Estimation of the intensity of non-homogeneous point processes via wavelets
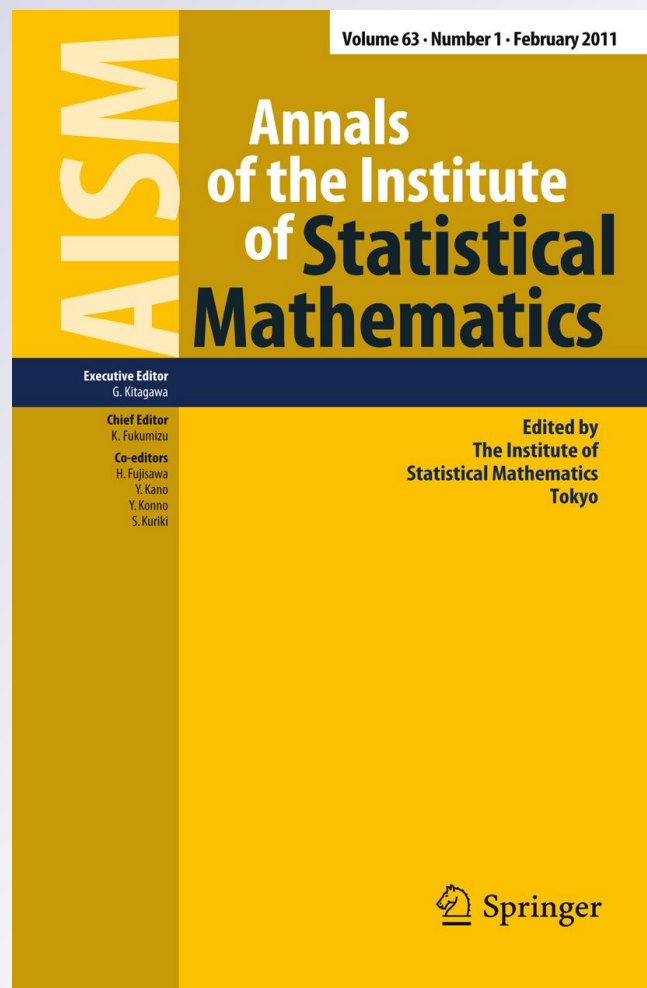
José Carlos Simon de Miranda & Pedro A. Morettin

Springer

Estimation of the intensity of non-homogeneous point processes via wavelets

José Carlos Simon de Miranda · Pedro A. Morettin

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 thresholded versions. Properties of the estimators for non-homogeneous Poisson processes 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 · 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 extensively discussed in several previous works, especially for Poisson point processes. Brillinger (1975) considers inference for general stationary point processes. Aalen (1978) introduces the so-called multiplicative intensity model, where it is assumed that the intensity of the point process can be written in the form $p_N(t) = \alpha(t)Y(t)$, where α is an unknown function and Y is a stochastic process which can be observed together with N. It is further assumed that α and the sample functions of Y be non-negative, left continuous

J. C. S. de Miranda · P. A. Morettin (✉)
Department of Statistics, Institute of Mathematics and Statistics,
C.P. 20570, São Paulo, SP 05311-970, Brazil
e-mail: pam@ime.usp.br

J. C. S. de Miranda
e-mail: simon@ime.usp.br

with right-hand limits. Non-parametric inference procedures are proposed to estimate functions like $\beta(t) = \int_0^t \alpha(s)\mathrm{d}s$ and related quantities.

Rathbun and Cressie (1994) consider inhomogeneous Poisson processes on a bounded Borel set $A \subset \mathbb{R}^d$, with intensity function $p_n(\mathbf{s}, \theta), \theta \in \Theta \subset \mathbb{R}^k$. Maximum likelihood and Bayes estimators of θ are proposed and shown to be asymptotically efficient and normally distributed, under several assumptions and special forms of processes. Further references are Kutoyants (1984) and Krickeberg (1982).

Helmers and Zitikis (1999) use a non-parametric approach for estimating the intensity of a Poisson process, while Helmers et al. (2005, 2007) use kernel-type estimators for the intensity function of cyclic and doubly periodic Poisson processes.

In our work we will use a non-parametric approach through wavelet expansions, as in Donoho et al. (1995, 1996) and Hall and Patil (1995, 1996). Wavelets provide a way of estimating intensities of 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.

Several works have dealt with point processes and wavelets. We mention Brillinger (1998), Timmermann and Nowak (1997), Kolaczyk (1999) and Besbeas et al. (2002). Patil and Wood (2004) consider a wavelet-based estimator of $\alpha(t)$ in Aalen's multiplicative model described above, under several assumptions on the process Y and the function α. Previously kernel estimators of α were proposed by Ramlau-Hansen (1983). These authors were mainly interested in developing mean integrated square error properties of the wavelet estimators.

General references on point processes are Snyder (1975), Daley and Vere-Jones (1988), Kingman (1993) and Kutoyants (1998).

In our work we expand in a wavelet series the restriction of the intensity of a point process to the interval where we know that the points of a trajectory of the underlying process lie. We propose unbiased estimators of the wavelet coefficients and derive their variances. Next, estimators of the intensity function are proposed and their properties analyzed. The probability density function of the empirical wavelet coefficients of a non-homogeneous Poisson process can also be derived.

The plan of the article is as follows. In Sect. 2 we provide some background on point processes and wavelets and set up the assumptions needed to establish the main results. The estimators of the intensity are introduced in Sect. 3 and an application is given in Sect. 4. The paper ends with some further comments in Sect. 5.

2 Background and assumptions

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 Point processes

We denote by $N(A)$ the number of events 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 $\{\ldots, \tau_{-2} \leq \tau_{-1} \leq \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots\}$ denote the times at which the events occur, $\tau_{-1} \leq 0 < \tau_0$. Then $N(t) = n$ if and only if $\tau_{n-1} \leq t < \tau_n$.

Provided that the set of probabilities of the form

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

for all $k \in \mathbb{N}^* = \{1, 2, \ldots\}$, and all $n_1, \ldots, n_k$ non-negative integers is consistent, 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 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 $(\alpha_i, \beta_i] \cap (\alpha_j, \beta_j] = \emptyset$, for all $i \neq j$,

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

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 $\Lambda(t) = \lambda t$, for a given positive constant λ.

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 cumulative intensity function $\Lambda(t)$.

Define $dN(t) = N(t + dt) - N(t)$. A basic assumption is that the measures $E\{dN(t_1) \cdots dN(t_k)\}$ exist and are boundedly finite, for all $t_1, \ldots, t_k$.

