}\left(\frac{1}{2^{2\alpha}}\right)^{J+1}\chi_{[0,1]}(\alpha)$$

$$+\lambda^2\left(\sum_{j=j_*+1}^{-1}2^j+2^{j_*+1}\max\{(1+\varepsilon T^{1/2})^2,(1-\varepsilon T^{1/2})^2\}+2^{J+1}-1\right)\text{ess sup}_{[0,T]}p.$$

Letting $\varepsilon\to 0$ we get (27). Moreover if $\ell i\in\mathbb{Z}$ we have

$$k_1=\int_0^{2^{\ell i}T}\frac{1}{T}\mathrm{d}x+\sum_{j=\ell i}^{-1}\int_0^{2^jT}\frac{1}{T}\mathrm{d}x=\left(\sum_{j=\ell i}^{-1}2^j+2^{\ell i}\right)=\sum_{j<-1}2^j=1$$

and (27) follows.

■

Next proposition gives us a way to choose the "optimal" value for J.

Proposition 5.2. *Under the hypotheses of Theorems 5.2 or 5.3, given $\lambda\geq 0$, we can choose J such that*

$$\|p-E(\hat{p}^T_{J,\lambda})\|^2\leq g(\lambda)=\min(A,B),$$

with

$$A=k_0\left(\frac{1}{2^{2\alpha}}\right)^{\lfloor a\rfloor+1}\chi_{(0,1]}(\alpha)+\lambda^2(k_1+2^{\lfloor a\rfloor\chi_{(0,1]}(\alpha)+1}-1)\operatorname*{ess\,sup}_{[0,T]}p,$$

$$B=k_0\left(\frac{1}{2^{2\alpha}}\right)^{\lceil a\rceil+1}\chi_{(0,1]}(\alpha)+\lambda^2(k_1+2^{\lceil a\rceil\chi_{(0,1]}(\alpha)+1}-1)\operatorname*{ess\,sup}_{[0,T]}p,$$

$$k_0=\frac{K^2M^2|\text{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1-2^{-2\alpha})},$$

k_1 given in the proof of Theorem 5.2 or Theorem 5.3,

$$a=\frac{\ell n(2\alpha k_0/\text{ess sup}_{[0,T]}p)-2\ell n\lambda}{(2\alpha+1)\ell n2}-1,$$

and, in case of $\alpha\leq 1$, J will be $\lfloor a\rfloor$ or $\lceil a\rceil$, depending on which of these two values minimizes $g(\lambda)$; otherwise, that is, $\alpha>1$, J will be zero and $A=B=g(\lambda)=\lambda^2(k_1+1)\text{ess sup}_{[0,T]}p$.

Proof From Theorem 5.2 or Theorem 5.3, we have for $\alpha\in(0,1]$

$$\|p-E(\hat{p}^T_{J,\lambda})\|^2\leq k_0\left(\frac{1}{2^{2\alpha}}\right)^{J+1}+\lambda^2(k_1+2^{J+1}-1)\text{ess sup}_{[0,T]}p=f(J,\lambda).$$

Extend $f : \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}$ to $f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$. Then

$$\frac{\partial f}{\partial J}(J,\lambda) = k_0 \left(\ell n \left(\frac{1}{2^{2\alpha}}\right)\right)\left(\frac{1}{2^{2\alpha}}\right)^{J+1} + \lambda^2(\ell n 2)2^{J+1}\text{ess sup}_{[0,T]}p,$$

$$\frac{\partial^2 f}{\partial J^2}(J,\lambda) = k_0 \left(\ell n \left(\frac{1}{2^{2\alpha}}\right)\right)^2\left(\frac{1}{2^{2\alpha}}\right)^{J+1} + \lambda^2(\ell n 2)^2 2^{J+1}\text{ess sup}_{[0,T]}p,$$

hence the second derivative is positive for all J and λ. The first derivative is zero if and only if

$$(J+1)(2\alpha+1) = log_2 \left(\frac{2\alpha k_0}{\lambda^2 \text{ess sup}_{[0,T]}p}\right),$$

which yields the value of a above. Since, for a given λ, $f(J,\lambda)$ has only a critical point which is a minimum point, $f(J,\lambda)$ will assume its minimum value over the integers for $J = \lceil a \rceil$ or $J = \lfloor a \rfloor$. If $\alpha > 1$, then $\|p - E(\hat{p}^T_{J,\lambda})\|^2 \leq \lambda^2(k_1 + 2^{J+1} - 1)\text{ess sup}_{[0,T]}p$ the minimum of which occurs for $J = 0$. ■

5.2. Inferential sequences for the estimated intensities. We close this section with some properties of $\hat{p}_J$ and $\hat{p}^T_{J,\lambda}$.

Theorem 5.4. *If N satisfies Assumption B, then*

(i) $\hat{p}_J$ *is an asymptotically unbiased estimator for the intensity function* p.

For N under Assumption A:

(ii) $Var(\hat{p}_J) = \sum_{Ze(\ell i)_J} \sum_{Ze(\ell i)_J} (\int \int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)dudv)\psi_\eta\psi_\xi$
$+ \sum_{Ze(\ell i)_J} \sum_{Ze(\ell i)_J} (\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt)\psi_\eta\psi_\xi$.

