Multiresolution analysis of point processes and statistical thresholding for Haar wavelet-based intensity estimation

Youssef Taleb and Edward A K Cohen

Department of Mathematics, Imperial College London,
South Kensington Campus, London, U.K.*

Abstract

We take a wavelet based approach to the analysis of point processes and the estimation of the first order intensity under a continuous time setting. A Haar wavelet multiresolution analysis of a point process is formulated which motivates the definition of homogeneity at different scales of resolution, termed J-th level homogeneity. Further to this, the activity in a point process' first order behavior at different scales of resolution is also defined and termed L-th level innovation. Likelihood ratio tests for both these properties are proposed with asymptotic distributions provided, even when only a single realization of the point process is observed. The test for L-th level innovation forms the basis for a collection of statistical strategies for thresholding coefficients in a wavelet based estimator of the intensity function. These thresholding strategies outperform the existing local hard thresholding strategy on a range of simulation scenarios. The presented methodology is applied to NetFlow data to demonstrate its effectiveness at characterizing multiscale behavior on computer networks.

1 Introduction

The development of wavelet theory has been one of the most significant advances in signal and image processing. Wavelets' ability to decompose an object at different scales makes them ideal for understanding underlying structures in random processes. Based on their success in analyzing time series (Percival and Walden, 2000), there has been an ever increasing interest in applying wavelets to point processes (e.g. Brillinger, 1997; Cohen, 2014). Representing a point process as $N(A)$, a random integer indicating the number of events that have occurred in the set $A \subset \mathbb{R}$, one may use the notation $N(t)$ to be equal to

*Correspondence should be addressed to e.cohen@imperial.ac.uk

$N((0,t])$ for $t > 0$, $-N((t,0])$ for $t < 0$ and $N(0) = 0$ (Daley and Vere-Jones, 1988). Wavelets have most commonly been used to estimate the first order intensity (rate) function $\lambda : \mathbb{R} \to \mathbb{R}_+$ defined as $\lambda(t) = \mathrm{dE}\{N\}/\mathrm{d}t$. This is based on the fact we can represent any $L^2(\mathbb{R})$ function as a linear combination of basis functions. Namely, for some $j_0 \in \mathbb{Z}$ and an orthogonal wavelet basis generated by father and mother pair (ϕ, ψ),

$$\lambda(t) = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k}(t) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k}(t) \tag{1}$$

where $\phi_{j_0,k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, provided $\lambda \in L^2(\mathbb{R})$. To estimate λ, the task becomes estimating the set of coefficients $\{\alpha_{j_0,k} \equiv \langle\lambda, \phi_{j_0,k}\rangle; k \in \mathbb{Z}\}$ and $\{\beta_{j,k} \equiv \langle\lambda, \psi_{j,k}\rangle; j \geq j_0, k \in \mathbb{Z}\}$, where $\langle f_1, f_2\rangle = \int_{\mathbb{R}} f_1(t)f_2^*(t)\mathrm{d}t$ is the usual inner product on $L^2(\mathbb{R})$ and f_2^* is the complex conjugate of f_2. This can be achieved by computing the stochastic integrals $\widehat{\alpha}_{j_0,k} = \int_{\mathbb{R}} \phi_{j_0,k}(t)\mathrm{d}N(t) = \sum_{\tau_i\in\mathcal{E}} \phi_{j_0,k}(\tau_i)$ and $\widehat{\beta}_{j,k} = \int_{\mathbb{R}} \psi_{j,k}(t)\mathrm{d}N(t) = \sum_{\tau_i\in\mathcal{E}} \psi_{j,k}(\tau_i)$, where $\mathcal{E}$ is the set of random event times of the process. Both $\widehat{\alpha}_{j_0,k}$ and $\widehat{\beta}_{j,k}$ can easily be shown to be unbiased estimators of $\alpha_{j_0,k}$ and $\beta_{j,k}$, respectively. Restricting the wavelet reconstruction up to some maximum resolution $J \geq j_0$ in (1), one can construct the estimator

$$\widehat{\lambda}^J(t) = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{J}\sum_{k\in\mathbb{Z}} \widehat{\beta}_{j,k}\psi_{j,k}(t) \tag{2}$$

which is asymptotically unbiased as $J \to \infty$ under standard regularity assumptions on N (de Miranda and Morettin, 2011). As in the classical wavelet regression setting (Donoho, 1993), or when using wavelets to estimate probability density functions (Härdle et al., 1998), it is then typical that shrinkage or thresholding procedures are applied to the coefficients to reduce the variance of the estimator $\widehat{\lambda}^J$.

Estimating the intensity of a point process has of course been addressed numerous times in either parametric (e.g. Rathbun and Cressie, 1994) or non-parametric methods (e.g. Brillinger, 1975; Aalen, 1978; Ramlau-Hansen, 1983; Helmers and Zitikis, 1999; Patil and Wood, 2004). In the specific case of wavelet based estimation, a non-parametric method, the approaches can be split into discrete-time and continuous-time methods. Discrete time methods (e.g. Timmermann and Nowak, 1999; Kolaczyk, 1999; Kolaczyk and Dixon, 2000; Fryzlewicz and Nason, 2004) typically apply a discrete wavelet transform (DWT) to the aggregated process $\{N_t; t \in \mathbb{Z}\}$, where $N_t \equiv N(t+1) - N(t)$ and then perform a shinkage procedure. Besbeas et al. (2004) offers a comprehensive review of discrete time methods and provides a simulation study comparing various thresholding schemes.

Under the continuous time framework, the setting of this paper, Brillinger (1997) proposes the estimator in (2), as well as an estimator for the second-order intensity. The shrinkage procedure $\widehat{\beta}_{jk} \to w(\widehat{\beta}_{j,k}/s_{j,k})$ is proposed where $s_{j,k}$ is an estimate of the standard error in $\widehat{\beta}_{j,k}$ and $w(u) = (1 - u^{-2})_+$ is the

Tukey function. Although applied to California earthquake data, the properties of the estimator are not studied in any detail. (de Miranda, 2008) offers the first proper treatment of the continuous time formulation, providing the characteristic and density functions for the estimators of the coefficients $\{\alpha_{j_0,k}; k \in \mathbb{Z}\}$ and $\{\beta_{j,k}; j \geq j_0, k \in \mathbb{Z}\}$ in terms of the basis (ϕ, ψ) under Haar wavelets as well as any continuous compactly supported wavelet of known closed form. This result is theoretically interesting but cannot be readily exploited as wavelets that fulfil all these criteria are rare and exotic. This work is extended in (de Miranda and Morettin, 2011) to provide first and second order moments for the linear (no thresholding) intensity estimator for any compactly supported wavelet of known closed form. With $\mathbb{1}_A(x)$ representing the characteristic function of the set A, they also propose a hard threshold $\widehat{\beta}_{j,k} \to \widehat{\beta}_{j,k}(1 - \mathbb{1}_{[-\omega s_{j,k}, \omega s_{j,k}]}(\widehat{\beta}_{j,k}))$ (ω typically set to 3) but it is given little treatment. Further thresholding procedures in continuous time have been proposed in Bigot et al. (2013) under a Meyer wavelet basis and in Reynaud-Bouret and Rivoirard (2010) under any biorthogonal wavelet basis. Both of these estimators are shown to achieve near optimal performance in the asymptotic setting that M, the number of observed independent realizations of the point process, goes to infinity. Further, the thresholding procedure of Reynaud-Bouret and Rivoirard (2010) does not require a compactly supported and bounded intensity to achieve asymptotic optimality. However, both thresholds are proportional to $\log(M)$ and are therefore only non-zero when $M > 1$, which questions their applicability to practical situations where one may only ever be able to observe a single realization. A thresholding procedure that can be applied in the $M = 1$ setting and for which the statistical properties are still tractable is therefore clearly desirable.

In this paper we go beyond solely estimating the intensity and provide methodology for characterizing the multiscale properties of the point process. We consider a wavelet based multiresolution analysis of a temporal point process; the motivation of which is both theoretical and practical. We demonstrate properties such as homogeneity can be explored and characterized through a multiresolution approach. We further propose statistical thresholding procedures for estimating the intensity in a data driven way. Statistical thresholding has previously been considered in Abramovich and Benjamini (1995) in the classical wavelet regression setting. Here we adapt it for point processes and show it is capable of providing estimates with just a single realization of the process ($M = 1$), while being grounded in a statistically principled and tractable framework.

In Section 2 we provide a background to wavelet estimation of point process intensities. We extend existing results to show that the linear wavelet estimator of λ has a scaled Poisson distribution under a Poisson process and the Haar wavelet basis . Then in Section 3 we develop the theoretical framework for a wavelet-based multiresolution analysis of a point process. Considering the first order properties of a point process to be due to activity on different scales, under the Haar basis we define different levels of homogeneity, which we term J-th level homogeneity in reference to the particular scale J at which we are

analyzing the point process. We provide a likelihood ratio test (LRT) for these different levels of homogeneity for the class of Poisson processes, providing the asymptotic distribution for the LRT statistic under the null hypothesis. We then consider a further test for whether the intensity function exhibits activity at a particular scale, which we term L-th level innovation. Again, we provide a LRT for this property for the class of Poisson processes under the Haar wavelet basis.

In Section 4, we demonstrate how the LRT for L-th level innovation can be used as a method of statistical thresholding for wavelet coefficients, for which we propose three different forms: local, intermediate and global. Importantly, we demonstrate that under our LRT framework increasing M and increasing the intensity of the process are equivalent to one another, and hence indistinguishable in the asymptotic analysis. We are therefore able to use the asymptotic distributions to draw reliable inference and threshold the intensity in the $M = 1$ setting. We provide a comprehensive simulation study comparing the three different statistical thresholding procedures presented in this paper with the hard thresholding procedure given in de Miranda and Morettin (2011). We demonstrate that one or more of the proposed statistical thresholding procedures outperform this hard thresholding in almost all circumstances. In Section 5 we apply the presented methodology to real NetFlow data. In doing so, we demonstrate its effectiveness at revealing and characterizing multiscale behaviour on computer networks which could have a powerful impact in cyber-security, amongst other applications.

Further discussion on the LRTs, including boundary cases, can be found in Supplementary Material Section S1, all proofs are provided in Supplementary Material Section S2, and results of a comprehensive simulation study can be found in Supplementary Material Section S3, with a link to access the MATLAB code written to replicate the simulation study.

We focus on the Haar wavelet basis as the notion of homogeneity and innovation within its associated multiresolution analysis is natural, interpretable and tractable, albeit producing discontinuous estimates. Extensions of J-th level homogeneity and L-th level innovation to other wavelet bases are proposed in Supplementary Material Section S4. Similarly, a discussion on how the estimation and statistical thresholding procedures presented in this paper can be extended to Daubechies D4 wavelets can be found in Supplementary Material Section S4.1.

2 Wavelets and Estimation of the Intensity

In this section we provide a brief background to wavelet estimation of point process intensities. We will restrict ourselves to simple point processes with no fixed atoms, i.e. point processes that satisfy $N(\{t\}) \in \{0, 1\}$ almost surely for all $t \in \mathbb{R}$, and the probability of observing a point at any pre-specified location is zero.

2.1 Wavelets and multiresolution analysis

We summarize here essential definitions and results on wavelets that need to be stated prior to their application to the intensity function. The theory presented here follows the work of Meyer (1992).