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,

$$E\left\{ \int \varphi_1(t_1) dN(t_1) \cdots \int \varphi_k(t_k) dN(t_k) \right\} = \int \varphi_1(t_1) \cdots \varphi_k(t_k) E\{dN(t_1) \cdots dN(t_k)\}.$$

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

$$E\left\{\int \varphi(t)\mathrm{d}N(t)\right\} = \int \varphi(t)E\mathrm{d}N(t).$$

We also observe that $E\mathrm{d}N(t) = E(N(t+\mathrm{d}t) - N(t)) = EN(t+\mathrm{d}t) - EN(t) = \mathrm{d}EN(t)$.

2.2 Intensity and product density

Suppose that there exist 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 $t \in \mathbb{R}$, we have the relation

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

and also the existence of the limit

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

uniformly in t. Inequality (1) implies that

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

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:

$$\forall \varepsilon > 0, \exists \delta > 0, \forall t \in \mathbb{R}, \forall \Delta \subset \mathbb{R}, t \in \Delta, (0 < |\Delta| < \delta)$$
$$\Rightarrow |o_{t,\Delta}(|\Delta|)| \leq \frac{\varepsilon}{2}|\Delta| \text{ and } o_{t,\Delta}(0) = 0,$$

that is,

$$\forall \varepsilon > 0, \exists \delta > 0, (0 < z < \delta) \Rightarrow \sup_{\substack{t \in \mathbb{R}, \Delta \subset \mathbb{R} \\ t \in \Delta, |\Delta| = z}} |o_{t,\Delta}(z)| \leq \frac{\varepsilon}{2}z < \varepsilon z \text{ and } o_{t,\Delta}(0) = 0.$$

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 $|o_{t,\Delta}(|\Delta|)| \leq o(|\Delta|)$. To simplify the 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 for all vectors $(t_1, \ldots, t_n) \in \mathbb{R}^m$, with $t_i \neq t_j$ for $i \neq j$, $1 \leq i, j \leq m$, there exists a positive real number δ and a constant $k_{\delta,m}$ such that for all intervals $\Delta_1, \ldots, \Delta_m$, $t_i \in \Delta_i$, of the real line with lengths $0 < |\Delta_i| < \delta$, $1 \leq i \leq m$, and all integers $n_i \geq 1$ both properties below are valid:

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

and, denoting $\Delta = (\Delta_1, \ldots, \Delta_m)$ the m-tuple of intervals and $|\Delta| = (|\Delta_1|, \ldots, |\Delta_m|) \in (\mathbb{R}_+^*)^m$ the vector of their lengths, there exists the limit

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

Here $\mathbb{R}_+^*$ denotes the set of strictly positive real numbers.

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

$$\lim_{|\Delta| \to 0} \frac{1}{\prod_{i=1}^m |\Delta_i|} E\left\{\prod_{i=1}^m N(\Delta_i)\right\} = p_m(t_1, \ldots, t_m), \tag{5}$$

p_m is called *product density of order m*. Relation (4) implies that

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

for $o_{t,\prod_{i=1}^m \Delta_i}(\Delta)$ an infinitesimal such that if we denote by $\mathcal{E}^m$ the set $\{(t_1, \ldots, t_m) \in \mathbb{R}^m | t_i = t_j \text{ for some } i \neq j\}$, and, for $t \in \mathbb{R}^m \backslash \mathcal{E}^m$, $\mathcal{A}_{t,z} = \{\prod_{i=1}^m \Delta_i \subset \mathbb{R}^m : t \in \prod_{i=1}^m \Delta_i, |\Delta_i| = z_i, 1 \leq i \leq m\}$, where $z = (z_1, \ldots, z_m) \in (\mathbb{R}_+^*)^m$, then

$$\sup_{\mathcal{A}_{t,z}} \left|o_{t,\prod_{i=1}^m \Delta_i}(z)\right| = o_t(z),$$

is another infinitesimal which is independent of $\prod_{i=1}^m \Delta_i \subset \mathbb{R}^m$, and satisfies $(o_t(\Delta))/(\prod_{i=1}^m |\Delta_i|) \to 0$ when $|\Delta| \to 0$.

Again, to simplify the 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 $t_1 \neq t_2$, where $t_1 \in \Delta_1$, $t_2 \in \Delta_2$, by

$$q_2(t_1, t_2) = \lim_{|\Delta| \to 0} \frac{\text{Cov}\,(N, N)(\Delta_1 \times \Delta_2)}{|\Delta_1||\Delta_2|}.$$

Whenever $p_2(t_1, t_2)$, $p_1(t_1)$ and $p_2(t_2)$ exist, we write

$$
\begin{aligned}
q_2(t_1, t_2) &= \lim_{|\Delta| \to 0} \frac{\text{Cov}\,(N, N)(\Delta_1 \times \Delta_2)}{|\Delta_1||\Delta_2|} \\
&= \lim_{|\Delta| \to 0} \frac{E(N(\Delta_1)N(\Delta_2))}{|\Delta_1||\Delta_2|} - \lim_{|\Delta| \to 0} \frac{E(N(\Delta_1))}{|\Delta_1|} \frac{E(N(\Delta_2))}{|\Delta_2|} \\
&= p_2(t_1, t_2) - p_1(t_1)p_2(t_2).
\end{aligned}
$$

2.3 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

$$\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0, \tag{6}$$

$$\int_{-\infty}^{\infty} |\,\psi(t)\,|\,\mathrm{d}t < \infty, \tag{7}$$

and may also satisfy

$$\int_{-\infty}^{\infty} \frac{|\,\hat{\psi}(\omega)\,|^2}{|\,\omega\,|}\mathrm{d}\omega < \infty, \tag{8}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$, that is, $\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \psi(t)\mathrm{e}^{-i\omega t}\,\mathrm{d}t$.

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 $a > 0$ 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

$$\psi_{j,k}(t) = \psi^{(a,b)}(t) = 2^{j/2}\psi(2^j t - k), \tag{9}$$

constitute an orthonormal basis for $L^2(\mathbb{R})$. See Daubechies (1992) and Meyer (1992).

There are many different forms of $\psi(t)$ all of which satisfy the conditions (6)–(8). The oldest and simplest example of a function ψ for which the $\psi_{j,k}$ defined by (9) constitute an orthonormal basis for $L^2(\mathbb{R})$ is the Haar function,

$$\psi^{(H)}(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

From (10), we have:

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

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 $\int_{-\infty}^{\infty} \phi(t)\mathrm{d}t = 1$. 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

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

and

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

respectively.