If N is a NIC point process satisfying Assumption A, then

(iii) $Var(\hat{p}_J) = \sum_{Ze(\ell i)_J} \sum_{Ze(\ell i)_J} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt\right)\psi_\eta\psi_\xi$, *and*

(iv) $\widehat{Var}(\hat{p}_J) = \sum_{Ze(\ell i)_J} \sum_{Ze(\ell i)_J} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)dN(t)\right)\psi_\eta\psi_\xi$ *is an unbiased estimator for* $Var(\hat{p}_J)$.

Proof

(i) $\lim_{J\to\infty} E(\hat{p}_J) = \lim_{J\to\infty} E(\sum_{\eta,j\leq J} \hat{\beta}_\eta\psi_\eta) =$
$= \lim_{J\to\infty} \sum_{\eta,j\leq J} E(\hat{\beta}_\eta\psi_\eta) = \lim_{J\to\infty} \sum_{\eta,j\leq J} \beta_\eta\psi_\eta = \sum_\eta \beta_\eta\psi_\eta = p$, in $L^2[0,T]$.

(ii), (iii) and (iv): it is sufficient to repeat the argument of the proof of Theorem 4.4, observing that $\eta, \xi \in Ze(\ell i)_J$. ■

Theorem 5.5. (Inferential sequence for p_J.) *Let $\eta = (\eta_1, ..., \eta_{2^n}) \in (Ze(\ell i)_J)^{2^n}$ and let N be a NIC point process under Assumption A. Let, for all $n \geq 1$,*

$$V_n(\hat{p}_J) = \sum_{\eta \in (Ze(\ell i)_J)^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} p dt\right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}$$

and

$$\hat{V}_n(\hat{p}_J) = \sum_{\eta \in (Ze(\ell i)_J)^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} dN \right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}.$$

Then, $V_n(\hat{p}_J)$ and $\hat{V}_n(\hat{p}_J)$ are, respectively, sequences of variances and estimators such that

(i) $E(\hat{p}_J) = p_J$, $V_1(\hat{p}_J) = \mathrm{Var}(\hat{p}_J)$

(ii) $V_{n+1}(\hat{p}_J) = \mathrm{Var}(\hat{V}_n(\hat{p}_J))$

(iii) $\hat{V}_n(\hat{p}_J)$ *is an unbiased estimator for* $V_n(\hat{p}_J)$.

(iv) $\hat{V}_n(\hat{p}_J)$ *is non-negative.*

Briefly, the sequence $(\hat{p}_J, V_n(\hat{p}_J), \hat{V}_n(\hat{p}_J))$ constitutes an inferential sequence of stochastic processes for $p_J = \sum_{\eta \in Ze(\ell i)_J} \beta_\eta \psi_\eta$, the approximation of the intensity by wavelets up to the J-th scale.

Proof

(i) Immediate; Theorem 5.4 *(iii)* and *(iv)*.

(ii) Just replace $Ze(\ell i)$ by $Ze(\ell i)_J$ in the proof of Theorem 4.5.

(iii) Same argument of the proof of Theorem 4.5.

(iv) Just replace $Ze(\ell i)$ by $Ze(\ell i)_J$ in the proof of Theorem 4.5. ■

The preceding Theorems 5.4 and 5.5 give us all the relevant information about $\hat{p}_J$, that is, its asymptotical unbiasedness, its variance function in the general case and the inferential sequence for p_J in case of NIC point processes. Now, we turn our attention to the threshold case. Observe that preposition 5.3 deals with the case of an arbitrary threshold function. Due to this, only upper bounds for the variance function were obtained. Assuming the hard threshold, we can derive the equalities in proposition 5.4 and obtain the important result for NIC point processes given by Theorem 5.6, i.e., an inferential sequence for $p^T_{J,\lambda}$.

Proposition 5.3. *If N satisfies Assumption A, then, for all threshold functions, we have*

(i) $\|E\hat{p}_J - E\hat{p}^T_{J,\lambda}\|^2 \to 0$, *when* $\lambda \to 0$, *for p and $\psi_{(0,0)}$ satisfying the hypothesis of Theorem 5.2 or 5.3.*

(ii) $Var(\hat{p}^T_{J,\lambda}) \leq \sum_{\xi,\eta \in Ze(\ell i)_J} \left(\int\int_C |\psi_\eta(u)\psi_\xi(v)| E(dN(u)dN(v)) \right. |\psi_\eta\psi_\xi|$
$+ \sum_{\xi,\eta \in Ze(\ell i)_J} \left(\int_0^T |\psi_\eta(t)\psi_\xi(t)| p(t)dt \right) |\psi_\eta\psi_\xi| + \sum_{\xi,\eta \in Ze(\ell i)_J} \|\psi_\eta p\|_1 \|\psi_\xi p\|_1 |\psi_\eta\psi_\xi|.$

If N is also a NIC point process, we have:

(iii) $Var(\hat{p}^T_{J,\lambda}) \leq \sum_{\eta,\xi \in Ze(\ell i)_J} \left(\int_0^T |\psi_\eta(t)\psi_\xi(t)| p(t)dt + 2\|\psi_\eta p\|_1 \|\psi_\xi p\|_1 \right) |\psi_\eta\psi_\xi|.$

(iv) $\widehat{MVar}(\hat{p}^T_{J,\lambda}) =$

$$\sum_{\eta,\xi \in Ze(\ell i)_J} \left(2\int_0^T |\psi_\eta(t)| dN(t) \int_0^T |\psi_\xi(t)| dN(t) - \int_0^T |\psi_\eta(t)\psi_\xi(t)| dN(t) \right) |\psi_\eta\psi_\xi|$$

is an unbiased estimator for the right hand side of the above inequality.