Definition 1. *A **multiresolution approximation** of $L^2(\mathbb{R}^n)$ is an increasing sequence V_j, $j \in \mathbb{Z}$, of closed linear subspaces of $L^2(\mathbb{R}^n)$ with the following properties:*

1. $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$, $\bigcup_{j=-\infty}^{\infty} V_j$ *is dense in* $L^2(\mathbb{R}^n)$;
2. *for all* $f \in L^2(\mathbb{R}^n)$ *and* $j \in \mathbb{Z}$, $f(\cdot) \in V_j \iff f(2\cdot) \in V_{j+1}$;
3. *for all* $f \in L^2(\mathbb{R}^n)$ *and* $k \in \mathbb{Z}^n$, $f(\cdot) \in V_0 \iff f(\cdot - k) \in V_0$;
4. *there exists a function* $g \in V_0$, *such that the sequence* $g(\cdot - k), k \in \mathbb{Z}^n$, *is a Riesz basis of the space* V_0.

It is also shown in Meyer (1992) that for a Riesz basis $g(\cdot - k), k \in \mathbb{Z}^n$ of V_0, the sequence $\phi(\cdot - k), k \in \mathbb{Z}^n$ defined by $\Phi(\xi) = G(\xi)(\sum_{k\in\mathbb{Z}^n} |G(\xi + 2k\pi)|^2)^{-1/2}$ is the canonical orthonormal basis of V_0, where Φ and G are the Fourier transforms of ϕ and g, respectively. ϕ is called either the father wavelet or scaling function. In this paper, we are concerned with point processes on the real line, and therefore we focus on the space $L^2(\mathbb{R})$. Defining W_j to be the orthogonal complement of V_j in V_{j+1}, Definition 1 allows us to write

$$L^2(\mathbb{R}) = \overline{V_{j_0} \oplus \bigoplus_{j=j_0}^{\infty} W_j} \quad \text{or} \quad L^2(\mathbb{R}) = \overline{\bigoplus_{j=-\infty}^{\infty} W_j}. \tag{3}$$

The spaces V_j each have the basis $\{\phi_{j,k}(x) := 2^{j/2}\phi(2^j x - k), k \in \mathbb{Z}\}$ and are called the approximation spaces. The spaces W_j are called detail spaces and each have the orthonormal basis $\{\psi_{j,k}(x) := 2^{j/2}\psi(2^j x - k), k \in \mathbb{Z}\}$, where $\psi(x)$ is called the mother wavelet and is constructed from the father wavelet. The mappings $f(\cdot) \to 2^{j/2} f(2^j \cdot - k)$ are called dyadic transformations. Consequently, a fundamental result from (3) is that for any $j_0 \in \mathbb{Z}$, the set $\{\phi_{j_0,k}; k \in \mathbb{Z}\} \cup \{\psi_{j,k}; j \geq j_0, k \in \mathbb{Z}\}$ forms an orthonormal basis for $L^2(\mathbb{R})$. Furthermore, for any $j_0 \in \mathbb{Z}$ a function $f \in L^2(\mathbb{R})$ can be decomposed as

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j_0,k}\rangle \phi_{j_0,k}(x) + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x). \tag{4}$$

This identity, which illustrates the idea of multiscale analysis, will be used to decompose the first order intensity of a point process. In practice, a function

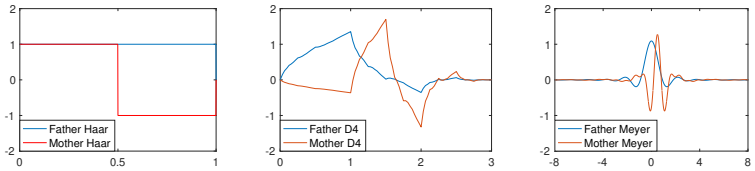


Figure 2.1: *Representation of three different wavelets. The Haar wavelet has a compact support and a closed form expression, the Daubechies D4 wavelet has a compact support only and the Meyer wavelet has a closed form expression only.*

$f \in L^2(\mathbb{R})$ is often approximated by its projection onto a specific approximation space $V_J = V_{j_0} \oplus \bigoplus_{j=j_0}^{J-1} W_j$, with $J > j_0$. Expansion (4) is then reduced to:

$$f^J(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{J,k}\rangle \phi_{J,k}(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j_0,k}\rangle \phi_{j_0,k}(x) + \sum_{j=j_0}^{J-1} \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k}(x). \tag{5}$$

As we increase J, the function $f^J \in V_J$ approximates f with ever increasing accuracy such that $\|f^J - f\|_2 \to 0$ as $J \to \infty$, where $\|\cdot\|_2 = \sqrt{\langle \cdot, \cdot \rangle}$ is the L^2 norm.

2.2 Continuous time wavelet estimator of the intensity

Consider a point process with a piecewise continuous intensity function $\lambda \in L^2(\mathbb{R})$, typically restricted to a finite length observation window $[0, T)$. We write the following wavelet expansion for this intensity (de Miranda and Morettin, 2011; Brillinger, 1997):

$$\lambda(t) = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k} \phi_{j_0,k}(t) + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} \beta_{j,k} \psi_{j,k}(t), \tag{6}$$

where $j_0 \in \mathbb{Z}$ is fixed and called the coarse resolution level, $\alpha_{j_0,k} = \langle \lambda, \phi_{j_0,k}\rangle$ and $\beta_{j,k} = \langle \lambda, \psi_{j,k}\rangle$. We are required to estimate the coefficients $\alpha_{j_0,k}$ and $\beta_{j,k}$ which we do so with $\widehat{\alpha}_{j_0,k} = \int \phi_{j_0,k}(t) \mathrm{d}N(t) = \sum_{\tau_i \in \mathcal{E}} \phi_{j_0,k}(\tau_i)$ and $\widehat{\beta}_{j,k} = \int \psi_{j,k}(t) \mathrm{d}N(t) = \sum_{\tau_i \in \mathcal{E}} \psi_{j,k}(\tau_i)$, where $\mathcal{E} = \{\tau_i, 1 \leq i \leq N(T)\}$ are the event times for one realization of a point process N on the time interval $[0, T)$. The observation window $[0, T)$ is often arbitrary or dictated by the application of interest. Hence the general linear estimator of the intensity function based on its wavelet expansion is:

$$\widehat{\lambda}(t) = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{j_0,k} \phi_{j_0,k}(t) + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} \widehat{\beta}_{j,k} \psi_{j,k}(t). \tag{7}$$

In the temporal case, $\mathrm{d}N(t)$ can also denote the differential process $N(t+\mathrm{d}t) - N(t)$. For a compactly supported wavelet function, Campbell's theorem (Daley and Vere-Jones, 1988, Chapter 6) gives us

$$\mathrm{E}\{\widehat{\alpha}_{j_0,k}\} = \int \phi_{j_0,k}(t) E\left\{\mathrm{d}N(t)\right\} = \int \phi_{j_0,k}(t)\lambda(t)\mathrm{d}t = \alpha_{j_0,k}$$
$$\mathrm{E}\{\widehat{\beta}_{j,k}\} = \int \psi_{j,k}(t) E\left\{\mathrm{d}N(t)\right\} = \int \psi_{j,k}(t)\lambda(t)\mathrm{d}t = \beta_{j,k},$$

showing the coefficient estimators to be unbiased. This is a linear estimator as it involves no shrinkage of the coefficients.

For obvious computational reasons, we can not in practice use an infinite wavelet basis to reconstruct the intensity (the intensity may only be fully reconstructed when we know that its decomposition is actually finite). Therefore, we firstly have to choose a maximum resolution level J. This maximum level plays a role in the bias-variance tradeoff of the estimator. Low values of J result in a smooth (high bias, low variance) estimator, whereas large values of J result in a noisy (low bias, high variance) estimator. The linear estimator then becomes the estimator of the projection of λ onto the space $V_J = V_{j_0} \oplus \bigoplus_{j=j_0}^{J-1} W_j$, and is denoted $\widehat{\lambda}^J$ from now on. In the case of linear estimation (no thresholding), the choice of j_0 is not relevant since it suffices to estimate the father wavelet coefficients from the approximation space V_J. Also, with compactly supported wavelets and events restricted to a finite length observation window $[0, T)$, the subset of translation indexes $k \in \mathbb{Z}$ satisfying $\widehat{\beta}_{j,k} \neq 0$ is finite. A non-linear estimator is obtained by adding a coefficient shrinkage term, determined from a thresholding strategy. The use of shrinkage methods in the classical wavelet regression setting is well studied (e.g. Donoho et al., 1995) and is used as a smoothing method to suppress contributing terms from fine scales which typically contain noise. For point process intensity estimation, while we do not have a noise term per se, shrinkage strategies are again desirable for smoothing, with fine scale terms typically having high variance. Thresholding also requires to choose the coarsest level of resolution j_0, ideally to some optimal value, since j_0 is also involved in the bias-variance tradeoff. When $j_0 \geq 0$, all the mother wavelet coefficients at the coarse levels $0 \leq j \leq j_0$ are kept. Donoho and Johnstone (1994) and Donoho et al. (1995) typically use $j_0 = 5$. From a further simulation study in Abramovich and Benjamini (1995), it is suggested that the choice for j_0 should be dependent on both the smoothness of the estimated function and the noise level.

When reconstructing the intensity of a point process, we have two desirable properties for a wavelet function. The first is that it should have a closed-form expression; it will be shown that this is required to compute the continuous linear estimator of the intensity function in an exact fashion, although in general not necessary for wavelet analysis and reconstruction of intensities. Second, the wavelet should be compactly supported; this is because invariably we can only observe the point process on a finite interval and therefore compactly supported

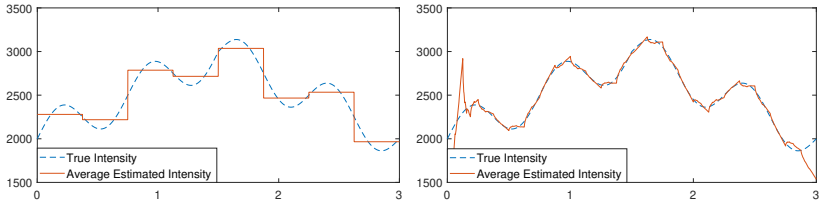


Figure 2.2: *Estimation of an example intensity with Haar and D4 wavelets obtained with an average over 1000 realizations of a point process on* $[0, 3]$*. We choose* $J = 3$ *here. The intensity is the sum of a triangular and a sine function.*

wavelets allow us to only consider a finite set of dyadic translations. In Figure 2.1 we show three examples of wavelet families; these are the Haar, Daubechies D4 and Meyer wavelets. Each family exhibits either one or both characteristics.