For the Haar wavelet,

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

hence,

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

and

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

Consequently,

$$\psi(t) = \sqrt{2}\left(\left(1/\sqrt{2}\right)\phi(2t) - \left(1/\sqrt{2}\right)\phi(2t - 1)\right).$$

Similarly to the mother wavelet case, dilated and translated versions of the father wavelet are written as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k). \tag{11}$$

Except for some special cases, there are no analytic formulas for computing wavelet functions. An important result guarantees, for all r, the existence of orthonormal bases for $L^2(\mathbb{R})$ of the form $2^{j/2}\psi_{(r)}(2^j x - k)$, $j, k \in \mathbb{Z}$, 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.

2.4 Assumptions

We make now 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). Denote by $\overline{\mathcal{L}}^m$ the class of Lebesgue integrable functions over bounded intervals of $\mathbb{R}^m$.

Assumption 1 We assume that the point process N is such that its expectation measure is absolutely continuous with respect to Lebesgue measure, $EN \ll \ell$, that is, there exists $\mathrm{d}EN/\mathrm{d}\ell \in \overline{\mathcal{L}}^1$ and also the following relation holds: for all $t \in \mathbb{R}$ and for all interval $\Delta \subset \mathbb{R}$, $t \in \Delta$, $EN(\Delta) = P\{N(\Delta) = 1\} + o_{t,\Delta}(|\Delta|)$.

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 1 *Let N be a point process that satisfies Assumption 1. 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 $f = \dfrac{dEN}{d\ell}$, $\varphi(x) = \int_c^x f(y)dy$, $\Delta = |a, b|$, $a < b$, $h_1 = b - t$ and $h_2 = t - a$. Here $|a, b|$ denotes any one of the intervals (a, b), $(a, b]$, $[a, b)$ or $[a, b]$. 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,

$$\frac{\varphi(t + h_1) - \varphi(t - h_2)}{h_1 + h_2} = \frac{\varphi(t + h_1) - \varphi(t)}{h_1} \frac{h_1}{h_1 + h_2} + \frac{\varphi(t - h_2) - \varphi(t)}{-h_2} \frac{h_2}{h_1 + h_2}$$

$$= (f(t) + o_t(h_1)) \frac{h_1}{h_1 + h_2} + (f(t) + o_t(-h_2)) \frac{h_2}{h_1 + h_2}$$

$$= f(t) + \left(o_t(h_1) \frac{h_1}{h_1 + h_2} + o_t(-h_2) \frac{h_2}{h_1 + h_2} \right),$$

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

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

Thus, $p_N(t) = \dfrac{dEN}{d\ell}$ a.e.$[\ell]$. $\qquad \square$

Remark Notice that if N satisfies Assumption 1, then

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

holds for all $A \in \Lambda_{\mathbb{R}^2}$, where $\Lambda_{\mathbb{R}^2}$ is the σ-algebra of Lebesgue measurable sets in $\mathbb{R}^2$, 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 the diagonal, $E(N \times N)|_D : \Lambda_D \to \mathbb{R}$, is the induced measure over the diagonal by the measure EN over the real line through π_1, that is, $E(N \times N)|_D = EN\pi_1$.

Definition 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 $\text{Cov}(N(A), N(B)) = 0$.

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

An example showing that the class of NIC point processes is not the Poisson class is the following. Let N be a Poisson point process on $[0, 1]$. Let u_1, u_2 be two distinct real numbers in $[0, 1]$. Denote δ_x the unit point mass measure at $x \in [0, 1]$.

Let X and Y be two dependent random variables with values on the non-negative integers such that $\mathrm{Cov}(X, Y) = 0$. Assume that N and (X, Y) are independent. Then $M = N + X\delta_{u_1} + Y\delta_{u_2}$ is a NIC point process which is not Poisson. For more information on these point processes, see de Miranda (2009).

For point processes satisfying Assumption 1, we have the following proposition.

Proposition 1 *If N satisfies Assumption 1 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.$[\ell]$.   □

Using the remark above we have the following proposition, whose proof will be omitted.

Proposition 2 *If N satisfies Assumption 1 then, for all functions φ_1 integrable with respect to the covariance measure $\mathrm{Cov}\,(N, N)$, we have:*

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

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 *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 1, and assuming that $E(N \times N)$ is boundedly finite, we have*

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

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

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

If N is NIC then

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

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

We will need in some instance the following condition to be satisfied.

Assumption 2 The restriction of the intensity function p_N to $[0, T]$ is essentially α-Hölder in $A \subset \mathbb{R}$, that is, there exist two constants, K and α, and a set $D \subset \mathbb{R}$, $\ell(D) = 0$, such that for all x and y in $A \backslash D$ we have $|f(x) - f(y)| \le K|x - y|^\alpha$, $\alpha > 0$.

3 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 $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$ be an orthonormal wavelet basis of the form $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ or $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - kT)$ 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 and let $\{\phi_{l,k}, \psi_{j,k} : j, k \in \mathbb{Z}, j \ge l, j, l \in \mathbb{Z}\}$ be an orthonormal wavelet basis that contains all the scales beyond some fixed integer l.

We adopt the following notation. Let $_d\mathbb{Z} = \{z \in \mathbb{Z} : z \ge d\}$, $d \in \mathbb{Z}$ and $Ze(l) = \mathbb{Z} \cup (_l\mathbb{Z} \times \mathbb{Z})$.

Let us use Greek letters for indexes in $Ze(l)$ and we shall write $\psi_\eta = \phi_{l,\eta}$ if and only if $\eta \in \mathbb{Z}$ and $\psi_\eta = \psi_{j,k}$ if and only if $\eta = (j, k) \in {_l\mathbb{Z}} \times \mathbb{Z}$.

Thus, the wavelet expansion

$$f(t) = \sum_{k\in\mathbb{Z}} \delta_k \phi_{l,k}(t) + \sum_{j\in {_l\mathbb{Z}}} \sum_{k\in\mathbb{Z}} \beta_{j,k} \psi_{j,k}(t)$$

will be simply written

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

with the coefficients α_η given by

$$\int_{-\infty}^{\infty} f\psi_\eta \mathrm{d}t = \int_\mathbb{R} \left(\sum_\xi \alpha_\xi \psi_\xi \right) \psi_\eta \mathrm{d}t = \sum_\xi \int_\mathbb{R} \alpha_\xi \psi_\xi \psi_\eta \mathrm{d}t$$

$$= \sum_\xi \alpha_\xi \langle \psi_\xi, \psi_\eta \rangle = \alpha_\eta.$$

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

$$p = \sum_{\eta} \beta_{\eta} \psi_{\eta}, \tag{12}$$

with

$$\beta_{\eta} = \int_{\mathbb{R}} p\psi_{\eta} \mathrm{d}t = \int_0^T p\psi_{\eta} \mathrm{d}t. \tag{13}$$

The main purpose is to estimate p through the expansion (12) and for this we need to estimate the wavelet coefficients β_{η} given by (13). From now on assume $l = 0$.