Observe that (iii) and (iv) give bounds that are independent of λ.

Proof (i) Due to the argument used in the proofs of Theorems 5.2 and 5.3, we can write

$$\|E(\hat{p}_J) - E(\hat{p}^T_{J,\lambda})\|^2 =$$

$$= \| \sum_{\eta \in Ze(\ell i)_J} E(\hat{\beta}_\eta - E(T(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta))\psi_\eta\|^2 \leq \lambda^2(k_1 + 2^{J+1} - 1)\text{ess sup}_{[0,T]}p.$$

Therefore, the left hand side of the inequality tends to zero as $\lambda \to 0$.

(ii) Write $\hat{T}_\eta$ to represent $T(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})$ and $\hat{x}_\eta = \hat{T}_\eta\hat{\beta}_\eta$. Then

$$\mathrm{Var}(\hat{p}^T_{J,\lambda}) = E(\hat{p}^T_{J,\lambda} - E(\hat{p}^T_{J,\lambda}))^2 = E(\sum_{Ze(\ell i)_J} \hat{x}_\eta\psi_\eta - \sum_{Ze(\ell i)_J} E(\hat{x}_\eta)\psi_\eta)^2 =$$

$$= E(\sum_{\eta,\xi \in Ze(\ell i)_J} (\hat{x}_\eta - E(\hat{x}_\eta))(\hat{x}_\xi - E(\hat{x}_\xi))\psi_\eta\psi_\xi) = \sum_{\eta,\xi \in Ze(\ell i)_J} \mathrm{Cov}(\hat{x}_\eta, \hat{x}_\xi)\psi_\eta\psi_\xi.$$

Therefore,

$$\mathrm{Var}(\hat{p}^T_{J,\lambda}) \leq \sum_{\eta,\xi \in Ze(\ell i)_J} |\mathrm{Cov}(\hat{x}_\eta, \hat{x}_\xi)||\psi_\eta\psi_\xi|.$$

Now,

$$|\mathrm{Cov}(\hat{x}_\eta, \hat{x}_\xi)| \leq |E(\hat{T}_\eta\hat{T}_\xi\hat{\beta}_\eta\hat{\beta}_\xi)| + |E(\hat{T}_\eta\hat{\beta}_\eta)||E(\hat{T}_\xi\hat{\beta}_\xi)|$$

$$\leq E|\hat{T}_\eta\hat{T}_\xi\hat{\beta}_\eta\hat{\beta}_\xi| + E|\hat{T}_\eta\hat{\beta}_\eta|E|\hat{T}_\xi\hat{\beta}_\xi|$$

$$\leq E|\hat{\beta}_\eta\hat{\beta}_\xi| + E|\hat{\beta}_\eta|E|\hat{\beta}_\xi|,$$

since $|T_\eta| \leq 1$ and $|T_\xi| \leq 1$. From $\hat{\beta}_\eta = \int \psi_\eta dN(t)$ we have $|\hat{\beta}_\eta| \leq \int |\psi_\eta| dN(t)$ and, consequently, $E|\hat{\beta}_\eta| \leq \int |\psi|_\eta p dt = \|\psi_\eta p\|_1$. Analogously, $E|\hat{\beta}_\xi| \leq \|\psi_\xi p\|_1$.

From

$$\hat{\beta}_\eta\hat{\beta}_\xi = \int\int \psi_\eta(u)\psi_\xi(v)dN(u)dN(v)$$

we have

$$E|\hat{\beta}_\eta\hat{\beta}_\xi| \leq \int\int_C |\psi_\eta(u)\psi_\xi(v)|E(\mathrm{d}N(u)\mathrm{d}N(v)) + \int\int_{D_1} |\psi_\eta(u)\psi_\xi(v)|E(\mathrm{d}N(u)\mathrm{d}N(v)),$$

$D_1 = [0,T]^2 \cap D$.