2.2.1 Haar estimator

The Haar mother and father wavelets are defined as

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \le t < 1/2 \\ -1 & \text{if } 1/2 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \phi(t) = \begin{cases} 1 & \text{if } 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases}.$$

These wavelets can be extended to the support $[0, T]$ with an orthonormality preserving rescaling $\psi_T(t) = T^{-1/2}\psi(t/T)$ and dyadic transforms of the type $\psi_{T,j,k}(t) = 2^{j/2}\psi_T(2^j t - kT)$. Henceforth, we will drop the subscript $_T$ and assume all wavelets are scaled for the support $[0, T]$. At each scale J the supports of $\phi_{J,k}$ and $\psi_{J,k}$ are of length $T/2^J$. Now consider a point process N on $[0, T)$. By construction, the interiors of Haar wavelets' supports are disjoint across all translations for a fixed scale, which implies that we only need 2^J wavelet coefficients when reconstructing on V_J, when $J \ge 0$. The linear estimator of the intensity function based on its Haar wavelet expansion becomes:

$$\widehat{\lambda}^J(t) = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \widehat{\beta}_{j,k}\psi_{j,k}(t) = \sum_{k=0}^{2^J-1} \widehat{\alpha}_{J,k}\phi_{J,k}(t),$$

for $j_0 \ge 0, J \ge j_0$.

Remark 1. *Under the Haar wavelet basis, at scale* $J \ge 0$ *and a translation* $0 \le k \le 2^J - 1$ *we have* $\alpha_{J,k} = \frac{1}{\sqrt{2}}(\alpha_{J+1,2k} + \alpha_{J+1,2k+1})$.

See proof in Supplementary Material S2.1. The linear estimator based on the Daubechies D4 wavelets is discussed in Supplementary Material Section S4.1.1.

2.2.2 Distribution of $\widehat{\lambda}^J$

In the case of Haar wavelets, we are able to derive the distribution of the estimator $\widehat{\lambda}^J$. The approximation space of interest, V_J, naturally induces a subdivision $S_J = \{s_k^J\}_{k=0}^{2^J-1}$ of the interval $[0,T)$. The elements of this subdivision, $s_k^J = [T\frac{k}{2^J}, T\frac{k+1}{2^J})$, are the non-zero intervals of the Haar wavelets at scale J and form 2^J disjoint subintervals of $[0,T)$. The Haar reconstruction of the intensity λ^J and its linear estimator $\widehat{\lambda}^J$ are piecewise constant functions, with forms $\lambda^J(t) = \sum_{k=0}^{2^J-1} \lambda_k^J \mathbb{1}_{s_k^J}(t)$ and $\widehat{\lambda}^J(t) = \sum_{k=0}^{2^J-1} \widehat{\lambda}_k^J \mathbb{1}_{s_k^J}(t)$, respectively. Hence we can establish the exact distribution for this estimator under a Poisson process model.

Proposition 1. *Under the Haar wavelet basis and for an inhomogeneous Poisson process N of intensity λ on $[0,T)$, $\widehat{\lambda}_0^J, ..., \widehat{\lambda}_{2^J-1}^J$ are independent random variables distributed as*

$$\widehat{\lambda}_k^J \sim \frac{2^J}{T}\text{Pois}(\mu_k^J), \quad 0 \leq k \leq 2^J - 1,$$

where $\mu_k^J = \int_{s_k^J} \lambda(t)\mathrm{d}t$.

The proof can be found in Supplementary Material S2.2. The result can also naturally be extended to any other point process with a square integrable intensity function for which the distribution of the event counts in any time interval is known (e.g. a binomial point process). It follows that $\mathrm{E}\{\widehat{\lambda}_k^J\} = \lambda_k^J = \frac{2^J}{T}\mu_k^J$, for all $0 \leq k \leq 2^J - 1$. We will now use Proposition 1 to develop likelihood ratio tests for two newly defined multiscale properties of a Poisson process.

3 A New Testing Protocol for Multiscale Properties of Poisson Processes

In this section, we will develop the theoretical framework for a wavelet-based multiresolution analysis of a point process. Considering the first order properties of a point process to be due to activity on different scales, under the Haar basis we define different levels of homogeneity under a multiresolution framework. We call this J-th level homogeneity, and provide a likelihood ratio test for it for the class of Poisson processes.

Under a compactly supported wavelet family, we then consider a more general setting to describe any activity of the intensity function at a particular scale, which we term L-th level innovation. We provide a likelihood ratio test for this property for the class of Poisson processes under the Haar basis. In Section 4, we will demonstrate how this test can be used as a method of thresholding coefficients in our wavelet estimator of the intensity function. In this section, it will be always assumed that $\lambda \in L^2(\mathbb{R})$.

3.1 Global behaviour: J-th level homogeneity

We use the Haar wavelet basis (rescaled if T is different than 1), because of its intuitive interpretation, its simplicity to implement and its amenability to statistical analysis. We consider the projection of the intensity on the Haar approximation space $V_J = V_{j_0} \oplus \bigoplus_{j=j_0}^{J-1} W_j$. With Haar wavelets, the reconstruction of the intensity at scale J is a piecewise constant function, and hence we can define a wavelet reconstruction vector $(\lambda_0^J, \lambda_1^J, \cdots, \lambda_{2^J-1}^J)^T$ where λ_k^J is the value of λ^J on the subinterval $s_k^J \in S_J, , k = 0, ..., 2^J - 1$. We use this formulation to define a property we call J-th level homogeneity.

Definition 2. *A point process N on $[0,T)$ with intensity λ is considered **level J homogeneous** if the reconstruction of the intensity at resolution J with Haar wavelets, or its projection on V_J, is constant on $[0,T)$. That is, $\lambda_0^J = \lambda_1^J = ... = \lambda_{2^J-1}^J$.*

Jth-level homogeneity was introduced in Taleb and Cohen (2016) in terms of the projection of the intensity on V_{J+1}. We propose that it is instead more convenient to base it on V_J, i.e. every point process is level 0 homogeneous as the projected intensity λ_0^0 on V_0 is always a constant on $[0,T)$. The concept of J-th level homogeneity goes side by side with the idea of a Haar multiresolution analysis of the intensity function, providing a natural way of studying on what scales the intensity function appears constant and hence the point process homogeneous, and on what scales the intensity function exhibits variability. If we define H_J as the class of level J homogeneous point processes, we have $H_J \supset H_{J+1}$. Indeed we know from Remark 1 that $\alpha_{J,k} = \frac{1}{\sqrt{2}}(\alpha_{J+1,2k} + \alpha_{J+1,2k+1})$ for Haar wavelets, and therefore $\lambda_0^J = \lambda_1^J = \cdots = \lambda_{2^J-1}^J$ if $\alpha_{J+1,0} = \alpha_{J+1,1} = \cdots = \alpha_{J+1,2^{J+1}-1}$.

Proposition 2. *Let N be a point process with a locally square integrable intensity λ. Then λ is constant almost everywhere on $[0,T)$ (i.e. $\lambda(t) = \lambda_0^0 = \frac{1}{T}\int_0^T \lambda(t)\mathrm{d}t$ almost everywhere) if and only if $N \in H_J$ for all $J \geq 0$.*

See Supplementary Material S2.3 for the proof. To avoid any confusion, we say that a point process with intensity λ is strictly homogeneous on $[0,T)$ when $\lambda(t) = \lambda_0^0$ for all $t \in [0,T)$. Proposition 2 illustrates how strict homogeneity can be loosely interpreted as the limit extension of Jth-level homogeneity. Furthermore, Definition 2 naturally leads us to define Jth-level inhomogeneity.

Definition 3. *A point process N on $[0,T)$ with intensity λ is considered **level J inhomogeneous** if it is level $J-1$ homogeneous and not level J homogeneous.*

We immediately remark that a level J inhomogeneous point process is not level j homogeneous for all $j \geq J$. Since J-th level homogeneity and inhomogeneity are based on the projection of the intensity function on the Haar approximation space V_J, they are descriptors of the first-order behaviour of the

point process when viewed at a particular scale. For instance, a point process may appear homogeneous when viewed at a coarse scale but show inhomogeneity when viewed at a finer resolution. An extension of J-th level homogeneity to other wavelets is proposed in Appendix S4.

3.2 Testing J-th level homogeneity

As the scope of this work is to analyse point processes in a multiscale fashion, we are not interested in testing the strict homogeneity of a Poisson process, which is the limit case for Definition 2 and has been thoroughly addressed in previous studies (e.g. Bain et al., 1985; Ng and Cook, 1999). We are instead aiming to statistically determine the resolution level where inhomogeneous behaviour appears. Recall that the choice of Haar wavelets implies that the wavelet reconstruction λ^J of the intensity λ, as well as the intensity estimator $\widehat{\lambda}^J$, are piecewise constant functions on the dyadic partition S_J. Although a piecewise analysis has also been carried out in Fierro and Tapia (2011) as a basis for a similar LRT, the wavelet approach presented here gives a natural, multiresolution scheme for defining the subdivision of the process. We begin by considering the LRT for equal means of scaled Poisson distributions, the results of which we can then utilize to test J-th level homogeneity of Poisson processes. This provides a comprehensive and rigorous treatment of the ideas first proposed in Taleb and Cohen (2016).

3.2.1 LRT for equal means of scaled Poisson distributions

Let $\mathbb{X} = \{\mathbf{X}_m\}_{m=1}^M$ be a set of iid scaled Poisson random vectors, each with independent components of form $\mathbf{X}_m = (X_{m,i})_{i=1}^P$, $X_{m,i} \sim \delta\text{Pois}(\mu_i)$. The scale parameter $\delta > 0$ is known and fixed so $\mathbf{X}_m$ is parametrized by the vector $(\mu_i)_{i=1}^P$. We consider testing the null hypothesis $H : \mu_1 = \cdots = \mu_P = \mu_c$ against the alternative hypothesis K that states H is not true. The LRT statistic is defined as

$$r = \frac{\sup_{\mu_c > 0} \mathcal{L}(\mathbb{X}; \mu_c, ..., \mu_c)}{\sup_{\{\mu_i\}_{i=1}^P, \sum \mu_i > 0} \mathcal{L}(\mathbb{X}; \mu_1, ..., \mu_P)}, \tag{8}$$

where $\mathcal{L}(\mathbb{X}; \mu_1, ..., \mu_P)$ is the likelihood of the data $\mathbb{X}$ given parameter vector $(\mu_i)_{i=1}^P$.

Proposition 3. *Let $R = -2\log(r)$, with r being the likelihood ratio statistic defined in (8). Then we have*

$$R = 2M \sum_{i=1}^{P} \bar{\mu}_i \log\left(\frac{\bar{\mu}_i}{\bar{\mu}_c}\right),$$

where $\bar{\mu}_c = \frac{1}{\delta MP} \sum_{i=1}^{P} \sum_{m=1}^{M} X_{m,i}$ *is the maximum likelihood estimator (MLE) for* μ_c*, the constant mean under the null hypothesis* H*, and* $\bar{\mu}_i = \frac{1}{\delta M} \sum_{m=1}^{M} X_{m,i}$ *is the MLE for* μ_i $(i = 1, ..., P)$*, under the alternative hypothesis* K*.*

See Supplementary Material S2.4 for the proof. If there exists at least one index i such that $\bar{\mu}_i = 0$, we use the convention $0 \log(0) = 0$. Further discussion on the absence of points within intervals can be found in Supplementary Material S1.2. Now let d_H be the number of free parameters under the null hypothesis H and let d_K be the number of free parameters under the alternative hypothesis K, then under the null hypothesis and regularity conditions on the likelihood functions that are met here, $R \to \chi^2_{d_K - d_H}$ as sample size $M \to \infty$ (see Wilks, 1938; Van der Vaart, 2000). In this setting, $d_K = P$ and $d_H = 1$. In practice, the $M = 1$ case is frequently encountered, and therefore we establish a more general and applicable result for the asymptotic distribution of R.