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 $q_2(u, v)\mathrm{d}u\mathrm{d}v$ by $\mathrm{dCov}\,(N, N)$ in the statements of the theorems and propositions that follow.

3.1 Estimation of the wavelet coefficients

We propose the following estimator of β_{η}:

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

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

Theorem 2 *If N satisfies Assumption* 1, *then*

(i)   *the estimator $\hat{\beta}_{\eta}$ is unbiased.*
(ii) *for all η and ξ, assuming the existence and local integrability of q_2, we have*

$$\mathrm{Cov}(\hat{\beta}_{\eta}, \hat{\beta}_{\xi}) = \int \int_C \psi_{\eta}(u)\psi_{\xi}(v)q_2(u, v)\mathrm{d}u\mathrm{d}v$$
$$+ \int_0^T \psi_{\eta}(u)\psi_{\xi}(u)p(u)\mathrm{d}u, \tag{14}$$

where $C = [0, T]^2 \backslash \{(x, x) \in \mathbb{R}^2 : 0 \le x \le T\}$.
(iii) *In particular, for q_2 locally integrable,*

$$\mathrm{Var}(\hat{\beta}_{\eta}) = \int \int_C \psi_{\eta}(u)\psi_{\eta}(v)q_2(u, v)\mathrm{d}u\mathrm{d}v + \int_0^T \psi_{\eta}^2(u)p(u)\mathrm{d}u. \tag{15}$$

Proof (i) Since

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

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

(ii) Apply Proposition 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 (14) becomes

$$\text{Cov}(\hat{\beta}_\eta, \hat{\beta}_\xi) = \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt = E \int_0^T \psi_\eta(t)\psi_\xi(t)dN(t) \qquad (16)$$

and (15) reduces to

$$\text{Var}(\hat{\beta}_\eta) = \int_0^T \psi_\eta^2(t)p(t)dt = \int_0^T \psi_\eta^2(t)E\{dN(t)\} = E \int_0^T \psi_\eta^2(t)dN(t). \quad (17)$$

This leads us to propose the following expressions as estimators of (16) and (17),

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

and

$$\widehat{\text{Var}}(\hat{\beta}_\eta) = \int_0^T \psi_\eta^2(t)dN(t), \qquad (18)$$

respectively, which are obviously unbiased.

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

$$V_{\xi,0} = \beta_\xi, \;\; \hat{V}_{\xi,0} = \hat{\beta}_\xi, \;\; V_{\xi,n+1} = \text{Var}(\hat{V}_{\xi,n}), \;\; \hat{V}_{\xi,n} = \int_0^T \psi_\xi^{2^n} dN(t), \;\; n \geq 0.$$

By direct substitution of (12) into (17) we obtain

$$\text{Var}(\hat{\beta}_\xi) = \int_0^T \psi_\xi^2(t) \sum_\eta \beta_\eta \psi_\eta(t)dt = \sum_\eta \beta_\eta \int_0^T \psi_\xi^2(t)\psi_\eta(t)dt \qquad (19)$$

Defining

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

we have that (19) can be written as

$$V_{\xi,1} = \text{Var}(\hat{\beta}_\xi) = \sum_\eta \beta_\eta K_{\xi,2}^\eta. \qquad (20)$$

Now, let us compute the variance of the estimator (18):

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

Thus, by Proposition 3 with $f = g = \psi_\xi^2$ we have

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

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

$$V_{\xi,2} = \int_0^T \psi_\xi^4(t)p(t)\mathrm{d}t = \int_0^T \psi_\xi^4(t)E\mathrm{d}N(t) = E\int_0^T \psi_\xi^4(t)\mathrm{d}N(t).$$

Now

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

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

$$K_{\xi,4}^\eta = \int_0^T \psi_\xi^4(t)\psi_\eta(t)\mathrm{d}t,$$

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

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

Defining

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

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

we get the following result.

Theorem 3 *If N is a NIC point process, satisfying Assumption 1, then*

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

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

Proof Immediate, using Proposition 3. with $f = g = \psi_\xi^{2^n}$ and the fact that $q_2(u,v) = 0$. $\square$

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 $\tau_0, \tau_1, \ldots, \tau_{m-1}$, this expression reduces to $\sum_{i=0}^{m-1} \psi_\xi^{2^n}(\tau_i)$.

We observe that Theorem 3 is formal and has its full meaning for the cases where the series converge. Nonetheless, finite approximations will always have their practical meaning. See de Miranda (2005) for a general treatment of sequences like the one above.

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

3.2 Estimation of the intensity function

We are now in position to estimate the intensity function p. We will consider first a linear estimate based on a maximum scale and then a thresholded estimator.

In the first situation, we restrict the scales up to a maximum scale J. Let $_d\mathbb{Z}_e = \{z \in \mathbb{Z} | d \leq z \leq e\}$ and $Ze(l)_J = \mathbb{Z} \cup (_l\mathbb{Z}_J \times \mathbb{Z})$.

Define the estimated intensity function by

$$\hat{p}_J = \sum_{\eta \in Ze(l)_J} \hat{\beta}_\eta \psi_\eta = \sum_{j \leq J} \hat{\beta}_\eta \psi_\eta, \tag{21}$$

noticing that when η is an ordered pair it is represented by (j, k). We will use $l = 0$ and $J \geq 0$.

Then we have the following result.