Now, since N is under Assumption A,

$$\int\int_{D_1} |\psi_\eta(u)\psi_\xi(v)|E(\mathrm{d}N(u)\mathrm{d}N(v)) = \int_{D_1} |\psi_\eta\psi_\xi| dE(N \times N)|_D = \int_0^T |\psi_\eta(u)\psi_\xi(u)|p(u)du.$$

Therefore

$$|\mathrm{Cov}(\hat{x}_\eta, \hat{x}_\xi)| \leq \int\int_C |\psi_\eta(u)\psi_\xi(v)|E(\mathrm{d}N(u)\mathrm{d}N(v)) +$$

$$+\int_0^T |\psi_\eta(u)\psi_\xi(u)|p(u)du + \|\psi_\eta p\|_1\|\psi_\xi p\|_1$$

and

$$\mathrm{Var}(\hat{p}_{J,\lambda}) \leq \sum_{\eta,\xi\in Ze(\ell i)_J}\left(\int\int_C |\psi_\eta(u)\psi_\xi(v)|E(\mathrm{d}N(u)\mathrm{d}N(v))\right)|\psi_\eta\psi_\xi|$$

$$+\sum_{\eta,\xi\in Ze(\ell i)_J}\left(\int_0^T |\psi_\eta(u)\psi_\xi(u)|p(u)du + \|\psi_\eta p\|_1\|\psi_\xi p\|_1\right)|\psi_\eta\psi_\xi|.$$

(iii) For a NIC point process N we have $\frac{\mathrm{d}E(N\times N)}{\mathrm{d}\ell_2} = p_2(u,v) = p(u)p(v),\ u\neq v$. Let $p^*(u,u) = p^2(u)$ and $p^*(u,v) = p_2(u,v),\ u\neq v$. Then

$$\int\int_C |\psi_\eta(u)\psi_\xi(v)|p_2(u,v)dudv = \int_0^T\int_0^T |\psi_\eta(u)\psi_\xi(v)|p^*(u,v)dudv =$$

$$= \int_0^T |\psi_\eta(u)|p(u)du\int_0^T |\psi_\xi(v)|p(v)dv = \|\psi_\eta p\|_1\|\psi_\xi p\|_1.$$

The conclusion follows by direct substitution into expression (ii).

(iv) Since $\|\psi_\eta p\|_1 = \int_0^T |\psi_\eta|pdt = E\int_0^T |\psi_\eta|dN(t)$, we can write

$$\|\psi_\eta p\|_1\|\psi_\xi p\|_1 =$$

$$E\left(\int_0^T |\psi_\eta|dN(t)\int_0^T |\psi_\xi|dN(t)\right) - \mathrm{Cov}\left(\int_0^T |\psi_\eta|dN(t), \int_0^T |\psi_\xi|dN(t)\right).$$

By Proposition 3.3 for NIC point processes, the covariance in the right hand side of the above equation is equal to $\int_0^T |\psi_\eta||\psi_\xi|pdt = E\int_0^T |\psi_\eta\psi_\xi|dN(t)$, hence

$$\|\psi_\eta p\|_1\|\psi_\xi p\|_1 = E\left(\int_0^T |\psi_\eta|dN(t)\int_0^T |\psi_\xi|dN(t)\right) - E\int_0^T |\psi_\eta\psi_\xi|dN(t),$$

and

$$\int_0^T |\psi_\eta(t)\psi_\xi(t)|p(t)dt + 2\|\psi_\eta p\|_1\|\psi_\xi p\|_1 =$$

$$= E\left(2\int_0^T |\psi_\eta|dN(t)\int_0^T |\psi_\xi|dN(t) - \int_0^T |\psi_\eta\psi_\xi|dN(t)\right).$$

Summing for all η e ξ in $Ze(\ell i)_J$, the proposition is established.

■

Proposition 5.4. *Let N satisfying Assumption A. If we choose $T(x,y) = 1$ when $|x| \geq y$ and $T(x,y) = 0$ otherwise, and if we let*

$$TZe(\ell i)_J = \{\eta \in Ze(\ell i)_J | T(\hat{\beta}_\eta, \lambda\sqrt{Var(\hat{\beta}_\eta)}) = 1\},$$

then

(i) $Var(\hat{p}^T_{J,\lambda}) = \sum_{\eta,\xi \in TZe(\ell i)_J} \left(\int\int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)dudv + \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt\right)\psi_\eta\psi_\xi.$

If N is also a NIC point process, we have (ii) and (iii) below:

(ii) $Var(\hat{p}^T_{J,\lambda}) = \sum_{\eta,\xi \in TZe(\ell i)_J}(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt)\psi_\eta\psi_\xi.$

(iii) $\widehat{Var}(\hat{p}^T_{J,\lambda}) = \sum_{\eta,\xi \in TZe(\ell i)_J}(\int_0^T \psi_\eta(t)\psi_\xi(t)dN(t))\psi_\eta\psi_\xi$ *is an unbiased estimator for* $Var(\hat{p}^T_{J,\lambda})$.