Theorem 1. *Let* $\mathbf{X}_1, .., \mathbf{X}_M$ $(M \geq 1)$ *be independent and identically distributed* P *dimensional random vectors where each* $\mathbf{X}_m = (X_{m,1}, ..., X_{m,P})^T$ *is constructed from independent components* $X_{m,i} \sim \delta \text{ Pois}(\mu_i)$*. Let* $R = -2 \log(r)$ *where* r *is the likelihood ratio statistic defined in (8). Then the distribution of statistic* R *is invariant to simultaneous changes in parameters* M *and* μ_i *provided that all products* $\mu_i M$ *,* $1 \leq i \leq P$*, remain constant. Furthermore, if* $\mu_1 = ... = \mu_P = \mu_c$*, then* $R \xrightarrow{d} \chi^2_{P-1}$ *as* $\mu_c M \to \infty$*.*

See Supplementary Material S2.6 for the proof[1]. It will now be shown that this result illustrates the practical advantage of Haar wavelets as it ensures that only one realization of the process is enough to conduct a LRT for J-th level homogeneity.

3.2.2 LRT for J-th level homogeneity of a Poisson process

Now let $\{N_m, m = 1, ..., M\}$ be a collection of $M \geq 1$ independent realizations of the same Poisson process N. Let $\Lambda = \{\mathbf{\Lambda}_m\}_{m=1}^{M}$ be the set of M independent random vectors where $\mathbf{\Lambda}_m = \left(\widehat{\lambda}^J_{m,k}\right)_{k=0}^{2^J-1}$ is the vector of all subinterval estimates of the intensity from N_m. From Proposition 1, $\mathbf{\Lambda}_m$ is a vector of independent scaled Poisson random variables and is therefore parametrized by the vector $\left(\lambda^J_k\right)_{k=0}^{2^J-1}$. We look to test the null hypothesis H which states N is level J homogeneous, i.e. $\lambda^J_0 = \cdots = \lambda_{2^J-1} = \lambda^J_c$ for some $\lambda^J_c > 0$, against the alternative hypothesis K which states H is not true. The LRT statistic in this

[1] It has been shown in Feng et al. (2012) that the classic asymptotic distributional result for the test statistic R does not hold if we are restricting ourselves to the $M = 1$ case and low values of μ_c ($\mu_c \leq 10$ in their study). This refutes the opposite claim in Brown and Zhao (2002), which possibly resulted from a confusion between the number of parameters P and the number M of independent realizations of the Poisson vector.

case is given as:

$$r^J = \frac{\sup_{\lambda_c^J>0} \mathcal{L}(\Lambda; \lambda_c^J, ..., \lambda_c^J)}{\sup_{\{\lambda_k^J\}_{k=0}^{2^J-1}, \sum \lambda_k^J>0} \mathcal{L}(\Lambda; \lambda_0^J, ..., \lambda_{2^J-1}^J)},$$

where $\mathcal{L}(\Lambda; \lambda_0^J, ..., \lambda_{2^J-1}^J)$ is the likelihood of the data Λ given parameter vector $\left(\lambda_k^J\right)_{k=0}^{2^J-1}$. Now using Proposition 3 we can write

$$R^J = -2\log(r^J) = 2\frac{M}{\delta^J}\sum_{k=0}^{2^J-1} \bar{\lambda}_k^J \log\left(\frac{\bar{\lambda}_k^J}{\bar{\lambda}_c^J}\right),$$

where $\delta^J = 2^J/T$, statistic $\bar{\lambda}_c^J$ is the maximum likelihood estimator (MLE) of λ_c^J and $\bar{\lambda}_k^J$ is the MLE for λ_k^J $(k = 0, ..., 2^J - 1)$, under the alternative hypothesis K. In this particular setting we have $d_K = 2^J$ and $d_H = 1$, giving R as asymptotically χ^2 distributed with $2^J - 1$ degrees of freedom under the conditions of Theorem 1. We reject J-th level homogeneity at significance level α if $R > c_\alpha$ where c_α, the critical value, is the upper $100(1-\alpha)\%$ point of the $\chi^2_{2^J-1}$ distribution.

3.2.3 Simulation study

Here, we demonstrate the LRT for J-th level homogeneity through simulations. We consider a class of inhomogeneous Poisson processes on a time interval $[0, T)$. These processes share a similar piecewise triangular intensity represented in Figure 3.1 and are defined as following:

$$\lambda(t) = \lambda_0\left(1 + (-1)^{i(t)}\xi\left(\theta(t) - \frac{1}{2}\right)\right),$$

where $i(t) \in \left\{0, \ldots, 2^{V+1} - 1\right\}$ is the index of the subinterval $s_{i(t)}^{V+1} = [\frac{i(t)}{2^{V+1}}T, \frac{i(t)+1}{2^{V+1}}T)$ in which t belongs, and $\theta(t)$ is defined by $t = (i(t) + \theta(t))T/2^{V+1}$. The absolute value of the gradient is $\xi\lambda_0 2^{V+1}/T$ and 2^V is the number of "triangles". The intensity takes values between $\lambda_0\frac{2-\xi}{2}$ and $\lambda_0\frac{2+\xi}{2}$ and its mean value λ_0^0 is the parameter $\lambda_0 > 0$. By construction, the quantity $\mu_0^0 = \int_0^T \lambda(t)dt = T\lambda_0$ does not depend on V, the process is level $V + 1$ homogeneous and level $V + 2$ inhomogeneous. We set the significance level of our test at $\alpha = 0.05$, with $M = 1$, i.e. we observe just a single realization. The empirical type 1 error and power of the LRT (over 10000 simulations) at different values of J are shown in Figure 3.1 as a function of λ_0, with $\lambda_0 \in [1000, 50000]$. In the example represented in Figure 3.1 where the process is level 2 homogeneous, the empirical type 1 error lies close to the 5% level as expected. When $J \geq 3$ and J-th level homogeneity no longer holds, the empirical power converges to 1 when $\lambda_0 \to \infty$. This behavior is expected as well. Indeed, this intensity model is proportional to λ_0

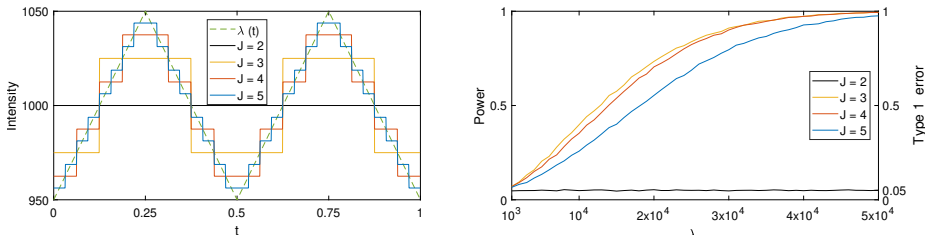


Figure 3.1: *Left: Haar wavelet reconstruction of a piecewise triangular intensity with* $V = 1$, $\xi = 0.1$ *and* $T = 1$ *at resolutions* $J \in \{2, 3, 4, 5\}$. *Right: Empirical type 1 error* ($J = 2$) *and power* ($J \in \{3, 4, 5\}$) *for this piecewise triangular intensity as a function of* λ_0.

and therefore its Haar reconstruction at any scale J satisfies $\lambda_k^J \propto \lambda_0$ as well as $\mu_k^J \propto \lambda_0$. Since statistic R^J tends to infinity as M increases towards infinity when $J \geq 3$ and a fixed λ_0, then the power of the LRT converges to 1. Hence the observed convergence of the empirical power to 1 when M is fixed and λ_0 increases towards infinity as ensured by Theorem 1. Similarly, the value of parameter ξ influences the speed of this convergence. Moreover, we note the power decreases as we increase J because the mass of the null distribution $\chi^2_{2^J-1}$ is displaced to the right as J increases, making it harder for the test to distinguish between the two hypotheses.

We also consider two scenarios where the parameter ξ is now dependent on λ_0 (variable ripple). The results for these two scenarios is given in Figure 3.2. When $\xi = \xi_0/\lambda_0$, the power decreases as λ_0 increases. Since $\lambda(t)$ takes values in the interval $\lambda_0 \frac{2-\xi}{2}$ and $\lambda_0 \frac{2+\xi}{2}$, the inhomogeneity from $J \geq 3$ due to the structure of the intensity is less detectable by the LRT as its amplitude decreases too quickly with ξ. When the parameter ξ is instead equal to $\xi_0/\sqrt{\lambda_0}$, the power stays maximal for the values of λ_0 considered in the simulation study. The amplitude of the intensity model decreases slowly enough as λ_0 increases such that the inhomogeneity is always detected by the LRT.

3.3 Local behaviour: L-th level innovation

In Section 2.1, we presented the decomposition $L^2(\mathbb{R}) = V_{j_0} \oplus \overline{\bigoplus_{j=j_0}^{\infty} W_j}$ where W_j is the orthogonal complement of V_j in V_{j+1} and often called the detail or innovation space. With J-th level homogeneity we focused on the behavior displayed on any space V_j, which brings together contributions from several resolutions. Projecting λ on W_j for increasing $j \geq j_0$, we explore the intensity function in progressively finer resolutions. To characterize this, we introduce the concept of L-th level innovation.

Definition 4. *Let* N *be a point process on* $[0, T)$ *with a square integrable intensity* λ. *We say that* N *possesses a* ***level L innovation*** *under the Haar basis if and only if there exists at least one index* $k \in \mathbb{Z}$ *such that* $\beta_{L,k} = \langle \lambda, \psi_{L,k} \rangle \neq 0$.

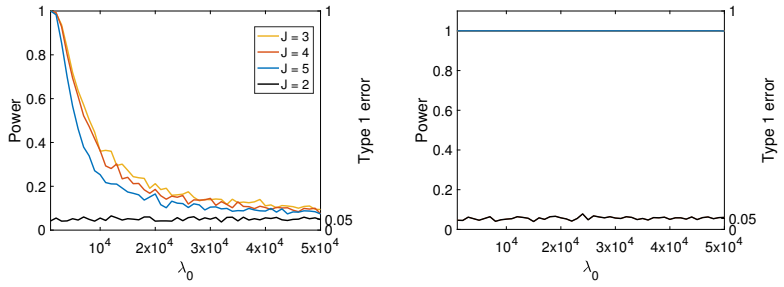


Figure 3.2: *Left: Empirical type 1 error ($J = 2$) and power ($J \in \{3, 4, 5\}$) for the piecewise triangular intensity where $\xi = \xi_0/\lambda_0$, with $\xi_0 = 1000$. Right: Empirical type 1 error ($J = 2$) and power ($J \in \{3, 4, 5\}$) for the piecewise triangular intensity where $\xi = \xi_0/\sqrt{\lambda_0}$, with $\xi_0 = \sqrt{1000}$.*

Since a point process on $[0, T)$ with constant intensity is level J homogeneous for all $J \geq 0$, it also displays no L-th level innovation irrespective of $L \geq 0$. With a constant intensity on observation window $[0, T)$, wavelets with non compact support will always produce an infinite number of non-zero wavelet coefficients and unbiasedness of their estimators is not guaranteed. Furthermore, compactly supported wavelets whose support is only partially contained within $[0, T)$ will also admit non-zero wavelet coefficients. This is why we restrict ourselves to Haar wavelets in the definition of L-th level innovation. Extensions of L-th level innovation to other wavelets is considered in Appendix S4. We further comment that although defined according to a specific scale, L-th level innovation also has an inherent temporal component. The translation index of non-zero coefficients given by wavelets in W_L indicates the time localization of the corresponding innovation.