Theorem 4 *If N satisfies Assumption 1, then*

  (i)   *$\hat{p}_J$ is an asymptotically unbiased estimator for the intensity function p.*
  (ii) *Additionally if q_2 exists and it is locally integrable, then*

$$\text{Var}(\hat{p}_J) = \sum_{Ze(0)_J} \sum_{Ze(0)_J} \left(\int \int_C \psi_\eta(u)\psi_\xi(v)q_2(u, v)dudv \right) \psi_\eta \psi_\xi$$

$$+ \sum_{Ze(0)_J} \sum_{Ze(0)_J} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt \right) \psi_\eta \psi_\xi.$$

  If N is a NIC point process, then
  (iii) *$\text{Var}(\hat{p}_J) = \sum_{Ze(0)_J} \sum_{Ze(0)_J} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)dt \right) \psi_\eta \psi_\xi$, and*
  (iv) *$\widehat{\text{Var}}(\hat{p}_J) = \sum_{Ze(0)_J} \sum_{Ze(0)_J} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)dN(t) \right) \psi_\eta \psi_\xi$ is an unbiased estimator for $\text{Var}(\hat{p}_J)$.*

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) $\text{Var}(\hat{p}_J) = E(\sum_{\eta \in Ze(0)_J} (\hat{\beta}_\eta - \beta_\eta)\psi_\eta)^2 = E\left(\sum_{\xi \in Ze(0)_J} \sum_{\eta \in Ze(0)_J} (\hat{\beta}_\eta - \beta_\eta)\right.$
$(\hat{\beta}_\xi - \beta_\xi)\psi_\eta\psi_\xi) = \sum_{\xi \in Ze(0)_J} \sum_{\eta \in Ze(0)_J} \text{Cov}(\hat{\beta}_\eta, \hat{\beta}_\xi)\psi_\eta\psi_\xi.$
Therefore, in the general case,

$$\text{Var}\,(\hat{p}_J) = \sum_{\eta \in Ze(0)_J} \sum_{\xi \in Ze(0)_J} \left(\int \int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)\mathrm{d}u\mathrm{d}v \right.$$
$$+ \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)\mathrm{d}t \Bigg) \psi_\eta\psi_\xi.$$

(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 $p(t)\mathrm{d}t = E\mathrm{d}N(t)$.   □

The next result gives a bound for the squared integrated bias of $\hat{p}_J$, measured in the L^2 norm, in the case of p under Assumption 2.

Denote by $\text{ess sup}_A f$ ($\text{ess inf}_A f$) the essential supremum (infimum) of a function f defined on the set A. For the ease of notation we will write $\text{ess sup}_\eta f$ instead of $\text{ess sup}_{\text{supp}\psi_\eta} f = \text{ess sup}_{x \in \text{supp}\psi_\eta} f(x)$, where $\text{supp}\,\psi_\eta$ indicates the support of the wavelet ψ_η.

Theorem 5 *Let $\{\psi_\eta | \eta \in Ze(0)\}$, 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 1. Suppose that p, the intensity function of N restricted to $[0, T]$, satisfies Assumption 2. Then,*

$$\|E(\hat{p}_J) - p\|^2 \leq \frac{K^2 M^2 |\text{supp}\psi_{(0,0)}|^{2(\alpha+1)}}{(1 - 2^{-2\alpha})} \left(\frac{1}{2^{2\alpha}}\right)^{J+1} I_{(0,1]}(\alpha), \qquad (22)$$

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

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

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

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

$$= \sum_{\eta \in \mathbb{Z}^2, j > J} \beta_\eta^2.$$

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,

$$\beta_\eta = \int \psi_\eta p \mathrm{d}t = \int \psi_\eta^+ p \mathrm{d}t - \int \psi_\eta^- p \mathrm{d}t$$

$$\leq \int \psi_\eta^+ \mathrm{ess} \sup_\eta p \mathrm{d}t - \int \psi_\eta^- \mathrm{ess} \inf_\eta p \mathrm{d}t$$

$$= \int (\psi_\eta^+ - \psi_\eta^-) \mathrm{ess} \inf_\eta p \mathrm{d}t + \int \psi_\eta^+ (\mathrm{ess} \sup_\eta p - \mathrm{ess} \inf_\eta p) \mathrm{d}t$$

$$= \mathrm{ess} \inf_\eta p \int \psi_\eta \mathrm{d}t + (\mathrm{ess} \sup_\eta p - \mathrm{ess} \inf_\eta p) \int \psi_\eta^+ \mathrm{d}t$$

$$\leq 0 + K(b_\eta - a_\eta)^\alpha (b_\eta - a_\eta) \mathrm{ess} \sup_\eta \psi_\eta^+$$

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

Analogously,

$$\beta_\eta \geq \int \psi_\eta^+ \mathrm{ess} \inf_\eta p \mathrm{d}t - \int \psi_\eta^- \mathrm{ess} \sup_\eta p \mathrm{d}t$$

$$\geq 0 - K(b_\eta - a_\eta)^\alpha \int \psi_\eta^- \mathrm{d}t \geq -K(b_\eta - a_\eta)^{\alpha+1} 2^{j/2} |\mathrm{ess} \inf_{(0,0)} \psi_{(0,0)}|.$$

Let $M = \max\{\mathrm{ess} \sup_{[0,T]} \psi_{(0,0)}, -\mathrm{ess} \inf_{[0,T]} \psi_{(0,0)}\}$. Then we can write

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

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

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

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

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

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

Now, if $\alpha > 1$ and $A \subset \mathbb{R}$ is a real line interval, every essentially α-Hölder function is constant on $A \backslash D$, and, in this case, p is essentially constant so that we have $\beta_\eta = \int \psi_\eta p \mathrm{d}t = 0$ for all $J \geq 0$, hence $\|p - E(\hat{p}_J)\|^2 = \sum_{\eta \in \mathbb{Z}^2, j>J} \beta_\eta^2 = 0$. This completes the proof. $\qquad \square$

The preceding theorem guarantees at least an exponential decay with J for the squared integrated bias of $\hat{p}_J$.

We consider now a non-linear thresholded estimator. A threshold function $\delta : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a measurable function such that $0 \leq \delta(x, y) \leq 1$ and, for each y, $\delta(x, y) = 1$ if $|x| \geq y$, $\delta(x, y)$ is non-decreasing over $[0, y]$ and non-increasing over $[-y, 0]$. If $\delta(x, y) = 1$, when $|x| \geq y$ and $\delta(x, y) = 0$ otherwise, we say that δ is a

hard threshold. This is the kind of threshold that we will use. We may also use soft thresholds.

Define the estimator as:

$$\hat{p}_{J,\lambda}^{\delta} = \sum_{\eta \in Ze(0)_J} \delta\left(\hat{\beta}_\eta, \lambda\sqrt{\widehat{\text{Var}}(\hat{\beta}_\eta)}\right)\hat{\beta}_\eta\psi_\eta = \sum_{j \le J}\delta\left(\hat{\beta}_\eta, \lambda\sqrt{\widehat{\text{Var}}(\hat{\beta}_\eta)}\right)\hat{\beta}_\eta\psi_\eta. \quad (23)$$

The parameter λ may be chosen according to several criteria. Other thresholding methodologies exist. See Morettin (1999) for details.

We have the following result.