Proof Since

$$\hat{p}^T_{J,\lambda} = \sum_{\eta \in Ze(\ell i)_J} T(\hat{\beta}_\eta, \lambda\sqrt{\mathrm{Var}(\hat{\beta}_\eta)})\hat{\beta}_\eta\psi_\eta = \sum_{\eta \in TZe(\ell i)_J} \hat{\beta}_\eta\psi_\eta,$$

it is sufficient to repeat the argument used in the proof of Theorem 4.4. ■

Theorem 5.6. (Inferential sequence for $p^T_{J,\lambda}$.) *Let N be a NIC point process under Assumption A and let $T(x,y) = 1$ when $|x| \geq y$ and $T(x,y) = 0$ otherwise. Let $TZe(\ell i)_J = \{\xi \in Ze(\ell i)_J | T(\hat{\beta}_\xi, \lambda\sqrt{Var(\hat{\beta}_\xi)}) = 1\}$ and $\eta = (\eta_1, \ldots, \eta_{2^n}) \in (TZe(\ell i)_J)^{2^n}$. Then,*

$$V_n(\hat{p}^T_{J,\lambda}) = \sum_{\eta \in (TZe(\ell i)_J)^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} p dt\right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}$$

and

$$\hat{V}_n(\hat{p}^T_{J,\lambda}) = \sum_{\eta \in (TZe(\ell i)_J)^{2^n}} \left(\int_0^T \prod_{\ell=1}^{2^n} \psi_{\eta_\ell} E dN\right) \prod_{\ell=1}^{2^n} \psi_{\eta_\ell}, \text{ for all } n \geq 1,$$

are, respectively, sequences of variances and estimators such that

(i) $E(\hat{p}^T_{J,\lambda}) = p^T_{J,\lambda}$, $V_1(\hat{p}^T_{J,\lambda}) = \mathrm{Var}(\hat{p}^T_{J,\lambda})$,

(ii) $V_{n+1}(\hat{p}^T_{J,\lambda}) = Var(\hat{V}_n(\hat{p}^T_{J,\lambda}))$,

(iii) $\hat{V}_n(\hat{p}^T_{J,\lambda})$ *is an unbiased estimator for* $V_n(\hat{p}^T_{J,\lambda})$.

(iv) $\hat{V}_n(\hat{p}^T_{J,\lambda})$ *is non-negative.*

That is, the sequence $(\hat{p}^T_{J,\lambda}, V_n(\hat{p}^T_{J,\lambda}), \hat{V}_n(\hat{p}^T_{J,\lambda}))$, $n \geq 1$, constitutes an inferential sequence of stochastic processes for the function $p^T_{J,\lambda}$, wavelet threshold approximation of the intensity till the J-th order scale, with $L : \mathbb{R}_+ \times Ze(\ell i)_J \to \mathbb{R}_+$ defined by $L(\lambda, \eta) = \lambda\sqrt{Var(\hat{\beta}_\eta)}$, $T(x,y) = 1$ when $|x| \geq y$ and $T(x,y) = 0$ otherwise.

Proof

(i) Immediate; Proposition 5.4 *(ii)* and *(iii)*.

(ii) , (iii) and (iv) Immediate. ■

6. AN APPLICATION

We will present here an application of the results obtained in the former sections. The intensity of a point process derived from the daily log-returns of the Dow-Jones Industrial Average will be estimated. To form a point process from these returns, we will agree that an event has occurred if and only if the absolute value of the log-return is greater than a given threshold level. A set of $T = 4225$ returns will be used, corresponding to the period of time from January 2nd 1986 to September 26, 2002, and the threshold level will be 0.01452 which corresponds to 1.28 times the standard deviation of these returns.

This procedure generates 558 events, which we assume to be a realization of a NIC point process with intensity $p_N(t)$. Since it is necessary to limit the number of wavelet coefficients that will be estimated and used for the synthesis of $\hat{p}$, our choice is made of a set of coefficients that encompass exactly all coefficients of all scales of order less than or equal to a positive number J. If the intensity were constant we would expect $(558/4225)c$ events within an interval of length c. Under this assumption one will expect to have $558/2^6 \approx 8$ events laying inside the support of each wavelet of the sixth scale and if the intensity at some time interval is half of the average intensity this number may drop to 4. Information based on a wavelet with few points laying within its support may be misleading. This heuristic argument led us to choose all wavelets until the fifth order for our synthesis procedure.

An important advantage of our estimation method is that we have direct access to the variance of $\hat{\beta}_\eta$, through $\widehat{\mathrm{Var}}(\hat{\beta}_\eta)$, for each η individually, and not by an estimation that depends on the whole set of wavelet coefficients of a given scale or any subset of the set of all wavelet coefficients. We observe that when one uses an estimator of $\mathrm{Var}(\hat{\beta}_\eta)$, for a given η, based on the variance of the values of all $\hat{\beta}_\xi$, that may belong to the same scale of β_η or to a bigger set of coefficients, what really is being done is to calculate an estimator of the variance of the coefficients within this set and most of this variance, probably, is due to the diversity of the indexes ξ's, that is, of all distinct β_ξ's in this set, and this variance may not have any or little relation with the variance of $\hat{\beta}_\eta$ for that particular η of interest.

It is worth noting that when the process is under the presence of noise it may happen that the whole set of coefficients is affected and the variance of the coefficients of higher-order scales is a measure of the intensity of the noise point process. In fact if the noise point process that is added to N is a homogeneous NIC point process with intensity λ, then the variance of the coefficients that belong to the J-th order scale is an asymptotically unbiased estimator of λ, that is, $E\{\mathrm{Var}\,(\hat{\beta}_{(0,J)}, \ldots, \hat{\beta}_{(2^J-1,J)})\} \to \lambda$, as $J \to \infty$. In this case we can still obtain the estimated intensity of the process N by synthesis based on the measured process and then subtracting from this estimated intensity the estimated intensity of the noise. See de Miranda (2003a).