Remark 2. *For the Haar wavelet, there is the following equivalence:*

- *A point process N is level J homogeneous and possesses a level J innovation.*
- *A point process N is level $J + 1$ inhomogeneous.*

This equivalence is immediate from applying Definitions 3 and 4 to the identity $V_{J+1} = V_J \oplus W_J$.

3.4 Testing L^{th}-level innovation

We are now interested in testing for L-th level innovation based on Definition 4 using the null hypothesis H: "*A point process N possesses no L-th level innovation under a wavelet family (ϕ, ψ)*". To do so, we consider the vector of empirical wavelet coefficients corresponding to the wavelet basis for W_L, which under the null hypothesis will be zero mean. As for J-th level homogeneity, we define a likelihood ratio test for L^{th}-level innovation under the Poisson process

model and Haar wavelets. This test will again be a special case of a more general setting for multivariate Poisson random variables.

If a point process is level $J+1$ inhomogeneous, then such a test should take place for any given scale $L > J$ (as by Remark 2 we know there already exists an innovation at level J). Consider a subdivision S_{L+1} of $[0,T)$ defined as in Section 2.2.2. Let $\{N_m\}_{m=1}^M$ be a collection of M independent realizations of the same Poisson process N on $[0,T)$ with intensity function λ, and let $\mathbb{X}_N = \{\mathbf{X}_m\}_{m=1}^M$ be a collection of M independent random vectors $\mathbf{X}_m = (X_{m,i})_{i=0}^{2^{L+1}-1}$, where $X_{m,i} = N_m(s_i^{L+1})$ is the event count for process N_m in $s_i^{L+1} \in S_{L+1}$. With $\widehat{\beta}_{L,k} = \sum_{\tau_i \in \mathcal{E}} \psi_{L,k}(\tau_i)$, for the Haar wavelets $\widehat{\beta}_{L,k} = \frac{2^{L/2}}{\sqrt{T}}(X_{m,2k} - X_{m,2k+1}), 0 \leq k \leq 2^L - 1$. Each count $X_{m,i}$ is distributed as $\text{Pois}(\mu_i)$ where $\mu_i = \int_{s_i^{L+1}} \lambda(t)\text{d}t$. Therefore, the estimators $\widehat{\beta}_{L,k}$, $k = 0, ..., 2^L - 1$ are independent realizations of a scaled Skellam distribution (or Poisson difference distribution), each with parameters μ_{2k} and μ_{2k+1}. Since $\widehat{\beta}_{L,k}$ has mean $\frac{2^{L/2}}{\sqrt{T}}(\mu_{2k} - \mu_{2k+1})$, Definition 4 is then equivalent to the following property: "*There exists* $k \in \{0, ..., 2^L - 1\}$ *such that* $\widehat{\beta}_{L,k}$ *is Skellam distributed with parameters* $\mu_{2k} \neq \mu_{2k+1}$". We can therefore build a likelihood ratio test for testing the null hypothesis H: "$\mu_{2k} = \mu_{2k+1}$ *for all* $k = 0, ..., 2^L - 1$".

Since there does not exist an explicit expression for the MLE of the parameter $\theta_k = \mu_{2k} - \mu_{2k+1}$ given Skellam distributed random variables (instead having to be numerically approximated (Alzaid and Omair, 2010)), it is more appealing to design a likelihood ratio test based on the event counts themselves. This leads us to first consider a LRT for the general setting of testing pairwise equality of means of Poisson distributions, which will then be used for the specific setting of testing L-th level innovation.

3.4.1 LRT for pairwise equality of Poisson means

We define here a LRT for the pairwise equality of the means of a multivariate Poisson distribution. Let $\mathbb{X} = \{\mathbf{X}_m\}_{m=1}^M$ be a set of iid Poisson random vectors, each with independent components of form $\mathbf{X}_m = (X_{m,i})_{i=1}^{2P}$, $X_{m,i} \sim \text{Pois}(\mu_i)$. We consider testing the null hypothesis $H : \mu_{2i-1} = \mu_{2i} = \mu_i^{pair}$, $1 \leq i \leq P$, against the alternative hypothesis K that states H is not true. The LRT statistic is defined as

$$r = \frac{\sup_{\{\mu_i^{pair}\}_{i=1}^P,\ \sum \mu_i^{pair} > 0} \mathcal{L}(\mathbb{X}; \mu_1^{pair}, \mu_1^{pair}, ..., \mu_P^{pair}, \mu_P^{pair})}{\sup_{\{\mu_i\}_{i=1}^{2P},\ \sum \mu_i > 0} \mathcal{L}(\mathbb{X}; \mu_1, \mu_2, ..., \mu_{2P-1}, \mu_{2P})}, \tag{9}$$

where $\mathcal{L}(\mathbb{X}; \mu_1, ..., \mu_{2P})$ is the likelihood of the data $\mathbb{X}$ given parameter vector $(\mu_i)_{i=1}^{2P}$.

Proposition 4. *Let $R = -2\log(r)$, with r being the likelihood ratio statistic defined in (9). Then*

$$R = 2M\left[\sum_{i=1}^{P}\bar{\mu}_{2i-1}\log\left(\frac{\bar{\mu}_{2i-1}}{\bar{\mu}_i^{pair}}\right)+\sum_{i=1}^{P}\bar{\mu}_{2i}\log\left(\frac{\bar{\mu}_{2i}}{\bar{\mu}_i^{pair}}\right)\right],$$

where $\bar{\mu}_i = \frac{1}{M}\sum_{m=1}^{M}X_{m,i}$ and $\bar{\mu}_i^{pair} = \frac{1}{M}\sum_{m=1}^{M}\widehat{\mu}_{m,i}^{pair}$ where $\widehat{\mu}_{m,i}^{pair} = \frac{1}{2}(X_{m,2i-1}+X_{m,2i})$. Statistic $\bar{\mu}_i^{pair}$ is the maximum likelihood estimator (MLE) of μ_i^{pair} $(i=1,...,P)$ under the null hypothesis H and $\bar{\mu}_i$ is the MLE for μ_i $(i=1,...,2P)$ under the alternative hypothesis K.

The proof can be found in Supplementary Material S2.5. From Wilks' Theorem (Wilks, 1938), we immediately have that under the null hypothesis R is asymptotically χ^2 distributed with $d_K - d_H = P$ degrees of freedom for a large sample size M (under the usual regularity assumptions). However, this result is not guaranteed when the true parameter vector lies on the boundary of the parameter space. This was not the case for the test in Section 3.2.1 since we must have $\mu_c > 0$, although it happens in this model when $\mu_i^{pair} = 0$. Further discussion on this particular case can be found in Supplementary Material S1.1. We now assume that $\mu_i^{pair} \neq 0$ for all $1 \leq i \leq P$. Similarly to Theorem 1, we can state an extension of Wilks' theorem for this LRT.

Theorem 2. *Let $\mathbf{X}_1,..,\mathbf{X}_M$ $(M \geq 1)$ be independent and identically distributed $2P$ dimensional random vectors where each $\mathbf{X}_m = (X_{m,1},...,X_{m,2P})^T$ is constructed from independent components $X_{m,i} \sim \text{Pois}(\mu_i)$. Let $R = -2\log(r)$ where r is the likelihood ratio statistic defined in (9). Then the distribution of statistic R is invariant to simultaneous changes in parameters M and μ_i provided all products $\mu_i M$, $1 \leq i \leq 2P$ remain constant. Furthermore, if $\mu_{2i-1} = \mu_{2i} = \mu_i^{pair}$ and $\mu_i^{pair} \neq 0$, $1 \leq i \leq P$, then $R \xrightarrow{d} \chi^2_P$ as $\mu_i^{pair} M \to \infty$, $1 \leq i \leq P$.*

The proof of Theorem 2 follows an analogous argument to that of Theorem 1 (see Supplementary Material S2.7). We again prove that in the asymptotic analysis of the distribution of R, M and the mean intensity are indistinguishable from their product and thus the results are applicable for only one realization of the random vector $\mathbf{X}$.

3.4.2 LRT for L-th level innovation

We can now apply the test developed in Section 3.4.1 to the task of testing L-th level innovation. The LRT statistic for testing the null hypothesis H: "$\mu_{2k} = \mu_{2k+1}$ *for all* $k = 0,...,2^L-1$" is

$$r^L = \frac{\sup_{\{\mu_k^{pair}\}_{k=0}^{2^L-1},\ \sum\mu_k^{pair}>0}\mathcal{L}(\mathbb{X};\mu_0^{pair},\mu_0^{pair},...,\mu_{2^L-1}^{pair},\mu_{2^L-1}^{pair})}{\sup_{\{\mu_k\}_{k=0}^{2^{L+1}-1},\ \sum\mu_k>0}\mathcal{L}(\mathbb{X};\mu_0,...,\mu_{2^{L+1}-1})}.$$

From Proposition 4 we have:

$$R^L = -2\log(r^L) = 2M\left[\sum_{k=0}^{2^L-1}\bar{\mu}_{2k}\log\left(\frac{\bar{\mu}_{2k}}{\bar{\mu}_k^{pair}}\right) + \sum_{k=0}^{2^L-1}\bar{\mu}_{2k+1}\log\left(\frac{\bar{\mu}_{2k+1}}{\bar{\mu}_k^{pair}}\right)\right].$$

Again, we refer to Supplementary Material S1.1 in the situation where one or several parameters μ_k^{pair} are equal to zero. In all other cases, we have $d_K = 2^{L+1}$ and $d_H = 2^L$, giving R as asymptotically χ^2 distributed with 2^L degrees of freedom under the conditions of Theorem 2. We reject the absence of a level L innovation at significance level α if $R > c_\alpha$ where c_α, the critical value, is the upper $100(1-\alpha)\%$ point of the $\chi^2_{2^L}$ distribution.