Proposition 4 *Let N satisfying Assumption* 1 *and choose δ as a hard threshold. Let*

$$Ze(0)_{J,\lambda}^{\delta} = \left\{\eta \in Ze(0)_J | \delta\left(\hat{\beta}_\eta, \lambda\sqrt{\widehat{\text{Var}T}(\hat{\beta}_\eta)}\right) = 1\right\}.$$

then:

(i) *If q_2 exists and it is locally integrable,*

$$\text{Var}(\hat{p}_{J,\lambda}^{\delta}) = \sum_{\eta,\xi \in Ze(0)_{J,\lambda}^{\delta}} \left(\int\int_C \psi_\eta(u)\psi_\xi(v)q_2(u,v)\text{d}u\text{d}v \right.$$

$$\left. + \int_0^T \psi_\eta(t)\psi_\xi(t)p(t)\text{d}t\right)\psi_\eta\psi_\xi.$$

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

(ii) $\text{Var}(\hat{p}_{J,\lambda}^{\delta}) = \sum_{\eta,\xi \in Ze(0)_{J,\lambda}^{\delta}} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)p(t)\text{d}t\right)\psi_\eta\psi_\xi.$

(iii) $\widehat{\text{Var}}(\hat{p}_{J,\lambda}^{\delta}) = \sum_{\eta,\xi \in Ze(0)_{J,\lambda}^{\delta}} \left(\int_0^T \psi_\eta(t)\psi_\xi(t)\text{d}N(t)\right)\psi_\eta\psi_\xi$ *is an unbiased estimator for* $\text{Var}(\hat{p}_{J,\lambda}^{\delta})$.

Proof Since

$$\hat{p}_{J,\lambda}^{\delta} = \sum_{\eta \in Ze(0)_J} \delta(\hat{\beta}_\eta, \lambda\sqrt{\widehat{\text{Var}}(\hat{\beta}_\eta)})\hat{\beta}_\eta\psi_\eta = \sum_{\eta \in Ze(0)_{J,\lambda}^{\delta}} \hat{\beta}_\eta\psi_\eta,$$

it is sufficient to repeat the argument used in the proof of Theorem 4. □

We remark that a theorem similar to Theorem 5 holds in this case: the squared bias is bounded by a sum of two terms, one corresponding to the exponential decay with J and another corresponding to the threshold. See de Miranda (2003) for details.

4 An application

We will present here an application that illustrates the use 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. A set of $T = 4{,}225$ returns will be used, corresponding to the period of time from January 2nd 1986 to September 26, 2002.

For time series models that can be written as $X(t) = m(t, X(t-1), X(t-2), \ldots) + s(t, X(t-1), X(t-2), \ldots)\epsilon_t$, $t \in \mathbb{Z}$, the point process, N_v, generated by the peaks over a threshold v, i.e, the point process whose occurrence times are those times for which $|X(t)| > v$, is such that for all bounded disjoint subsets A and B of $\mathbb{R}$ we have $\text{Cov}(N_v(A), N_v(B)) \to 0$ as $v \to \infty$. Here $m : \mathbb{R} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ and $s : \mathbb{R} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}_+^*$ are bounded functions, $s : \mathbb{R} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}_+^*$ is bounded bellow away from zero, and $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed innovations with zero mean and finite variance.

Log returns of financial time series have been successfully described by models that belong to or that, in practice, can be considered to belong to the class of models above. We so say meaning that in practice the boundedness conditions in the class above are irrelevant. If the log returns can be fitted by a model using unbounded functions, they so can be by a model with large bounds. The same is true for very small infimum of s.

Now, the higher the threshold v is the more confident we are in assuming our process is NIC or, at least, is well approximated by a NIC process. However, if we choose too a high threshold few log returns will generate occurrences of the point process and the estimation of the intensity will be jeopardized. A low threshold will generate more occurrences for the point process, but we will not be confident we can apply the results obtained for NIC processes. We remind that these point processes are not the same. In general different values of v generate different point processes. In this way, for a given data set, there will be values of v, and corresponding point processes, for which we will not be able to properly apply the technique developed in the previous sections. In our example of application, a parsimonious solution for this tradeoff was found in an "ad hoc" manner.

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, namely 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 $(558/4{,}225)c$ events within an interval of length c. Under this assumption one will expect to have $558/2^6 \approx 8$ events lying 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 lying 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 $\widehat{\text{Var}}(\hat{\beta}_\eta)$, 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 actually 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, to the difference among 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_R, then the variance of the coefficients that belong to the J-th order scale is an asymptotically unbiased estimator of λ_R, that is, $E\left(\mathrm{Var}\{\hat{\beta}_{(0,J)}, \ldots, \hat{\beta}_{(2^J-1,J)}\}\right) \rightarrow \lambda_R$, as $J \rightarrow \infty$. 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 (2003).

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

$$\psi_{(0,0)} = T^{-1/2}(I_{[0,T/2)} - I_{[T/2,T)}), \quad \phi_{(0,0)} = T^{-1/2}I_{[0,T)}, \quad T = 4{,}225$$

and

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

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

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

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

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

and

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

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

Analogously, we have obtained $\hat{\beta}_0$ and $\text{Var}\hat{\beta}_0$. 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, 558/4,225, that is, the mean value of p.

The threshold function chosen was $\delta(x, y) = 0$ for $|x| < y$ and $\delta(x, y) = 1$ for $|x| \geq y$.

Observe that the use of a hard threshold on $\hat{\beta}_\eta$ is equivalent to testing the hypothesis $\beta_\eta = 0$ against $\beta_\eta \neq 0$ based on $\hat{\beta}_\eta$. We will accept $\beta_\eta = 0$ if $|\hat{\beta}_\eta| < \lambda\sqrt{\widehat{\text{Var}}(\hat{\beta}_\eta)}$ and reject it if $|\hat{\beta}_\eta| \geq \lambda\sqrt{\widehat{\text{Var}}(\hat{\beta}_\eta)}$, and, in this case, we will assume the value of β_η is given by its estimate $\hat{\beta}_\eta$. We recall that for $\lambda = 3$ we have a "confidence level" of at least $1 - (1/3)^2 = 8/9$ or approximately 88, 8% whatever the distribution of $\hat{\beta}_\eta$ is. This is a consequence of Chebychev's inequality. In Fig. 1 we show the number of counts and in Fig. 2 the estimated intensity. We clearly see the non-homogeneous character of the point process. In Figs. 3 and 4 we have the estimated standard deviation and the respective threshold version. Figures 5 and 6 show the estimated intensity and thresholded estimated intensity, respectively, with their (non-negative) confidence bands. More specifically, Fig. 5 is the graph of $\hat{p}_J$ along with the lower and upper boundaries of the pointwise confidence band, $\max\{0, \hat{p}_J - \mu\sqrt{\widehat{\text{var}}(\hat{p}_J)}\}$ and $\hat{p}_J + \mu\sqrt{\widehat{\text{var}}(\hat{p}_J)}$, where $J = 5$ and $\mu = 3$. Similarly, Fig. 6 depicts $\hat{p}_{J,\lambda}^\delta$ and the borders $\max\{0, \hat{p}_{J,\lambda}^\delta - \mu\sqrt{\widehat{\text{var}}(\hat{p}_{J,\lambda}^\delta)}\}$ and $\hat{p}_{J,\lambda}^\delta + \mu\sqrt{\widehat{\text{var}}(\hat{p}_{J,\lambda}^\delta)}$ where $J = 5$, δ is the hard threshold, $\lambda = 3$ and $\mu = 3$. Since $\mu = 3$, Chebychev's inequality guarantees a "confidence level" of at least 88.8% for these pointwise confidence bands. See de Miranda (2003, 2005) for details. Again, these last figures confirm the non-homogeneity of the fitted NIC point process. 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.