We have used in this application the Haar wavelet system. Let I_A be the indicator function of a set A. Thus

$$\psi_{(0,0)} = T^{-1/2}(I_{[0,T/2)} - I_{[T/2,T)}) \,,\ \phi_{(0,0)} = T^{-1/2} I_{[0,T)} \,,\ T = 4225$$

and

$$\psi_{(i,j)} = \frac{2^{j/2}}{T^{1/2}} \left(I_{[iT/2^j,(2i+1)T/2^{j+1})} - I_{[(2i+1)T/2^{j+1},(i+1)T/2^j)} \right),$$

$$\psi^2_{(i,j)} = \frac{2^j}{T} I_{[iT/2^j,(i+1)T/2^j)}.$$

The estimators $\hat{\beta}_\eta$ and $\widehat{\mathrm{Var}}(\hat{\beta}_\eta)$ were obtained through the formulas

$$\hat{\beta}_{(i,j)} = \int \psi_{(i,j)} dN(t) = \sum_{\tau_k \in \mathrm{supp}\,\psi_{(i,j)}} \psi_{(i,j)}(\tau_k) =$$

$$= \frac{2^{j/2}}{T^{1/2}} \left(\#\{\tau_k \in [iT/2^j, (2i+1)T/2^{j+1})\} - \#\{\tau_k \in [(2i+1)T/2^{j+1}, (i+1)T/2^j)\} \right)$$

and

$$\widehat{\mathrm{Var}}(\hat{\beta}_{i,j}) = \int \psi^2_{i,j} dN(t) = \sum_{\tau_k \in \mathrm{supp}\,\psi_{i,j}} \psi^2_{i,j}(\tau_k) =$$

$$= \frac{2^j}{T} \left(\#\{\tau_k \in [iT/2^j, (i+1)T/2^j)\} \right).$$

Analogously, we have obtained $\hat{\beta}_0$ and $\mathrm{Var}\hat{\beta}_0$. Observe that

$$\hat{\beta}_0 \phi_{(0,0)} = \frac{1}{T^{1/2}} \frac{1}{T^{1/2}} \#\{\tau_k \in [0,T]\} = \frac{\#\{\tau_k \in [0,T]\}}{T}$$

is the mean intensity, 558/4225, that is, the mean value of p.

The threshold function chosen was $T(x,y) = 0$ for $|x| < y$ and $T(x,y) = 1$ for $|x| \geq y$. We recall that for $\lambda = 3$ we have (using Chebyshev's inequality) a "confidence level" of at least $1 - (1/3)^2 = 8/9$ or approximately 88, 8% whatever the distribution of $\hat{\beta}_\eta$ is.

In Figure 1 we show the number of counts and in Figure 2 the estimated intensity. We clearly see the non-stationary character of the process. In Figures 3 and 4 we have the estimated standard deviation and the respective threshold version, as given by Theorem 5.4 and Proposition 5.4. Figures 5 and 6 show the estimated intensity and thresholded estimated intensity, respectively, with their (non-negative) confidence bands. Again, these last figures confirm the non-homogeneity of the fitted NIC point process. In Figures 5 and 6 the bands are computed adding (and subtracting) μ times the standard deviation function to (respectively from) the intensity function, bounded inferiorly by zero. If we do not assume that N is a NIC point process, the estimated intensity and its threshold version are still the ones presented, but we cannot in this case compute the bands.

7. Further Comments

In this work we dealt with the problem of estimating the time-variable intensity of a nonhomogeneous point process on the real line, specializing for the case of a NIC point process. The generalization for point process on $\mathbb{R}^m$, using for example wavelets on $\mathbb{R}^m$ given by tensor products of wavelets on $\mathbb{R}$, can be directly done. A more general treatment is possible and this will be pursued elsewhere.

Another situation of interest might be that where a point process occurs under noisy conditions. We have a primary point process N that is the object of our study and to this it is summed another point process that will be called the noise process, R. The resulting point process M is the one effectively observed. We write $M = N + R$ and by this we mean that for all $A \subset \mathcal{B}_{\mathbb{R}}, M(A) = N(A) + R(A)$. It is also assumed that N and R are independent. The target is to estimate the intensity of N, which will depend on the estimate of the intensity of the noisy process. Similar results to those obtained here can be derived. See de Miranda (2003a).

References

Besbeas, P., De Feis, I. and Sapatinas, T. (2002). A comparative simulation study of wavelet shrinkage estimators for Poisson counts. *Rapporto Tecnico*, 240/02. Istituto per le Applicazioni del Calcolo "Mauro Picone"- Sezione di Napoli, CNR, Italy.

Brillinger, D.R. (1975). Statistical inference for stationary point processes. *Stochastic Processes and Related Topics*, **1**, 55–99.

Brillinger, D.R. (1978). Comparative aspects of the study of ordinary time series and point processes. *Developments in Statistics*, **1**, 33–133.