3.4.3 Simulation study

Let us now consider the triangular intensity model from Section 3.2.3 where we now introduce an additive perturbation in the form of a sine function with period $T/2^\nu, \nu \geq V+3$, and magnitude $A\lambda_0$. Again, T is the length of the process and λ_0 is the mean value of the rate. Therefore this intensity model has expression

$$\lambda_{\text{sine}}(t) = \lambda_0\left(1 + (-1)^{i(t)}\xi\left(\theta(t) - \frac{1}{2}\right)\right) + A\lambda_0\sin\left(\frac{2^{\nu+1}\pi}{T}t\right).$$

Similarly to the previous model, the quantity $\mu_0^0 = \int_0^T \lambda_{\text{sine}}(t)dt = T\lambda_0$ does not depend on V, the process is level $V+1$ homogeneous and level $V+2$ inhomogeneous. The sinusoidal term does not influence the values of the wavelets coefficients up to resolution ν. Hence a Poisson process N whose intensity is λ_{sine} possesses no innovations from levels 0 to V, $V+1$ innovation is introduced by the triangular part and another source of innovation is introduced at level ν from the sinusoidal term. The power of the test is studied for $L \geq \nu$. An example plot is given in Figure 3.3. We set the significance level of our test at $\alpha = 0.05$, with $M = 1$ and $\lambda_0 \in [1000, 50000]$ as in the LRT for J-th level homogeneity. The empirical type 1 error and power plots from 10000 simulations are shown in Figure 3.3 for $L = 1$ (type 1 error in the absence of innovation) and $L = 3, 4$ and 5 (power in the presence of innovation). We are interested in exploring the effects of the parameter λ_0 on the empirical type 1 error and power of the LRT for the absence of L-th level innovation. Again the empirical type 1 error lies close to the 5% level as expected when the conditions of Theorem 2 are met. We also observe that the empirical power converges to 1 as the magnitude of the perturbation increases through the product $A\lambda_0$. Since the intensity model is still proportional to λ_0, this is also justified from Theorem 2 as the equivalent behavior is expected when λ_0 is fixed and M increases towards infinity. Furthermore, it is noticeable that for a fixed λ_0, the power decreases as we increase L. This can be explained because increasing L displaces the mass of the null distribution $\chi^2_{2^L}$ further to the right, making it harder for the test to distinguish between the null hypothesis and the true state of nature.

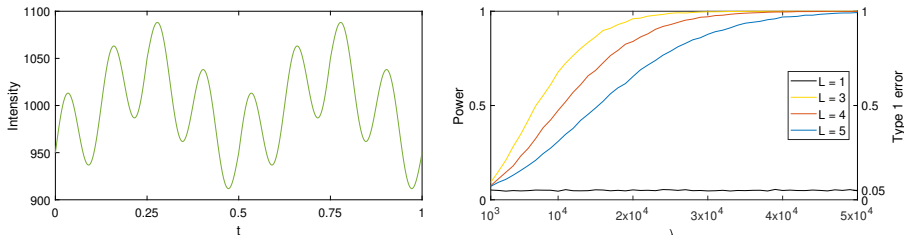


Figure 3.3: *Left: Triangular rate on* $[0,1]$ *with mean* $\lambda_0 = 1000$, $V = 1$, $\xi = 0.1$ *and an additive sine perturbation with* $\nu = 3$ *and magnitude* $A = 0.05$ *. Right: Empirical type 1 error* $(L = 1)$ *and power plots* $(L \in \{3,4,5\})$ *as a function of* λ_0 *with* $T = 1$, $V = 1$, $\xi = 0.1$ *and* $A = 0.05$. *See text in Section 3.4.3 for further details.*

4 Statistical Thresholding

As stated in Section 2.2, we can define a non-linear wavelet estimator of the intensity of a point process when a thresholding strategy is applied on the coefficient estimates. We initially define a general formulation for thresholding strategies in intensity estimation that we can adapt to different examples. To define a thresholding strategy, we need to choose a wavelet family for the estimation of the corresponding coefficients and a threshold operator that will be applied on the data. We consider a collection of compactly supported mother wavelets $\{\psi_{L,k}, k \in \mathcal{K}_L\}$, where $\mathcal{K}_L$ is the ordered finite subset of $\mathbb{Z}$ containing the translation indexes of the wavelets that are used as a basis for W_L, and further denote $K_L = |\mathcal{K}_L|$. For instance $\mathcal{K}_L = \{0, ..., 2^L - 1\}$ under the Haar basis if the intensity has support $[0,1)$ or $[0,T)$ with rescaled wavelets. Let $\{N_m, m = 1, ..., M\}$, $M \geq 1$, be a collection of independent realizations of the same point process N, we define $\widehat{\mathbf{B}}^L = (b_{m,i}) \in \mathbb{R}^{M \times K_L}$, where $b_{m,i} \equiv \widehat{\beta}^{(m)}_{L,k_i}$ is the estimator of the true wavelet coefficient β_{L,k_i} obtained from N_m.

We represent a thresholding operator $\mathcal{T} : \mathbb{R}^{M \times K_L} \to \mathbb{R}^{M \times K_L}$ with $\widehat{\mathbf{\Theta}}^L = \mathcal{T}(\widehat{\mathbf{B}}^L)$ being the output where each column of $\widehat{\mathbf{\Theta}}^L$ is the corresponding column of $\widehat{\mathbf{B}}^L$ if a thresholding criterion C is met, or a column of zeros if C is not met (see illustration in Figure 4.1). If the i-th column of $\widehat{\mathbf{B}}^L$ meets the criterion C and is therefore kept by the operator $\mathcal{T}$, then the estimator of β_{L,k_i} used in the final reconstruction of λ will be the sample mean $\frac{1}{M}\sum_{m=1}^{M} \widehat{\beta}^{(m)}_{L,k_i}$. A thresholding operator is applied between coarse and fine limits j_0 and J, respectively, resulting in a filtering of the information contained in the detail spaces $W_j, j_0 \leq j \leq J$. The effect of different choices for j_0 and J is explored in Appendices S3.1 and S3.2. Defining the $\mathbb{R}^{K_L}$ vector $\Psi_L(t) = (\psi_{L,k_1}(t), ..., \psi_{L,k_{K_L}}(t))^T$, where k_1 and k_{K_L} are respectively the first and last elements of the index set $\mathcal{K}_L$, and $\mathbf{1}_M = (1, ..., 1)^T$ the vector of ones of length M, the non-linear estimator can

$$\widehat{\mathbf{B}}^L = \begin{pmatrix} \widehat{\beta}_{L,1}^{(1)} & \widehat{\beta}_{L,2}^{(1)} & \widehat{\beta}_{L,3}^{(1)} & \widehat{\beta}_{L,4}^{(1)} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{L,1}^{(m)} & \widehat{\beta}_{L,2}^{(m)} & \widehat{\beta}_{L,3}^{(m)} & \widehat{\beta}_{L,4}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{L,1}^{(M)} & \widehat{\beta}_{L,2}^{(M)} & \widehat{\beta}_{L,3}^{(M)} & \widehat{\beta}_{L,4}^{(M)} \end{pmatrix} \Longrightarrow \widehat{\mathbf{\Theta}}^L = \mathcal{T}(\widehat{\mathbf{B}}^L) = \begin{pmatrix} \widehat{\beta}_{L,1}^{(1)} & 0 & \widehat{\beta}_{L,3}^{(1)} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{L,1}^{(m)} & 0 & \widehat{\beta}_{L,3}^{(m)} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{\beta}_{L,1}^{(M)} & 0 & \widehat{\beta}_{L,3}^{(M)} & 0 \end{pmatrix}$$

Figure 4.1: *Example output of a thresholding operator with* $\mathcal{K}_L = \{1, 2, 3, 4\}$.

be formulated as

$$\widehat{\lambda}_{\mathcal{T}}^J(t) = \frac{1}{M} \sum_{m=1}^{M} \sum_{k_i \in \mathcal{K}_{j_0}} \widehat{\alpha}_{j_0,k_i}^{(m)} \phi_{j_0,k_i}(t) + \frac{1}{M} \sum_{L=j_0}^{J} \mathbf{1}_M^T \widehat{\mathbf{\Theta}}^L \Psi_L(t) \ . \tag{10}$$

Similarly to the distinction made in Härdle et al. (1998) for density estimation, we define three procedures for thresholding. We are applying local thresholding if criterion C considers each column of $\widehat{\mathbf{B}}^L$ separately, global thresholding if C considers the entire matrix $\widehat{\mathbf{B}}^L$, and intermediate thresholding for other cases where C considers subsets of columns. The criteria C that we will propose here are based on variations of the previously defined L-th level innovation hypothesis test formulated in Section 3.4.2, and in doing so we assume that the conditions of Theorem 2 are always met for all $j_0 \leq L \leq J$. Our thresholding strategies hence take the form of multiple hypothesis testing procedures. It is consequently crucial to consider efficient ways of handling multiple hypothesis tests as ignoring this specificity could lead to a high number of truly zero coefficients to be kept in the reconstruction of λ.

When $M = 1$, a common setting, $\widehat{\mathbf{B}}^L$ and $\widehat{\mathbf{\Theta}}^L$ become row vectors with the i-th element of $\widehat{\mathbf{\Theta}}^L$ being $\widehat{\beta}_{L,k_i}(1 - \mathbb{1}_{[-\delta_{k_i}, \delta_{k_i}]}(\widehat{\beta}_{L,k_i}))$, where $\delta_{k_i} \geq 0$, $i = 1, ..., K_L$, are threshold levels that need to be chosen. (de Miranda and Morettin, 2011) propose $\delta_{k_i} = \omega\sqrt{\text{Var}(\widehat{\beta}_{L,k_i})}$, with ω typically equal to 3. This requires a crude estimator of the variance of the coefficient estimators. The authors notice an equivalence between this method and using $\widehat{\beta}_{L,k_i}$ as a test statistic for the null hypothesis $\beta_{L,k_i} = 0$. This employs Chebyshev's inequality and works on the assumption that $\widehat{\beta}_{L,k_i}$ is approximately Gaussian. This parallel is interesting enough for us to use this thresholding operator as a comparison point in our simulations.

4.1 Local thresholding with False Discovery Rate control

Under this thresholding procedure we apply a hypothesis test to each coefficient with the null hypothesis being that this coefficient is zero. In the case of Haar wavelets, the LRT for L-th level innovation defined in Section 3.4.2 can be

reduced to the case of a single coefficient without any change to its asymptotic properties. In particular, the reduced null hypothesis is now $H_0^{L,k}$:"$\beta_{L,k} = 0$" or equivalently "$\mu_{2k} = \mu_{2k+1}$", for any $k \in \{0, ..., 2^L - 1\}$ and $j_0 \leq L \leq J$.

Using a local thresholding operator with Haar wavelets requires a total of $Q = 2^{J+1} - 2^{j0}$ hypothesis tests for coarse and fine resolution scales j_0 and J, respectively. For this thresholding scheme, the criterion C considers individually the p-value of each test. A naive criterion C is that the coefficient is kept if the p-value for the corresponding test is lower than some fixed significance level α. However, in this case too few coefficients might be thresholded. The other approach that we explore here follows the statistical thresholding method of Abramovich and Benjamini (1995) which is based on the False Discovery Rate (FDR) defined in Benjamini and Hochberg (1995). Of the Q hypotheses being tested, we say that Q_0 are true null hypotheses and the total number of rejected hypotheses is R, of which F are falsely rejected. Note that Q_0 and F are unknown quantities. The FDR is the expectation of the ratio F/R, and is the quantity we look to control. Since the FDR approach to multiple testing produced lower mean squared errors compared to the universal hard threshold for certain types of signals in Abramovich and Benjamini (1995), it seems natural to carry it over to the Poisson intensity estimation model. This method positions itself between the naive approach where the error is only controlled at the very local level (coefficient-wise) and more constrained approaches like Bonferonni's correction where the error is instead simultaneously controlled among all tests (the family-wise error rate), with the latter being prone to power loss.