5 Further comments

In this work we dealt with the problem of estimating the time-variable intensity of a non-homogeneous 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. See de Miranda and Morettin (2005)

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
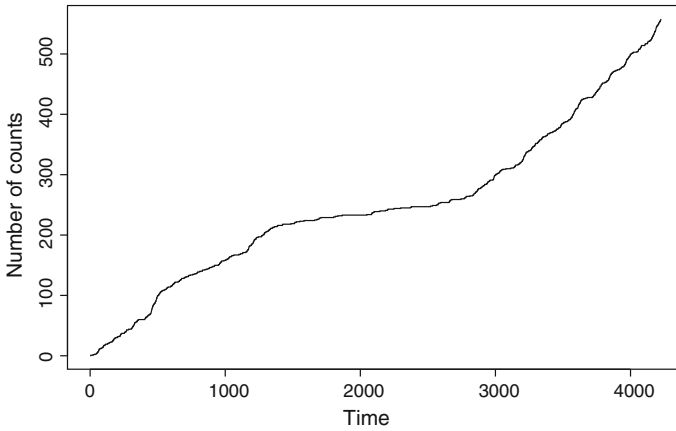
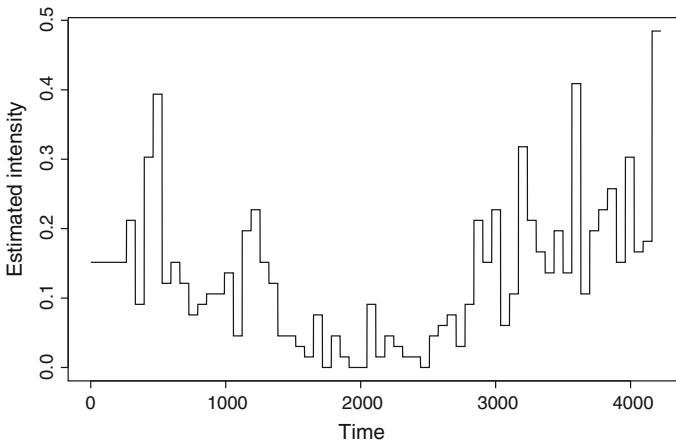**Fig. 1** Accumulated number of events

Fig. 2 Estimated intensity

depend on the estimate of the intensity of the noisy process. Similar results to those obtained here can be derived. See de Miranda (2003).

For the case of non-homogeneous Poisson processes on $\mathbb{R}^d$, it is possible to obtain the probability density function of the empirical wavelet coefficients $\hat{\beta}_\eta$. The key feature here is the independence in the internal probability structure. The following result is valid.

Proposition 5 *Suppose N is a Poisson process on $\mathbb{R}^d$ with intensity function p_N : $\mathbb{R}^d \to \mathbb{R}_+$. Let ψ_η be a continuous compactly supported wavelet. Denote $f_\eta : \mathbb{R} \to \mathbb{R}_+$ the probability density function of $\hat{\beta}_\eta = \int_{\mathbb{R}^d} \psi_\eta dN$. Then f_η is given by the principal value*
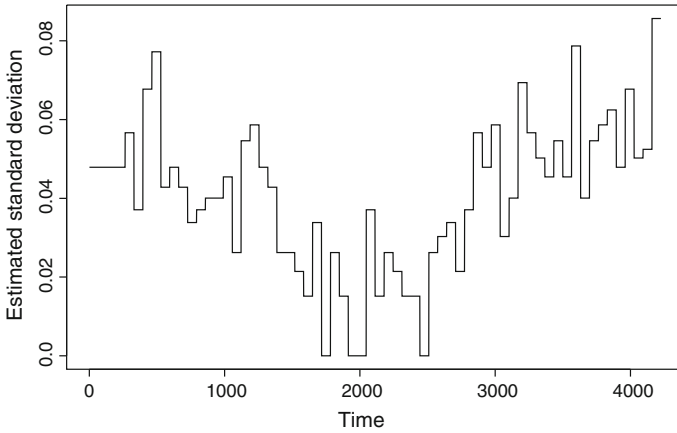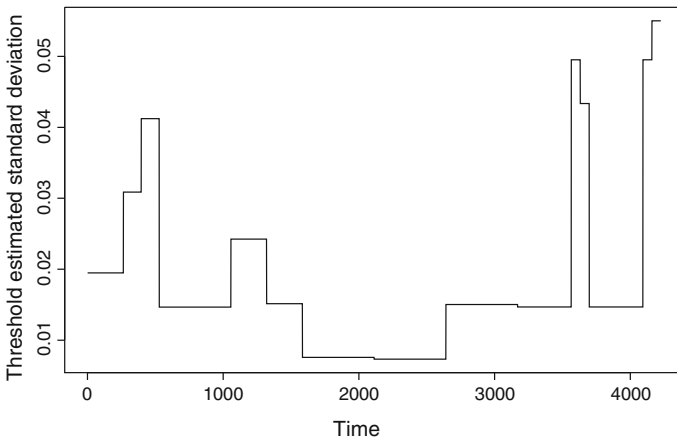
Fig. 3 Estimated standard deviation

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

$$f_\eta(y) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(\int_{\text{supp}\psi_\eta} p_N(x) \left(\cos\left(w\psi_\eta(x)\right) - 1\right) dx \right)$$

$$\times \cos\left(\int_{\text{supp}\psi_\eta} p_N(x) \sin\left(w\psi_\eta(x)\right) dx - wy \right) dw. \qquad (24)$$

From this result we can obtain the characteristic function of $\hat{\beta}_\eta$ as well as the variance, asymmetry and kurtosis coefficients. The case of the Haar wavelet has to be treated separately, since it does not satisfy the assumption of the proposition. In case we have n independent replications of the Poisson process, the estimators $\tilde{\beta}_\eta = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_\eta(i)$ and $\tilde{p}_J = \frac{1}{n}\sum_{i=1}^{n} \hat{p}_J(i)$, where $\hat{\beta}_\eta(i)$ and $\hat{p}_J(i)$ are the estimated wavelet coefficient and intensity obtained from the ith observation, inherit the unbiasedness of $\hat{\beta}_\eta(i)$ and the asymptotic unbiasedness with the scale J of $\hat{p}_J(i)$ respectively. As a
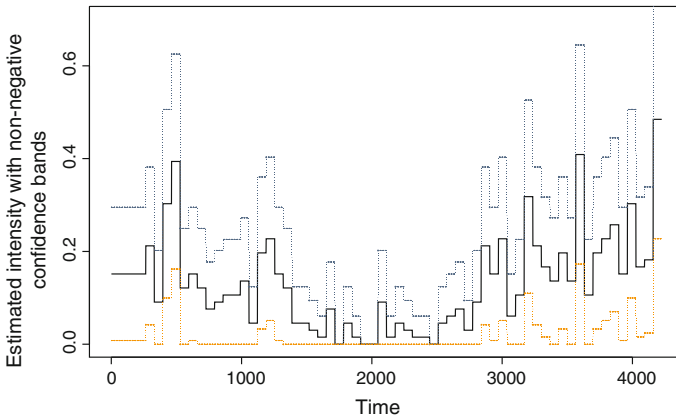
Fig. 5 Estimated intensity with non-negative bands using $\mu = 3$
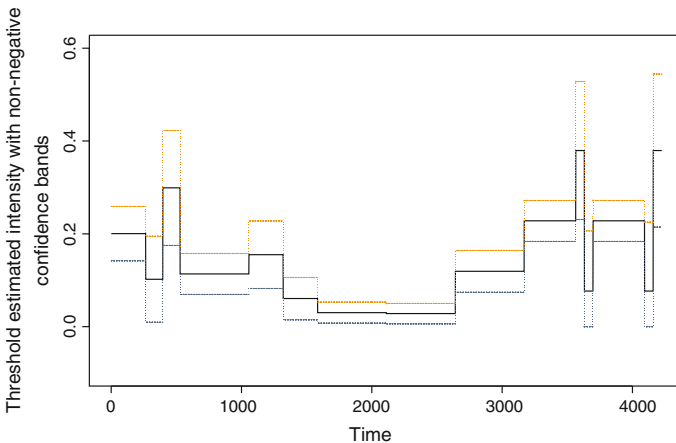

Fig. 6 Estimated intensity under threshold with non-negative bands using $\lambda = 3$ and $\mu = 3$

consequence of the finiteness of the first and second moments of the wavelet coefficient estimators $\hat{\beta}_\eta(i)$, the estimator $\tilde{\beta}_\eta$ is asymptotically normal. The asymptotic normality of $\tilde{\beta}_\eta$ is also valid for NIC point processes in case of independent replications. See de Miranda (2008) for the proofs and further details.

We finally remark that the intensity estimators suggested may be negative, a fact that occurs also with estimators of a density function via wavelets, as suggested for example by Donoho et al. (1996), but this seldom occurs. It can be proved that for the Haar wavelets the estimators are always positive. We could consider also expanding $\exp\{p\}$ in a wavelet series, but this would make the estimation procedure more complex.

References

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics, 6*, 701–726.

Besbeas, P., De Feis, I., 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. (1998). Some wavelet analysis of point processes data. In *The third first asilomar conference on signals, systems and computers* (pp. 1087–1091). New York: IEEE Computer Society.

Cramér, H., Leadbetter, M. R. (1967). *Stationary and related stochastic processes*. New York: Wiley.

Daley, D. J., Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer.

Daubechies, I. (1992). Ten lectures on wavelets. In *CBMS-NSF regional conference series in applied mathematics*. Philadelphia: SIAM.

de Miranda, J. C. S. (2003). 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. (2005). Sure inference analysis. Technical report no. 16. Department of Mathematics, University of São Paulo.

de Miranda, J. C. S. (2008). Probability density functions of the empirical wavelet coefficients of a wavelet multidimensional Poisson intensity estimator. *Brazilian Journal of Probability and Statistics, 22*, 157–164.

de Miranda, J. C. S. (2009). Non internally correlated point processes. Technical report no. 6. Department of Statistics, University of São Paulo.

de Miranda, J. C. S., Morettin, P. A. (2005). Estimation of the density of point processes on $\mathbb{R}^m$ via wavelets. Technical report no. 9. Department of Mathematics, University of São Paulo.

Donoho, D., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage: Asymptotia? (With discussion). *Journal of the Royal Statistical Society, Series B, 57*, 301–369.

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

Hall, P., Patil, P. N. (1995). Formulae for the mean integrated squared error of nonlinear wavelet-based density estimators. *The Annals of Statistics, 23*, 905–928.

Hall, P., Patil, P. N. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *Journal of the Royal Statistical Society, Series B, 58*, 361–378.

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

Helmers, R., Mangku, I. W., Zitikis, R. (2005). Statistical properties of a kernel-type estimator of the intensity function of a cyclic Poisson process. *Journal of Multivariate Analysis, 84*, 1–23.

Helmers, R., Mangku, I. W., Zitikis, R. (2007). A non-parametric estimator for the doubly periodic Poisson intensity function. *Statistical Methodology, 4*, 481–492.

Kingman, J. F. C. (1993). *Poisson processes*. Oxford: Clarendom Press.

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

Krickeberg, K. (1982). Processus ponctuels en statistique. In P. L. Hennequin (Ed.), *Lecture notes in mathematics* (Vol. 929, pp. 205–313). New York: Springer.

Kutoyants, Yu. A. (1984). *Parameter estimation for stochastic processes*. Berlin: Heldermann.

Kutoyants, Yu. A. (1998). Statistical inference for spatial point processes. In *Lecture notes in statistics* (Vol. 134). New York: Springer.

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

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

Patil, P. N., Wood, A. T. A. (2004). Counting process intensity estimation by orthogonal wavelet methods. *Bernoulli, 10*, 1–24.

Ramlau-Hansen, H. (1983). Smoothing counting process intensity by means of kernel functions. *The Annals of Statistics, 11*, 453–466.

Rathbun, S. L., 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., 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.