Brillinger, D.R. (1997). Some wavelet analyses of point processes data. The Third First Asilomar Conference on Signals, Systems and Computers. IEEE Computer Society, 1087–1091.

Cramér, H. and Leadbetter,M.R. (1967). *Stationary and Related Stochastic Processes.* New York: Wiley.

Daley, D.J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes.* New York: Springer.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM.

de Miranda, J.C.S. (2003a). On the estimation of the intensity of point processes via wavelets. PhD thesis. University of São Paulo. São Paulo, Brazil. (In Portuguese.)

de Miranda, J.C.S. (2003b). Some wide classes of point Processes - Interrelationships and three working properties. Preprint.

de Miranda, J.C.S. (2003c). Sure inference analysis. Preprint.

de Miranda, J.C.S. and Morettin, P.A. (2003a). Characterization of the intensity of point process. Preprint.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet shrinkage. *The Annals of Statistics*, **24**, 508–539.

Helmers, R. and Zitikis, R. (1999). On estimation of Poisson intensity functions. *Annals of the Institute of Statistical Mathematics*, **51**, 265–280.

Kolaczyk, E.D. (1999a). Bayesian multiscale models for Poisson processes. *Journal of the American Statistical Association*, **94**, 92–933.

Kolaczyk, E.D. (1999b). Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statistica Sinica*, **9**, 119–135.

Meyer, Y. (1992). *Ondelettes et Algorithmes Concurrents.* Hermann.

Morettin, P.A. (1999). *Waves and Wavelets.* São Paulo: University of São Paulo Press.

Priestley, M.B. (1996). Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis*, **17**, 85–103.

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.

Snyder, D.L. (1975). *Random Point Processes*. New York: Wiley.

Timmermann, K.E. and Nowak, R.D. (1997). Multiscale modelling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory*, **45**, 846–862.

José Carlos Simon de Miranda and Pedro A. Morettin
Department of Statistics
University of São Paulo
05311-970- São Paulo, SP
Brazil
e-mails: simon@ime.usp.br, pam@ime.usp.br