This procedure assumes independence of at least the Q_0 test statistics associated with the true null hypotheses. Under that setting the FDR is controlled by α, a global significance level. Since our Poisson intensity estimation model introduces dependence (between scales) among the test statistics, Benjamini and Yekutieli (2001) demonstrate that a conservative modification of α to $\alpha_Q = \alpha/(\sum_{i=1}^{Q} \frac{1}{i})$ allows us to extend the FDR control method for any joint distribution of the test statistics. The FDR is then bounded by $(Q_0/Q)\alpha$ which is lower than α. Now the thresholding procedure is as follows:

1. Determine the p-values $p_{L,k}$ of the LRT for the null hypothesis of each coefficient $H_0^{L,k}$:"$\beta_{L,k} = 0$", for all $j_0 \leq L \leq J$ and $k \in \mathcal{K}_L$ and sort them by increasing value to obtain the ordered indexed set $\mathcal{P} = \{p_1, \dots p_Q\}$, where Q is the total number of tests considered in the thresholding range. Note that Q does not depend on M.

2. For a given significance level α, find the largest index i that satisfies $p_i \leq (i/Q)\alpha_Q$ where $\alpha_Q = \alpha/(\sum_{i=1}^{Q} \frac{1}{i})$.

3. Criterion C states that the coefficients corresponding to the p-values smaller than or equal to p_i are kept.

4.2 Global thresholding with Holm-Bonferroni correction

The global thresholding strategy is based on the exact L-th level innovation test defined in Section 3.4.2. In this circumstance we test each level j, $j_0 \leq j \leq J$ with a single test. The total number of tests is now $Q = J - j_0 + 1$, significantly decreasing computational time when compared to the local thresholding method. Again, several approaches can be considered to control the multiplicity of errors arising from combining the results of multiple tests. One thing to notice is that swapping multiple univariate tests for a single multivariate test at each level L is already a way to address multiple hypothesis testing in this context. This choice reflects an emphasis on the detection of any significant information inside the detail space W_L regardless of its temporal location. This makes the thresholding easier to control statistically but may lead to an unnecessary number of coefficients kept in the end. Now since the number of tests here is linear with the maximum resolution J and thus limited in practice, the Holm-Bonferonni method Holm (1979), which is a uniformly more powerful method than Bonferonni correction, can be reasonably considered. Another interest here is that Holm-Bonferroni correction does not require independence of the test statistics. Now the procedure to determine the criterion C is the following:

1. Determine the p-value of the LRT for each null hypothesis H_0^L : " there is no L-th level innovation", $j_0 \leq L \leq J$, and sort them by increasing value to obtain the ordered indexed set $\mathcal{P} = \{p_1, \dots p_Q\}$, where Q is the total number of tests considered in the thresholding range. Again Q does not depend on M.

2. From a given significance level α, find the minimal index i that satisfies $p_i > \frac{\alpha}{Q+1-i}$. Denote this index i_m.

3. Accept the null hypotheses with p-values indexed from 1 to $i_m - 1$.

4. Criterion C states that if the null hypothesis H_0^L is accepted then $\widehat{\mathbf{\Theta}}^L = 0$, otherwise $\widehat{\mathbf{\Theta}}^L = \widehat{\mathbf{B}}^L$.

Using Holm-Bonferroni's correction, the familywise error rate of this global thresholding strategy, which is the probability or having at least one type 1 error for an individual test, is always less or equal to the given significance level α.

4.3 Intermediate thresholding based on recursive tests

The intermediate thresholding strategy uses the recursive testing approach proposed in Ogden and Parzen (1996). This method falls into the intermediate category since the number of coefficients tested together to determine Criterion C varies between 1 and $K_L = |\mathcal{K}_L|$ for each resolution level L. The procedure is the same at each level $j_0 \leq L \leq J$, and is as follows:

1. Test the null hypothesis H_0^L :"$\beta_{L,k} = 0$ for all $k \in \mathcal{K}_L$" using the LRT at significance level α.

2. If the null is rejected, find the index $i \in \mathcal{K}_L$ such that the sample mean $\frac{1}{M}\sum_m \widehat{\beta}_{L,i}^{(m)}$ has the largest absolute value. Remove the i-th component in the null hypothesis H_0^L to form a new null hypothesis $H_0^{L,-i}$.

3. Repeat steps 1 and 2 until the null is not rejected. Criterion C retains all the coefficients that have been removed from the original null hypothesis.

4.4 Simulation study

This study aims to compare the accuracy of different thresholding strategies by applying them on three Poisson process models on $[0, 1]$ with intensities that exhibit different behaviors and regularities. The chosen measures of accuracy are the mean root integrated squared error (MRISE) which is defined as

$$\mathrm{E}\left[\left(\int (\widehat{\lambda}^J(t) - \lambda(t))^2 \mathrm{d}t\right)^{1/2}\right],$$

and the mean integrated absolute error (MIAE) which is defined as

$$\mathrm{E}\left[\int \mid \widehat{\lambda}^J(t) - \lambda(t) \mid \mathrm{d}t\right].$$

We estimate the MRISE and the MIAE with

$$\widehat{MRISE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\left(\widehat{\lambda}_i^J(t_j) - \lambda(t_j)\right)^2\right)^{1/2}$$

and

$$\widehat{MIAE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} \mid \widehat{\lambda}_i^J(t_j) - \lambda(t_j) \mid\right).$$

In this study, we use $n = 10000$ repeat simulations and $t_j = (j-1)/m$ where $m = 1000$. The first two intensity models are based on the "*Blocks*" and "*Bumps*" test functions from Donoho and Johnstone (1994). The third function is a modification to that defined in Section 3.4.3. We will refer to this model as "*TriangleSine*" and it has expression

$$f_{\text{tsine}}(t) = \lambda_0\left(1 + (-1)^{i(t)}\xi\left(\theta(t) - \frac{1}{2}\right) + A\sin\left(\frac{2^{L+1}\pi}{T}t + \frac{1}{T}\right)\right),$$

where $i(t) \in \left\{0, \ldots, 2^{V+1} - 1\right\}$ is the index of the subinterval $s_{i(t)}^{V+1} = [\frac{i(t)}{2^{V+1}}T, \frac{i(t)+1}{2^{V+1}}T)$ in which t belongs, and $\theta(t)$ is defined by $t = (i(t) + \theta(t))T/2^{V+1}$.

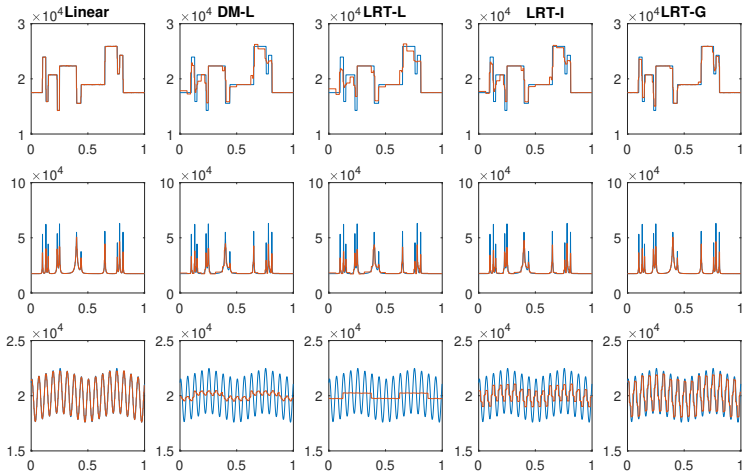


Figure 4.2: *Averaged reconstruction of the three intensity models "Blocks", "Bumps" and "TriangleSine", with* $A_0 = 10000, j_0 = 3, J = 7, M = 1$ *and significance level* $\alpha = 0.05$*. The true intensity is in blue and the reconstruction is in red.*

We set $T = 1$ and rescale these functions so that their integral on $[0, 1]$ are equal. Further, since the "*Blocks*" function can take negative values, we apply an upwards shift such that it is positive. The resulting intensities are

$$\lambda_{\text{blocks}}(t) = 1.75A_0 + 0.25A_0 \frac{f_{\text{blocks}}(t)}{\int_0^1 f_{\text{blocks}}} \qquad \lambda_{\text{bumps}}(t) = 1.75A_0 + 0.25A_0 \frac{f_{\text{bumps}}(t)}{\int_0^1 f_{\text{bumps}}} \qquad \lambda_{\text{tsine}}(t) = A_0 + A_0 \frac{f_{\text{tsine}}(t)}{\int_0^1 f_{\text{tsine}}}.$$

We are therefore ensuring that $\mathrm{E}\{N(1)\}$ is always equal to $2A_0$ for the three Poisson process models. The value of A_0 determines the highest resolution at which we can threshold the Haar wavelet coefficients. From the conditions of Theorem 2, the minimum value of the set $\left\{M\mu_i = M\int_{s_i^{J+1}} \lambda(t)\mathrm{d}t, \quad i = 0, ..., 2^{J+1} - 1\right\}$ should be high enough (for instance, greater than or equal to 50) for reliable likelihood ratio tests for L-th level innovation up to level J (and for smaller groups of wavelet coefficients in local and intermediate thresholding). Since we are demonstrating the presented methods for the $M = 1$ case this imposes that the minimum value of $\{\mu_i, i = 0, ..., 2^{J+1} - 1\}$ is greater than or equal to 50.

We now compare the MRISE and MIAE on these three intensity models for five thresholding strategies: statistical local, intermediate and global thresholding, as well as no thresholding (linear estimation) and the hard local thresholding of de Miranda and Morettin (2011). This study is restricted to continuous estimation of the intensity and $M = 1$ as we want to compare methods in a practical context. We included the linear estimation as it serves as a reference point and is also the $M = 1$ case for the methods presented in Reynaud-Bouret and

Table 1: *R-MRISE and R-MIAE values with* $A_0 = 10000, j_0 = 3, J = 7, M = 1$ *and significance level* $\alpha = 0.05$*. The number in bold indicates the best performing method.*

	Linear	**DM-L**	**LRT-L**	**LRT-I**	**LRT-G**
Blocks R-MRISE	1	0.6455	0.6937	**0.6402**	0.7701
Blocks R-MIAE	1	0.5156	0.5778	**0.5107**	0.7367
Bumps R-MRISE	1	1.0099	1.0538	**0.9659**	0.9996
Bumps R-MIAE	1	0.7632	0.7974	**0.7201**	1
TriangleSine R-MRISE	1	0.6887	0.6544	0.6747	**0.6000**
TriangleSine R-MIAE	1	0.7448	0.7224	0.7079	**0.6008**

Rivoirard (2010) and Bigot et al. (2013). We aim to study the influence of four parameters on this accuracy ranking: the starting resolution level j_0, the maximum resolution level J, the significance level α and the value of A_0. In Table 1 we provide the relative MRISE (R-MRISE) and relative MIAE (R-MIAE) values for one scenario where the estimated MRISE and MIAE for each thresholding strategy is divided by the value under absence of thresholding, which serves as a reference point. We refer to the method of de Miranda and Morettin (2011) as "DM-L" and our three statistical thresholding strategies as "LRT-L", "LRT-I" and "LRT-G" for the local, intermediate and global thresholding methods respectively. Intensity reconstructions averaged over 10000 simulations are shown in Figure 4.2 under the same setting and for all thresholding procedures as well. Bootstrapped 95% confidence intervals for the MRISE and the MIAE, plus further simulation studies can be found in Supplementary Material S3. The first conclusion in the setting of Table 1 is that we have statistical evidence that for all three intensity models at least one of LRT-I or LRT-G performs better than the linear and DM-L strategies. The statistical validity of this ranking relies on the absence of overlap between the 95% confidence intervals for the MRISE and MIAE of each method, as shown in Supplementary Material S3 Table 1. LRT-G performs better when innovations are well spread across time, whereas LRT-I leads in the case of abrupt changes. The same ranking is obtained when using the MIAE as an error measure, which gives consistency to the results. This was expected from the design of each strategy. For instance, the "*Blocks*" intensity has a sparse Haar wavelet decomposition with non-zero mother wavelets coefficients at high resolutions localized at the jumps. Therefore, this model favors LRT-L and LRT-I. Figure 4.2 shows the mean intensity estimate against the true intensity and therefore illustrates bias. We note as expected that the linear estimator is unbiased, although it has high variance which is accounted for in the MRISE. The improvement from the linear estimation is more significant for the "*Blocks*" and "*Bumps*" intensity models when the MIAE is used as an error measure. This is due to the MRISE giving more penalization to noisy estimators at higher resolutions than the MIAE.

5 Application to NetFlow Data

We apply the methodology presented in this paper to NetFlow data from a single router on the Imperial College London network. This data consists of 1.566×10^8 event times, each corresponding to the time a flow was sent or received by the router. The data was collected over a single 24 hour period that starts and ends at midnight. We therefore assume this to be a single realization ($M = 1$) of an underlying Poisson process that may or may not be homogeneous. In Section 4.4, to give validity to our approach, we proposed that the minimum event count across the half-support of a Haar wavelet at the highest level of resolution J should be at least 50. In the following data analysis we consider up to scale $J = 10$, at which the minimum event count is 5.9035×10^4. This puts us well within the setting where asymptotic distributions derived in this paper can be assumed and the power of the tests within the thresholding mechanism are high.

We start with testing level-1 homogeneity for which we obtain a p-value less than 10^{-300} indicating strong evidence to reject homogeneity. We also strongly reject the absence of level-L innovation for all levels between 1 and 10, with p-values again less than 10^{-300} in all cases. This indicates an underlying intensity function which is rapidly varying across even very small timescales (~ 1 minute). In Figure 5.1 we plot reconstructions of the intensity for different values of J using the LRT-I method with $j_0 = 3$ and $\alpha = 0.05$. From the simulation study in Supplementary Material S3, this method is preferred as it performed consistently well under this choice of parameters. Analyzing at $J = 4$ shows broad trends in network activity, including both human behavioural habits and general automated processes on the network. Specifically, between 03:00 and 07:00 the high activity through the router corresponds to automated system tasks, predominantly the backing up of servers. Human activity is then likely to be responsible for increasing activity from 09:00 until late afternoon. Netflow traffic then decreases from 18:00 until midnight as human activity on the network gradually decreases. The analysis of $J = 4$ and $J = 6$ reveal interesting spikes in activity at around 18:00 and 21:00. The spike at 18:00, for example, is likely to be a flurry of activity before people leave to go home. As we move up to scales 8 and then 10 we reveal regular, pronounced spiking in the activity on the network. Further analysis reveals these to be all of similar magnitudes and at regular 15 minute intervals. Routers are designed to manage their memory, which means that, at regular intervals, it will close some open flows, and start them again. Typical manufacturer choices for these intervals are 5, 10 or 15 minutes. This analysis would indicate 15 minutes for this particular router.

Our ability to be able to detect and characterize network behaviour in this multiscale fashion has potential usage in cyber-security applications where the characterization of "normal" network activity is key to being able to detect anomalous and potentially malicious activity.

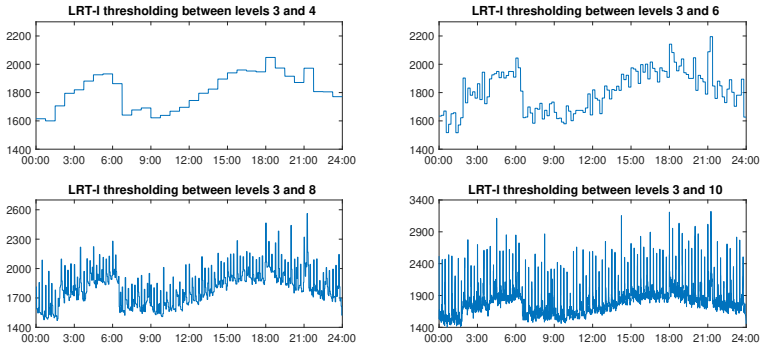


Figure 5.1: *Reconstruction of the NetFlow intensity with LRT-I, using parameters* $j_0 = 3, J \in \{4, 6, 8, 10\}, M = 1$ *and significance level* $\alpha = 0.05$.

6 Conclusion

The wavelet analysis of point processes in continuous time has been addressed through wavelet expansions of the first-order intensity. By defining a Haar wavelet multiresolution analysis on the point process, new multiscale properties, namely J-th level homogeneity and L-th level innovation, were introduced and tests for them formulated. Importantly, these tests can be applied when only a single realization of the process is observed. Tests for L-th level innovation formed the framework with which to perform thresholding of wavelet coefficients for intensity estimation.

The mean root integrated squared error and mean integrated absolute error of these methods were compared on simulated data for three different intensity models, revealing different accuracy rankings depending on the model. An important point here is that no thresholding method uniformly outperforms all others - although at least one of the statistical thresholding (LRT) methods outperforms the existing local hard thresholding method (DM-L) in all but one of the scenarios studied (see Supplementary Material S3). This seems reasonable and is consistent with the study of Antoniadis et al. (2001) for wavelet regression and Besbeas et al. (2004) for discrete time Poisson intensity estimation. The rule of thumb we offer is that LRT-G outperforms the other methods for intensity functions that exhibit smooth, large-scale changes in time. For intensity functions that exhibit abrupt, localized changes (i.e. possess a sparse wavelet representation), LRT-L and LRT-I strategies are to be preferred. It has been demonstrated that LRT-I thresholding when applied to NetFlow data exposes different behavior at different scales, and that this can be attributed with various human and automated activity. This illustrates the benefit and insight gained from a multiscale approach to analyzing point processes.

How to go about choosing the free-parameters α, j_0 and J in a data-driven way still needs to be addressed. The development of cross validation schemes in

the point process setting would make an interesting extension but falls outside the scope of this paper. Extensions of the presented theory and methodology can now be considered for the second-order intensity and multidimensional point processes.

Acknowledgements

The authors would like to thank Imperial College London ICT for providing the NetFlow data and Andy Thomas (Research Computing Manager, Department of Mathematics) for support. Youssef Taleb is funded by an EPSRC Doctoral Training Award.

References

Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes. *The Annals of Statistics 6*(4), 701–726.

Abramovich, F. and Y. Benjamini (1995). Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In *Wavelets and Statistics*, pp. 5–14. Springer.

Alzaid, A. A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society 33*(1), 17–45.

Antoniadis, A., J. Bigot, and T. Sapatinas (2001). Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study. *Journal of Statistical Software 6*(6), 1–83.

Bain, L. J., M. Engelhardt, and F. T. Wright (1985). Tests for an Increasing Trend in the Intensity of a Poisson Process: A Power Study. *Journal of the American Statistical Association 80*(390), 419–422.

Benjamini, Y. and Y. Hochberg (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics 29*(4), 1165–1188.

Besbeas, P., I. de Feis, and T. Sapatinas (2004). A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts. *International Statistical Review / Revue Internationale de Statistique 72*(2), 209–237.

Bigot, J., S. Gadat, T. Klein, and C. Marteau (2013). Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Electronic Journal of Statistics 7*(1), 881–931.

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

Brillinger, D. R. (1997). Some wavelet analyses of point process data. In *Proceedings of the Thirty-First Asilomar Conference on Signals, Systems and Computers*, pp. 1087–1091. IEEE.

Brown, L. D. and L. H. Zhao (2002). A new test for the Poisson distribution. *Sankhyā: The Indian Journal of Statistics, Series A 64*(A, Pt. 3), 1–29.

Cohen, E. A. K. (2014). Multi-wavelet coherence for point processes on the real line. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2649–2653. IEEE.

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

de Miranda, J. C. S. (2008). Probability Density Functions of the Empirical Wavelet Coefficients of Multidimensional Poisson Intensities. In *Functional and Operatorial Statistics*, pp. 231–236. Physica-Verlag HD.

de Miranda, J. C. S. and P. A. Morettin (2011). Estimation of the intensity of non-homogeneous point processes via wavelets. *Annals of the Institute of Statistical Mathematics 63*(6), 1221–1246.

Donoho, D. L. (1993). Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data. In *Proceedings of Symposia in Applied Mathematics*, pp. 173–205.

Donoho, D. L. and I. M. Johnstone (1994). Ideal spatial variation via wavelet shrinkage. *Biometrika 81*(3), 425–455.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard (1995). Wavelet Shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological) 57*(2), 301–369.

Feng, C., H. Wang, and X. M. Tu (2012). The asymptotic distribution of a likelihood ratio test statistic for the homogeneity of poisson distribution. *Sankhya A 74*(2), 263–268.

Fierro, R. and A. Tapia (2011). Testing Homogeneity for Poisson Processes. *Revista Colombiana de Estadística 34*(3), 421–432.

Fryzlewicz, P. and G. P. Nason (2004). A Haar-Fisz Algorithm for Poisson Intensity Estimation. *Journal of Computational and Graphical Statistics 13*(3), 621–638.

Härdle, W., G. Kerkyacharian, D. Picard, and A. Tsybakov (1998). *Wavelets, Approximation, and Statistical Applications*, Volume 129 of *Lecture Notes in Statistics*. Springer New York.

Helmers, R. and R. Zitikis (1999). On Estimation of Poisson Intensity Functions. *Annals of the Institute of Statistical Mathematics 51*(2), 265–280.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics 6*(2), 65–70.

Kolaczyk, E. D. (1999). Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statistica Sinica 9*(1), 119–135.

Kolaczyk, E. D. and D. D. Dixon (2000). Nonparametric Estimation of Intensity Maps Using Haar Wavelets and Poisson Noise Characteristics. *The Astrophysical Journal 534*(1), 490–505.

Meyer, Y. (1992). *Wavelets and Operators*, Volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press.

Ng, E. T. M. and R. J. Cook (1999). Adjusted Score Tests of Homogeneity for Poisson Processes. *Journal of the American Statistical Association 94*(445), 308–319.

Ogden, T. and E. Parzen (1996). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computational Statistics & Data Analysis 22*(1), 53–70.

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

Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics 11*(2), 453–466.

Rathbun, S. L. and N. Cressie (1994). Asymptotic Properties of Estimators for the Parameters of Spatial Inhomogeneous Poisson Point Processes. *Advances in Applied Probability 26*(1), 122–154.

Reynaud-Bouret, P. and V. Rivoirard (2010). Near optimal thresholding estimation of a Poisson intensity on the real line. *Electronic Journal of Statistics 4*(0), 172–238.

Taleb, Y. and E. A. K. Cohen (2016). A wavelet based likelihood ratio test for the homogeneity of poisson processes. In *2016 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 1–5.

Timmermann, K. E. and R. D. Nowak (1999). Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory 45*(3), 846–862.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge University Press.

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics 9*(1), 60–62.