The first author thanks **Our Lord and Savior Jesus Christ.**

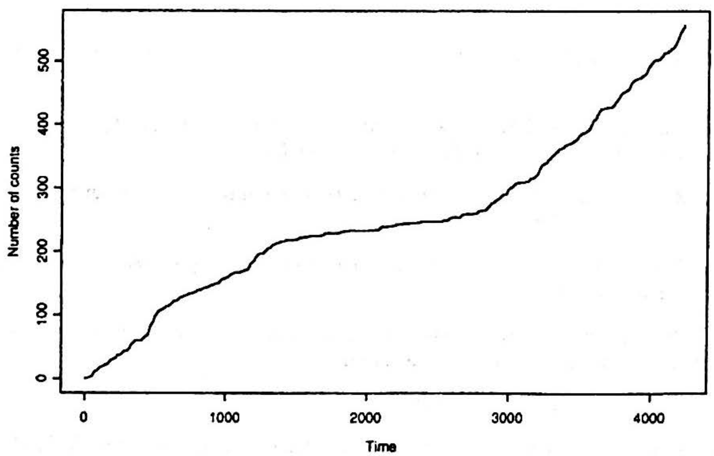

FIGURE 1. Accumulated number of events

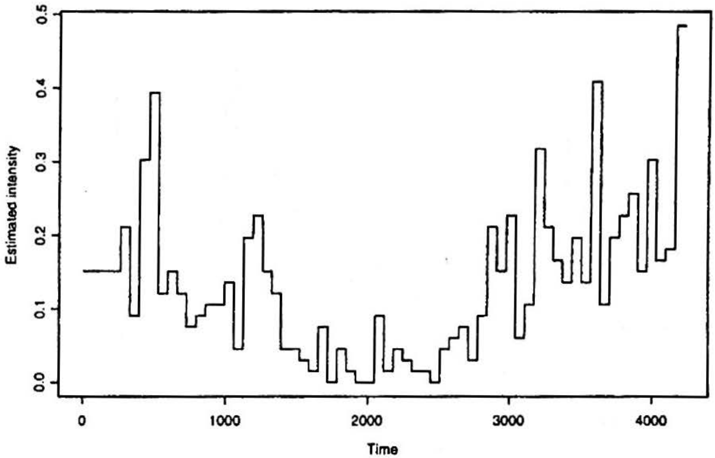

FIGURE 2. Estimated intensity

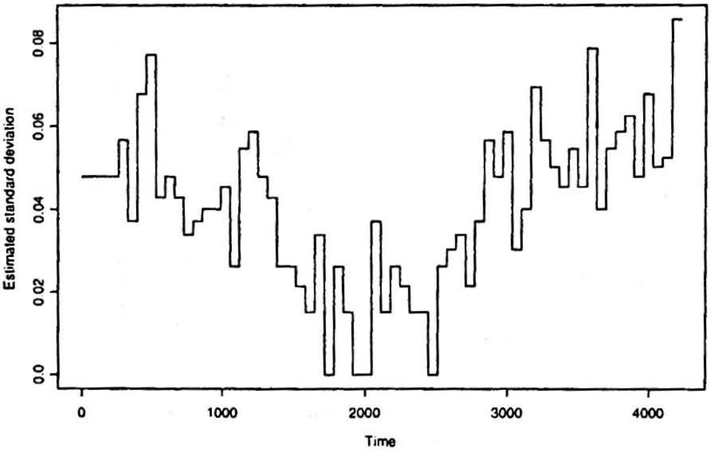

FIGURE 3. Estimated standard deviation

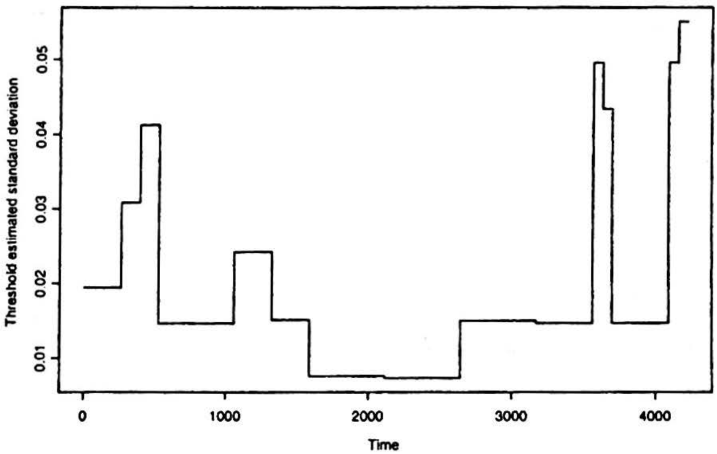

FIGURE 4. Estimated standard deviation under threshold ($\lambda = 3$)

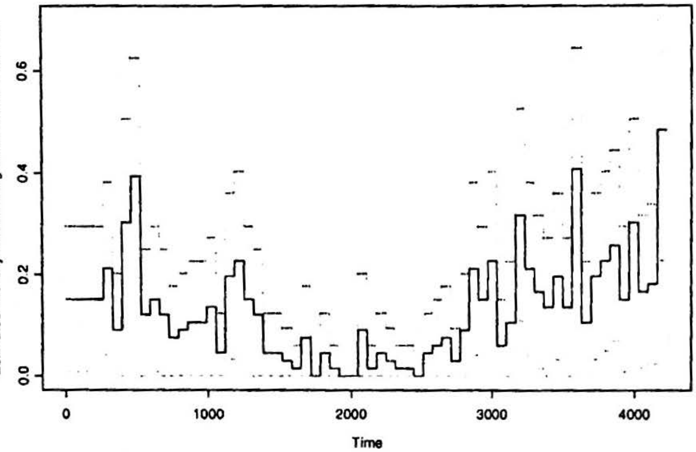

FIGURE 5. Estimated intensity with non-negative bands ($\mu = 3$)

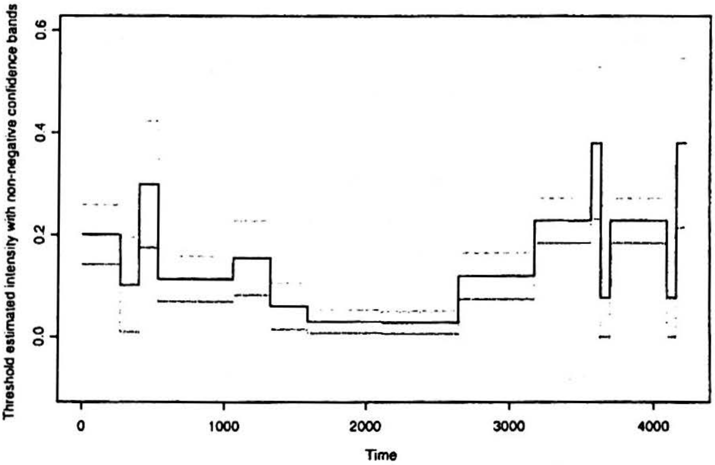

FIGURE 6. Estimated intensity under threshold with non-negative bands ($\lambda = 3, \mu = 3$)

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2006-01 - SILVA, M.F. DA, FERRARI, S.L.P., CRIBARI-NETO, F. Improved likelihood inference for the shape parameter in Weibull regression. 2006. 24p. (RT-MAE 2006-01)

2006-02 - LACHOS, V.H., BOLFARINE, H., ARELLANO-VALLE, R. B., MONTENEGRO, L. C. Inference in multivariate skew-normal regression models. 2006. 21p. (RT-MAE 2006-02)

2006-03 - BUENO, V. DA C. Redundancy allocation for a coherent system of dependent components. 2006. 16p. (RT-MAE 2006-03)

2006-04 - ESPINHEIRA, P.L., FERRARI, S. L. P., CRIBARI-NETO, F. On Beta regression residuals. 2006. 17p. (RT-MAE 2006-04)

2006-05 - BAZÁN, J. L., BOLFARINE, H., BRANCO, M. D. A generalized skew probit class link for binary regression. 2006. 22p. (RT-MAE 2006-05)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil