RT-MAT 2006 - 01

Adaptive Maximum Probability Estimation of Multidimensional Poisson Processes Intensity Function

José Carlos Simon de Miranda

Março 2006

Esta é uma publicação preliminar ("preprint").

Adaptive Maximum Probability Estimation of Multidimensional Poisson Processes Intensity Function

José Carlos Simon de Miranda
Department of Mathematics
University of São Paulo
São Paulo, São Paulo, Brazil
email: simon@ime.usp.br

ABSTRACT

We propose a non parametric methodology of estimation of the intensity for Poisson point processes on $\mathbb{R}^m$. We assume the observation region, $\mathcal{O}$, is a bounded $\mathbb{R}^m$ interval. The space of positive functions formed by composition of $L^2(\mathcal{O})$-functions with the exponential is endowed with a probability induced from another one defined on the set of wavelet coefficients. This is a convenient space for the intensity to belong to and we choose as our estimate for the intensity the function that maximizes the posterior probability, given a trajectory of the Poisson process on $\mathcal{O}$, by means of determining the wavelet coefficients of its logarithm. This approach presents the desired feature of furnishing everywhere non negative amplitude-scale invariant estimates of the intensity that exhibit not only a minimization of the energy, relative to the wavelet basis, but also a maximization of the entropy of the process on $\mathcal{O}$ conditional to the realization. A novel adaptive thresholding procedure based on jointly testing hypothesis, on the wavelet coefficients, and adjusting the priors' locations is given. We define exponentially decaying invariance and, as an example of the general estimating procedure above, we specialyse to self affine and self similar probability prior Poisson processes.

KEY WORDS

Poisson process, intensity, self affine probability, wavelet, maximum posterior probability, entropy.

1 Introduction

In this work we address the estimation of Poisson point processes intensity function. We assume that the process occurs on $\mathbb{R}^m$ and that we are given a realization of the process on a finite $\mathbb{R}^m$-interval. From this information we want to find a good estimate of the intensity function on this region. This topic has already been studied for both the one and two dimensional cases using step functions with congruent, intervals and rectangles, step supports or, equivalently, Haar wavelets by Timmermann and Novak [1], where they justify the use of multiscale bayesian methods for modelling point processes, and by Novak and Kolaczyk [2]. Their analysis is developed using beta mixing distributions for the priors. Heikkinen and Arjas [3] also used piecewise constant functions to approximate the intensity for spatial, two dimensional, point processes. Random step functions on the plane were generated using Voronoi tessellations of random point patterns to build the support of the steps of the step functions. A smoothing technique is then applied to the stepwise intensity in order to obtain their final estimated intensity. For related works we refer to Novak and Kolaczyk [4], Müller and Vidakovic [5], Barber, Nason and Silverman [6], Kolaczyk and Novak [7], Lam and Wornell [8] and Winter, Maître, Cambou and Legrand [9].

The use of multidimensional compactly supported wavelets brings some advantages in estimating the intensity. If Haar wavelets are used and we fix the higher scale so as the supports of its wavelets are equal to the intervals and rectangles used in [1], [2], we recover their analysis. The possible difficulties in higher dimensions with Voronoi tessellations are not present in the wavelet approach. The use of smooth wavelets directly produces smooth intensity estimates. In this work we do not restrict the distributions of the priors and, in this general setting, we obtain the estimators of the wavelet coefficients of the log intensity of Poisson processes on $\mathbb{R}^m$ that correspond to the maximum posterior probability intensity. The only requirement on the prior probabilities of wavelet coefficients is that their means and standard deviations are finite and square summable over the wavelet coefficient indexing set. This is a natural condition if we assume the log intensity belongs to $L^2(\mathcal{O})$. As special cases we have self affine prior probability distributions self similar prior distributions and self similar location invariant prior distributions. We work out the example of self similar prior probability distributions in the exponential family. The special case of self similar gaussian priors shows itself to be particularly simple. If it is possible to choose, using translation of probability measures in our modelling procedure, the mean value of the prior distributions of only those wavelet coefficients for which there are no observed occurrences in their corresponding wavelet supports, we have proved that, without decreasing the maximum posterior probability, the estimate of the intensity that will not only minimize the log intensity energy but also maximize the Poisson process entropy on $\mathcal{O}$ is the one that corresponds to setting a mode of each of these distributions to zero and consequently zeroing these wavelet coefficients. The minimization of energy maximization of entropy criterion suggests an adaptive thresholding procedure based on testing the hypothesis of wavelet coefficients being equal to zero. Due to the form of our estimator and the possibility in the general case to substitute the constant function on $\mathcal{O}$ by the scale function and to choose the location of its prior, our estimate of the intensity is automatically amplitude-scale invariant in the sense defined by Figueiredo and Novak [10] in the context of image denoising.

This article is organized as follows: Section 2 is devoted to build some worthwhile notations that will simplify a lot the rest of our work. Section 3 brings some lemmas; in section 4 we present some definitions that will be used to describe the probability structure on the wavelet coefficients and consequently of the point process; the main results under the general prior probability setting are presented in section 5 where we assume fixed wavelet coefficient priors. Adaptivity of the priors is introduced in Section 6, where we show the optimization of energy and entropy are attained without decreasing maximum probability by zeroing some wavelet coefficients. It is enhanced in Section 7 where the adaptive threshold is defined. The self similar exponentially decaying probability prior Poisson process is studied in Section 8 where we use the novel concept of exponentially decaying invariance to determine the hyperparameters associated to the decay of the priors; the gaussian case is also studied in this section. Some concluding remarks are presented in Section 9.

Finally we observe that we could perform similar statistical modelling and estimation using sines and cosines, splines, or orthonormal polynomials as bases of function spaces instead of wavelets and the formalism would remain essentially the same.

2 Some Notations

Let N be a point process on $\mathbb{R}^m$, with unknown intensity p_N.

Let $\{\psi_{j,i}|i,j \in \mathbb{Z}\}$ be an orthonormal wavelet basis of $L^2(\mathbb{R})$ of the form $\psi_{j,i}(t) = 2^{j/2}\psi(2^jt - i)$ or $\psi_{j,i}(t) = 2^{j/2}\psi(2^j(t-t_1) + t_1 - iT)$ for some mother wavelet ψ obtained, if necessary by the composition of a standard wavelet with an affine transformation, such that its support is $[t_1, t_2]$ with $T = t_2 - t_1$. Let ϕ be the father wavelet corresponding to ψ.

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

It is extremely pleasant to adopt the following notation. Let ${}_d\mathbb{Z} = \{z \in \mathbb{Z} : z \geq d\}, d \in \mathbb{Z} \cup \{-\infty\}$ and define $Ze(\ell i) = \begin{cases} \mathbb{Z} \cup ({}_{\ell i}\mathbb{Z} \times \mathbb{Z}) \text{ if } \ell i \in \mathbb{Z}, \\ \mathbb{Z}^2 \text{ if } \ell i = -\infty. \end{cases}$

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

Thus, the wavelet expansions $f(t) = \sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}}\delta_{ji}\psi_{j,i}(t)$ and $f(t) = \sum_{k\in\mathbb{Z}}\gamma_k\phi_{\ell i,k}(t) + \sum_{i\in\mathbb{Z}}\sum_{j\in\ell_i\mathbb{Z}}\delta_{ji}\psi_{j,i}(t)$ will be simply written $f = \sum_{\eta\in Ze(\ell i)}\alpha_\eta\psi_\eta$, for α_η given by $\int_{-\infty}^{\infty} f\psi_\eta dt = \int_{\mathbb{R}}(\sum_\xi \alpha_\xi\psi_\xi)\psi_\eta dt = \sum_\xi\int_{\mathbb{R}}\alpha_\xi\psi_\xi\psi_\eta dt = \sum_\xi \alpha_\xi < \psi_\xi, \psi_\eta >= \alpha_\eta$.

Let for all n, $1 \le n \le m$, $\{\psi_{n,j,i}|i,j \in \mathbb{Z}\}$, $\psi_{n,j,i}(t) = 2^{j/2}\psi_n(2^j t - i)$ or $\psi_{n,j,i}(t) = 2^{j/2}\psi_n(2^j(t - a_n) + a_n - iT_n)$ and $\{\phi_{n,\ell i,k}, \psi_{n,j,i} : i,k \in \mathbb{Z}, j \ge \ell i_n, j, \ell i_n \in \mathbb{Z}\}$ be orthonormal wavelet bases of $L^2(\mathbb{R})$ as above where $\operatorname{supp}\psi_n = [a_n, b_n]$ and $T_n = b_n - a_n$. For easy of notation we write $(Ze(\ell i))_n = Ze(\ell i)_n$. These bases are simply written as $\{\psi_{n,\eta_n}|\eta_n \in (Ze(\ell i))_n\}$ and they are also orthonormal bases of $L^2[a_n, b_n]$, $1 \le n \le m$. Taking tensor products we form the orthonormal base $\{\psi_\eta|\psi_\eta = \otimes_{n=1}^m \psi_{n,\eta_n}, \eta = (\eta_1, \ldots, \eta_m) \in \prod_{n=1}^m (Ze(\ell i))_n\}$ of $L^2(\mathbb{R}^m)$ and also of $L^2(\prod_{n=1}^m [a_n, b_n])$. Denote $\prod_{n=1}^m (Ze(\ell i))_n$ by $Ze(\ell i)$. In this way if $f \in L^2(\mathbb{R}^m)$ we have $f = \sum_{\eta\in Ze(\ell i)} \alpha_\eta\psi_\eta$ with $\alpha_\eta = \int_{\mathbb{R}^m} f\psi_\eta d\ell$.

Our aim is to obtain the restriction of p_N to $\prod_{n=1}^m [a_n, b_n] = [a,b] = \mathcal{O}$ based on the points of a trajectory of the process that are contained in this $\mathbb{R}^m$ interval. Define $p = \begin{cases} p_N & \text{if } x \in [a,b], \\ 0 & \text{otherwise.} \end{cases}$
We will agree that if the trajectory presents no occurrences then $p = 0$ on $\mathcal{O}$. From now on we assume that p never vanishes and that $ln\ p \in L^2([a,b])$. In practice this is not a severe restriction since the estimated values tend to be negligible where $p = 0$. Therefore for the wavelet expansion of p we have $\ln p = \sum_\eta \alpha_\eta\psi_\eta$, (1) with $\alpha_\eta = \int_{\mathbb{R}^m}(\ln p)\psi_\eta d\ell = \int_{[a,b]}(\ln p)\psi_\eta d\ell$. (2)

The main purpose is to estimate p through the expansion (1) and for this we need to estimate the wavelet coefficients α_η given by (2).

We use $O_F = (0, \ldots, 0) \in Ze(\ell i)$, $O_M = ((\ell i, 0), \ldots, (\ell i, 0)) \in Ze(\ell i)$. We write for $\eta \in Ze(\ell i)$, $j(\eta) = \ell i$ for $\eta \in \mathbb{Z}$ and $j(\eta) = j$ if $\eta = (j,i)$ and also, if $\eta \in Ze(\ell i)$, $j(\eta) = (j(\eta_1), \ldots, j(\eta_m))$ and $|j(\eta)| = \sum_{\ell=1}^m j(\eta_i)$.

From now on in this work we use $\ell i = 0$.

3 Some Lemmas

Lemma 1: Let $\{u_i\}_{i\in I}$ be a class of random variables. A sufficient condition for the series $\sum_{i\in I}|u_i|^2$ to almost surely converge is that both $\sum_{i\in I}(\mathbb{E}u_i)^2$ and $\sum_{i\in I} Var(u_i)$ converge.

Proof: $\mathbb{E}\sum_{i\in I}|u_i|^2 = \sum_{i\in I}\mathbb{E}|u_i|^2 = \sum_{i\in I}[(\mathbb{E}u_i)^2 + \operatorname{Var}(u_i)]$.

In this way, if $\sum_{i\in I}(\mathbb{E}u_i)^2$ and $\sum_{i\in I}\operatorname{Var}(u_i)$ converge, then $\mathbb{E}\sum_{i\in I}|u_i|^2 < \infty$ so that $P\left(\sum_{i\in I}|u_i|^2 = \infty\right) = 0$. ∎

Lemma 2: Let A and B be such that $A \cup B = Ze(\ell i)$ and $A \cap B = \emptyset$. Then the subset of $L^2(\mathcal{O})$, $E = E((\alpha_\eta)_{\eta\in A}) = \{f \in L^2(\mathcal{O})| f = \sum_{\eta\in A}\alpha_\eta\psi_\eta + \sum_{\eta\in B}\beta_\eta\psi_\eta, \beta\eta \in \mathbb{R}\}$ is convex.

Proof: Let $f_i = \sum_{\eta\in A}\alpha_\eta\psi_\eta + \sum_{\eta\in B}\beta_\eta^{(i)}\psi_\eta$, $i = 1,2$ belong to E. Then, for all p, $0 \le p \le 1$,

$$\begin{aligned}(1-p)f_1 + pf_2 &= \sum_{\eta\in A}\alpha_\eta\psi_\eta + \sum_{\eta\in B}[(1-p)\beta_\eta^{(1)} \\ &\quad + p\beta_\eta^{(2)}]\psi_\eta \in E.\end{aligned}$$
∎

Lemma 3: $E((\alpha_\eta)_{\eta\in A}) \subset L^2(\mathcal{O})$ is closed, thus complete.

Proof: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in E such that $f_n \to f \in L^2(\mathcal{O})$ as $n \to \infty$. Write $f_n = \sum_{\eta\in A}\alpha_\eta\psi_\eta + \sum_{\eta\in B}\beta^{(n)}\psi_\eta$ and $f = \sum_{\eta\in A}\gamma_\eta\psi_\eta + \sum_{\eta\in B}\delta_\eta\psi_\eta$. In this way, for all $\eta \in A$, $|\alpha_\eta - \gamma_\eta| = |\langle f_n - f, \psi_\eta\rangle| \le \|f_n - f\| \to 0$, as $n \to \infty$, so that $\gamma_\eta = \alpha_\eta$ and $f \in E$. Completeness follows from that of $L^2(\mathcal{O})$. ∎

Lemma 4: Let E be a convex set of functions from $\mathcal{A} \in \mathcal{B}_{\mathbb{R}^m}$ to $\mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be such that for all x and y, $(g(x) + g(y))/2 > g((x+y)/2)$. Define the functional $F : E \to \mathbb{R}$ by $F(f) = \int_{\mathcal{A}} g(f)d\ell$, $\ell(\mathcal{A}) \neq 0$. Then, for all f_1 and f_2 in E we have $(F(f_1) + F(f_2))/2 > F((f_1 + f_2)/2)$ and, consequently, if it exists, the function f_* which minimizes F is unique.

Proof: $F(f_1) + F(f_2) = \int_{\mathcal{A}} g(f_1) + g(f_2) d\ell > \int_{\mathcal{A}} 2g\left(\frac{f_1+f_2}{2}\right) d\ell = 2F\left(\frac{f_1+f_2}{2}\right)$.

Suppose $\exists f_1, f_2,\ f_1 \neq f_2,\ \forall f,\ (F(f) \geq F(f_1) \wedge F(f) \geq F(f_2))$. Then $\overline{f} = \frac{f_1+f_2}{2}$ is such that $F(\overline{f}) < F(f_1) = F(f_2)$. ∎

Lemma 5: Let E, $\mathcal{A}$ and F be as in lemma's 4 hypothesis. Suppose that there exists $g'' : \mathbb{R} \to \mathbb{R}_+^*$ if there is f_* such that $\forall v : \mathcal{A} \to \mathbb{R}$, $\partial/\partial_\lambda \int_{\mathcal{A}} g(f_* + \lambda v) d\ell|_{\lambda=0} = 0$, then f_* is the unique function that minimizes F.

Proof: Since g'' is positive, $\forall x, y,\ \frac{g(x)+g(y)}{2} > g\left(\frac{x+y}{2}\right)$ and we are under lemma's 4 hypothesis.

Now it suffices to show that $\forall v,\ \frac{\partial^2}{\partial\lambda^2} F(f_* + \lambda v)|_{\lambda=0} > 0$.

$$\frac{\partial^2}{\partial\lambda^2} \int_{\mathcal{A}} g(f_* + \lambda v) d\ell = \int_{\mathcal{A}} g''(f_* + \lambda v).v^2 d\ell$$

so that

$$\frac{\partial^2}{\partial\lambda^2} F(f_* + \lambda v)|_{\lambda=0} = \int_{\mathcal{A}} g''(f_*) v^2 d\ell > 0$$

and the result follows. ∎

4 Some definitions

Let $r : L^2(\mathcal{O}) \to \ell^2(Ze(\ell i))$ be the wavelet representation of a function on $L^2(\mathcal{O})$, i. e., $r(f) = (\alpha_\eta)_{\eta \in Ze(\ell i)}$ where $f = \sum_{\eta \in Ze(\ell i)} \alpha_\eta \psi_\eta$.

For each $\eta \in Ze(\ell i)$, take a probability on $\mathbb{R}$, P_η, and form the product probability space $\left(\prod_{\eta \in Ze(\ell i)} \mathbb{R}, \bigotimes_{\eta \in Ze(\ell i)} \mathcal{B}_{\mathbb{R}}, \overline{P} = \bigotimes_{\eta \in Ze(\ell i)} P_\eta\right)$. Denote $\mu(\eta) = \int_{\mathbb{R}} x dP_\eta$ and $\mathrm{Var}(\eta) = \int_{\mathbb{R}} (x - \mu(\eta))^2 dP_\eta$ and suppose that $\sum_{\eta \in Ze(\ell i)} \mu(\eta)^2 + \sum_{\eta \in Ze(\ell i)} \mathrm{Var}(\eta) < \infty$. Let $i : \ell^2(Ze(\ell i)) \to \prod_{\eta \in Ze(\ell i)} \mathbb{R}$ be the inclusion. By lemma 1, $\bigotimes_{\eta \in Ze(\ell i)} P_\eta \left(\left\{(x_\eta)_{\eta \in Ze(\ell i)} | \sum_{\eta \in Ze(\ell i)} (x_\eta)^2 = \infty\right\}\right) = 0$ so that $\bigotimes_{\eta \in Ze(\ell i)} P_\eta(i(\ell^2(Ze(\ell i)))) = 1$ and $\overline{P}$ can be pulled back to $\ell^2(Ze(\ell i))$ and $\ell^2(\mathcal{O})$. We will refer to both these probabilities simply as π.

Definition 1: A probability $\bigotimes_{\eta \in Ze(\ell i)} P_\eta$ is summable if and only if $\sum_{\eta \in Ze(\ell i)} (\mu(\eta))^2 + \mathrm{Var}(\eta) < \infty$.

Definition 2: We say that $\overline{P} = \bigotimes_{\eta \in Ze(\ell i)} P_\eta$ is self affine on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ with affinity function $s : Ze(\ell i) \to \mathbb{R} \times \mathbb{R}_+^*$, $s(\eta) = (s_1(\eta), s_2(\eta))$, if and if for all $\eta \in Ze(\ell i) - \{O_F\}$ and all $A \in \mathcal{B}_{\mathbb{R}}$, $P_\eta(A) = P_{O_M}\left(s_1(O_M) + \frac{s_2(O_M)}{s_2(\eta)}(-s_1(\eta) + A)\right)$.

Note that if $s_1(O_M)$ and $s_2(O_M)$ are the mean and standard deviation of P_{O_M} respectively then $s_1(\eta)$ and $s_2(\eta)$ are the mean, $\mu(\eta)$, and standard deviation, $\sigma(\eta)$, of P_η, $\eta \neq O_F$.

Definition 3: A self affine probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ is admissible if and only if its affinity function satisfies:

- $\sum_{\eta \in Ze(\ell i)} (s_1(\eta)^2 + s_2(\eta)^2) < \infty$.
- $s_1(O_M)$ is the mean of P_{O_M} and $s_2(O_M)$ is the standard deviation of P_{O_M}.

Clearly an admissible self affine probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ induces a probability on the space of wavelet coefficients and on $L^2(\mathcal{O})$ which also will be called self affine probabilities.

Definition 4: A self affine probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ is location invariant if its affinity function is such that for all η and ξ, if $j(\eta) = (j(\eta_1), \ldots, j(\eta_n)) = j(\xi)$ then $s(\eta) = s(\xi)$. That is, a location invariant self affine probability assigns the same probability distribution to all copies of $\mathbb{R}$ that belong to the same scale.

Definition 5: The decay of a probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ is the sequence

$$(d_J)_{J \in S}, d_J = \sum_{\eta, j(\eta) = J} (\mu(\eta)^2 + \sigma(\eta)^2)$$

where S denotes the set of scales $S = j(Ze(\ell i))$

Of course a self affine probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ is admissible if and only if $(d_J)_{J \in S}$ is summable, $s_1(O_M) = \mu(O_M)$ and $s_2(O_M) = \sigma(O_M)$.

Definition 6: A probability on $\prod_{\eta \in Ze(\ell i)} \mathbb{R}$ is over exponentially decaying, or decays over exponentially, with the scales if there exist $c > 0$ and $k > 0$ such that for all $J \in S$, $d_J \leq ce^{-k|J|}$, where $|J| = \sum_{i=1}^m J_i$. We will say that k is the decay rate. This probability is called exponentially decaying if $d_J = ce^{-k|J|}$.

Clearly $\sum_{J \in S} d_J < \infty$ for these probabilities.

Definition 7: Let P be a probability measure on $\mathbb{R}$. We say that the probability P_v is the translation of P by $v \in \mathbb{R}$ when for all $A \in \mathcal{B}_{\mathbb{R}}$, $P_v(A) = P(-v + A)$.

Both discrete and continuous probabilities, as well as their combinations, are all denoted P_η so that when we write $y = \arg\max_{x\in\mathbb{R}} f(x)P_\eta(x)$, $f : \mathbb{R} \to \mathbb{R}_+$, we search the value of y first among the atoms $x_i \in \mathbb{R}$, $i \in I \subset \mathbb{N}$ of P_η and y will assume the value x_i for which $f(x_i)P_\eta(x_i) \geq f(x_j)P_\eta(x_j)$ for all $j \in I$. If P_η has no atoms or for all atoms $f(x_i) = 0$ then we are in a common maximization setting.

An observation $x = \{x_1, \ldots, x_N\}$; $x_i \in \mathcal{O} \subset \mathbb{R}^m$ generates a partition A, B of $\mathcal{Z}e(\ell i)$ by $\eta \in A$ if there is an occurrence x_i in the support of ψ_η and $\eta \in B$ otherwise.

5 Main Results

Theorem 1: Let N be a Poisson point process with intensity function p and $x = \{x_1, \ldots, x_N\}$ be an observation on $\mathcal{O}$. Suppose that the prior probability π on the space of wavelet coefficients is obtained from $P_0 \otimes (\otimes P_\eta)_{\eta\in\mathcal{Z}e(\ell i)-\{O_F\}}$, a summable probability on $\prod_{\eta\in\mathcal{Z}e(\ell i)} \mathbb{R}$ by translation of P_0, i. e., $P_{O_F} = P_{0,v}$. Then the maximum posterior probability estimate of p is given by $p^* = e^{\sum_\eta \alpha^*_\eta \psi_\eta}$ where

1) $\forall \eta \in B$, $\alpha^*_\eta = m_\eta$ a mode of P_η.

2) $\forall \eta \in A - \{O_F\} = A^-, \alpha^*_\eta =$
$\arg\max_{\alpha_\eta} \left(e^{\alpha_\eta \sum_{i=1}^N \psi_\eta(x_i)} P_\eta(\alpha_\eta) \right)$.

3) $\alpha^*_{O_F}$ is determined by

$\int_{\mathcal{O}} \psi_{O_F} e^{\sum_{\eta\in A^-\cup B} \alpha^*_\eta\psi_\eta + \alpha_{O_F}\psi_{O_F}} d\ell = \sum_{i=1}^N \psi_{O_F}(x_i)$.

The value of v that corresponds to the maximum posterior probability is given by $\alpha^*_{O_F} - v = m_0$ a mode of P_0.

Proof: $P(p|x) \propto P(x|p)\pi(p) \propto$
$e^{-\mu} \prod_{i=1}^N p(x_i) P_{0,v}(\alpha_{O_F}) \prod_{\eta\in A^-\cup B} P_\eta(\alpha_\eta)$ where
$\mu = \int_{\mathcal{O}} e^{\sum_{\eta\in\mathcal{Z}e(\ell i)} \alpha_\eta\psi_\eta} d\ell$,

$$P(p|x) \propto e^{\sum_{i=1}^N \sum_{\eta\in\mathcal{Z}e(\ell i)} \alpha_\eta\psi_\eta(x_i)} e^{-\mu}$$

$$\cdot \left(\prod_{\eta\in A^-\cup B} P_\eta(\alpha_\eta) \right) P_{0,v}(\alpha_{O_F})$$

$$= \prod_{\eta\in A-\{O_F\}} \left(e^{\alpha_\eta \sum_{i=1}^N \psi_\eta(x_i)} P_\eta(\alpha_\eta) \right)$$

$$. e^{\alpha_{O_F} \sum_{i=1}^N \psi_{O_F}(x_i)} P_0(\alpha_{O_F}) e^{-\mu} \prod_{\eta\in B} P_\eta(\alpha_\eta).$$

Now $P(p|x)$ is proportional to the product of three factors that can be maximized separately.

$\prod_{\eta\in B} P_\eta(\alpha_\eta)$ is maximized by taking α^*_η a mode of P_η;

$\prod_{\eta\in A^-} e^{\alpha_\eta \sum_{i=1}^N \psi_i(x_i)} P_\eta(\alpha_\eta)$ is maximized by choosing $\alpha^*_\eta = \arg\max_{\alpha_\eta\in\mathbb{R}} e^{\alpha_\eta \sum_{i=1}^N \psi_\eta(x_i)} P_\eta(\alpha_\eta)$ and $e^{\alpha_{O_F} \sum_{i=1}^N \psi_\eta(x_i)} P_0(\alpha_{O_F} - v) e^{-\mu}$ is maximized by taking $\alpha^*_{O_F}$ such that $e^{\alpha_{O_F} \sum_{i=1}^N \psi_\eta(x_i)} e^{-\mu}$ attains its maximum and then choosing v such that $\alpha_{O_F} - v$ is a mode of P_0.

Thus $\alpha^*_{O_F}$ maximizes $\alpha_{O_F} \sum_{i=1}^N \psi_{O_F}(x_i) - \mu$ so that

$$\sum_{i=1}^N \psi_{O_F}(x_i) - \mu' = \sum_{i=1}^N \psi_{O_F}(x_i)$$
$$- \int_{\mathcal{O}} \psi_{O_F} e^{\sum_{\eta\in A^-\cup B} \alpha^*_\eta\psi_\eta + \alpha_{O_F}\psi_{O_F}} d\ell = 0.$$

Clearly, the solution for the equation above is a point of maximum since $-\mu'' = -\int_{\mathcal{O}} \psi^2_{O_F} e^{\sum_{\eta\in\mathcal{Z}e(\ell i)} \alpha^*_\eta\psi_\eta} d\ell < 0$. ∎

Corollary 1: Under theorem's 1 hypothesis, if $P_\eta << \ell$ and $\frac{dP_\eta}{d\ell} = f_\eta$ is differentiable then α^*_η is within the solutions of

1) $f'_\eta(\alpha_\eta) = 0$ if $\eta \in B$.

2) $\frac{f'_\eta(\alpha_\eta)}{f_\eta(\alpha_\eta)} = -\sum_{i=1}^N \psi_\eta(x_i)$ if $\eta \in A - \{O_F\}$.

In case there exists f''_η, we add $f''_\eta(\alpha_\eta) \leq 0$ to item 1 and $f''_\eta(\alpha_\eta)f_\eta(\alpha_\eta) - (f'_\eta(\alpha_\eta))^2 \leq 0$ to item 2.

Proof: Immediate. ∎

Corollary 2: If ψ_{O_F} is constant on $\mathcal{O}$ then $\mu = \int_{\mathcal{O}} p^* d\ell = N$.

Proof: $\mu' = \psi_{O_F}(x) \int_{\mathcal{O}} p^* d\ell = \sum_{i=1}^N \psi_{O_F}(x_i)$. ∎

Observe that, in this case, amplitude scale invariance with respect to the intensity values is equivalent to summing a random variable, $v \in \mathbb{R}$, with improper prior $\pi(v) \propto 1$ to the wavelet coefficient of ψ_{O_F}.

Corollary 3: Let π be a self affine prior on the wavelet coefficients with affinity function $s(\eta) = (\mu(\eta), \sigma(\eta))$ and suppose $\frac{dP_{O_M}}{d\ell} = f_{O_M}$ is twice differentiable. Denote f its re-scaled density with mean 0 and variance 1, i. e., $f(\alpha) = \frac{1}{\sigma(O_M)} f_{O_M}\left(\frac{\alpha - \mu(O_M)}{\sigma(O_M)}\right)$. Then we have

1) $\forall \eta \in B$, α^*_η is of the form $\sigma(\eta)\alpha + \mu(\eta)$ where α is within the solutions of $f'(\alpha) = 0$ and $f''(\alpha) \leq 0$.
2) $\forall \eta \in A^-$, α^*_η is of the form $\sigma(\eta)\alpha + \mu(\eta)$ where α is within the solutions of $\frac{f'(\alpha)}{f(\alpha)} = -\sigma(\eta)\sum_{i=1}^N \psi_\eta(x_i)$ and $f''(\alpha)f(\alpha) - (f'(\alpha))^2 \leq 0$.

Proof: Straightforward.

Example: The gaussian self affine case.

If $f(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-\alpha^2/2}$ then $f'(0) = 0$, $f''(0) < 0$, $f'(\alpha) = -\alpha f(\alpha)$ and $f''(\alpha) = -f(\alpha) + \alpha^2 f(\alpha)$ so that for all α, $f''(\alpha)f(\alpha) - (f'(\alpha))^2 = -(f(\alpha))^2 < 0$.

In this way, by corollary 3, the gaussian case reduces to the simple formulas:

$$\begin{aligned} \forall \eta \in B,\ \alpha^*_\eta &= \mu(\eta), \\ \forall \eta \in A^-,\ \alpha^*_\eta &= \sigma(\eta)^2 \sum_{i=1}^N \psi_\eta(x_i) + \mu(\eta) \end{aligned}$$

and $\alpha^*_{O_F}$ is obtained from the equation

$$\int_O \psi_{O_F} e^{\sum_{A^- \cup B} \alpha^*_\eta \psi_\eta + \alpha_{O_F}\psi_{O_F}} d\ell = \sum_{i=1}^N \psi_{O_F}(x_i).$$

6 Entropy and Energy

Now we are interested in improving the procedure of estimating the intensity by introducing some adaptiveness to our modelling.

By definition, the wavelets ψ_η with $\eta \in B$ have no occurrences, from the realization we have observed, in their supports. It seems to be quite sensible to set their coefficients equal to zero.

Since $\ln p = \sum_{\eta \in Ze(\ell i)} \alpha_\eta \psi_\eta$, the energy of $\ln p$ is $\sum_{\eta \in A} \alpha^2_\eta + \sum_{\eta \in B} \alpha^2_\eta$.

Suppose we have freedom to choose the mean value of P_η for $\eta \in B$ by translation of probability measure. Then we can minimize the energy of $\ln p$ in the detail scales while maximizing the posterior probability by taking $\hat{p} = e^{\sum_{\eta \in A} \alpha^*_\eta \psi_\eta}$ as the estimate of the intensity, where α^*_η for $\eta \in A$ are determined as in theorem 1, items 2 and 3. The coefficients α^*_η for $\eta \in B$ may be set to zero since we can substitute P_{η, m_η}, for P_η, by changing the mean value of P_η, so that one of its modes is now zero. Observe that the coefficients α^*_η, $\eta \in A - \{O_F\}$ remain unchanged but $\alpha^*_{O_F}$ is changed. Note that the maximization of the posterior probability is maintained. The following theorem shows that $\hat{p}$ also maximizes the entropy of N on $\mathcal{O}$, $(\int_{\mathcal{O}} p \ln p d\ell)$, given the observation, i. e., letting α^*_η unchanged for $\eta \in A - \{O_F\}$.

Theorem 2: Under theorem 1 hypothesis, if we have freedom to choose the mean value of P_η by translation of probability measure then $\hat{p} = e^{\sum_{\eta \in A} \alpha^*_\eta \psi_\eta}$ is the maximum posterior probability intensity that maximizes the entropy of N on $\mathcal{O}$.

Proof: For all $v \in \langle\{\psi_\eta | \eta \in B\}\rangle$ we have

$$\begin{aligned} &\frac{\partial}{\partial \lambda} \int_O (\hat{p} + \lambda v) \ln(\hat{p} + \lambda v) d\ell|_{\lambda=0} \\ &= \frac{\partial}{\partial \lambda} \int_O (\hat{p} + \lambda \sum_{\eta \in B} \gamma_\eta \psi_\eta) \ln(\hat{p} + \lambda \sum_{\eta \in B} \gamma_\eta \psi_\eta) d\ell|_{\lambda=0} \\ &= \int_O (\sum_{\eta \in B} \gamma_\eta \psi_\eta)(1 + \ln \hat{p}) d\ell \\ &= \sum_{\eta \in B} \gamma_\eta \int \psi_\eta (1 + \sum_{\xi \in A} \alpha^*_\xi \psi_\xi) d\ell \\ &= \sum_{\eta \in B} \gamma_\eta \left(\int \psi_\eta d\ell + \sum_{\xi \in A} \alpha^*_\xi \langle \psi_\eta, \psi_\eta \rangle \right) = 0. \end{aligned}$$

And the result follows from lemmas 2, 4, and 5. ∎

7 Adaptive Thresholding

In this section we develop a thresholding procedure for the wavelet coefficients. In the previous section we have seen that setting coefficients to zero decreases the energy of the log intensity and maximizes the entropy of the process on $\mathcal{O}$ given the observed trajectory. Following this reasoning and noting that the values of the wavelet coefficients we have are based on only one trajectory of the process, it is reasonable to test the hypothesis of these coefficients being equal to zero. We have supposed that we have freedom to choose the mean value of the prior probabilities by translation so as to set modes equal to zero when there are no occurrences on the support of the corresponding wavelets. Now we will also suppose we can do so for those wavelet coefficients that are not significantly different from zero. Due to reasons that deppend solely on the trajectory, which bears a random character, our procedure sets a coefficient equal to zero if either $\eta \in B$ or $\eta \in A - \{O_F\}$ and $\sum_{i=1}^{N} \psi_\eta(x_i) = 0$. This resumes to the condition $\sum_{i=1}^{N} \psi_\eta(x_i) = 0$. In this way we will test the hypothesis $\alpha_\eta = 0$ using the sum $\sum_{i=1}^{N} \psi_\eta(x_i)$ to build a statistic. If $\alpha_\eta = 0$ then $p = e^{\sum_{\xi \neq \eta} \alpha_\xi \psi_\xi}$ and $v_\eta = \sum_{i=1}^{N} \psi_\eta(x_i)$ is distributed as the random sum $\omega \to \sum_{i=1}^{N(\omega)} \psi_\eta(x_i(\omega)) \sim \sum_{i=1}^{N'} \psi_\eta(x_i)$ where the independent random variables N', X_i have distributions $N' \sim$ Poisson $\left(\int_{\mathrm{supp}\psi_\eta} p d\ell\right)$ and $X_i \sim \frac{\exp(\sum_{\xi \neq \eta} \alpha_\xi \psi_\xi(x))}{\int_{\mathrm{supp}\psi_\eta} \exp(\sum_{\xi \neq \eta} \alpha_\xi \psi_\xi(x)) dx}$. If we know that on $\mathrm{supp}\psi_\eta$ there have been $N_\eta \neq 0$ occurrences then the conditional statistic $v_\eta = \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} \psi_\eta(x_i)$ will have distribution $(v_\eta | N_\eta) \sim \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} \psi_\eta(X_i)$, where $\{X_i\}_{1 \le i \le N_\eta}$ are i.i.d. with distribution as above. Due to independence, $\frac{1}{N_\eta} \sum_{i=1}^{N_\eta} \psi_\eta(X_i) \xrightarrow{\mathcal{D}} N\left(E\psi_\eta(X), \frac{\mathrm{Var}\,\psi_\eta(X)}{N_\eta}\right)$ as $N \to \infty$. Using $\exp(\sum_{\xi \neq \eta} \alpha_\xi^* \psi_\xi)$ as an estimate for $\exp(\sum_{\xi \neq \eta} \alpha_\xi \psi_\eta)$ we access the distribution of v_η via simulation. Observe that the distribution of x_i depends on an infinite series and we have to limit the number of its terms. This can be done in a variety of ways which may include information from the practical situation under study as the precision of the occurrences location for example, may take advantage of the minimum precision requirement for the estimated intensity or, if it is too expensive or time consuming to obtain coefficients, penalty functions on the number of scales may be used. We propose here as a general guideline, for simulation and general applications where we doont have the requirements or clues as above, one of the following two methods to limit the number of wavelets: I) limit the higher scale in such a way that if this scale is increased then there are at most one occurrance point in each support of the new wavelets that belong to this new higher scale. II) limit the scale as above and use only those wavelets with at least two points from the observed trajectory in their support. Let us call the finite sets of wavelets corresponding to the choices above as I and II respectively. Now, using $\exp(\sum_{\xi \neq \eta} \alpha_\xi^* \psi_\xi)$ where $\eta \in I$ or $\eta \in II$ as an estimate for $\exp(\sum_{\xi \neq \eta} \alpha_\xi \psi_\xi)$ we access the distribution of v_η via simulation and then we test the hypothesis $\alpha_\eta = 0$ for some prescribed significance level, with the statistic $\hat{v}_\eta = \frac{1}{N_\eta} \sum_{i=1}^{N_\eta} \psi_\eta(x_i)$ we obtain from the realization we have. After performing this tests, we are left with a set of non zero coefficients which we will call $T_I(1) = \{\alpha_{\eta,1}^*\}_{\eta \in I_1}$ or $T_{II}(1) = \{\alpha_{\eta,1}^*\}_{\eta \in II_1}$. Our estimate of the intensity may have changed and we may perform again the hypothesis testing procedure on the coefficients that belong to $T_I(1)$ or $T_{II}(1)$ so as to generate new sets $T_I(2)$ or $T_{II}(2)$ and so on till we stabilize the sets $T_I(n)$ or $T_{II}(m)$, that is, $T_I(n) = T_I(n+1)$ and $T_{II}(m) = T_{II}(m+1)$. The final thresholded estimator of the intensity will then be $p^T = e^{\sum \alpha_\eta \psi_\eta}$ where $\alpha_\eta \in T_I(m)$ or $\alpha_\eta \in T_{II}(m)$.

8 Self Similar Exponentially Decaying Prior Poisson Process Estimation.

In this section we assume that we have an admissible self affine exponentially decaying prior for the wavelet coefficients of our Poisson process. In this way we have the decay $\sum_{j(\eta)=J} (\sigma(\eta)^2 + \mu(\eta)^2) = ce^{-k|J|}$ which is a function of two hyper parameters.

Definition: A class $\mathcal{C} = \{\overline{P}_{(c,k)}\}_{(c,k) \in \mathbb{R}_+^2}$ of exponentially decaying probabilities $\overline{P}_{(c,k)}$ on $\prod_{\eta \in \mathcal{Z}e(\ell i)} \mathbb{R}$

is called exponentially decay invariant when we associate an improper prior density $\pi(c,k)$ to $\mathbb{R}^2_+$ such that $\pi(c,k) \propto \frac{1}{c}$.

This is a natural invariance requirement if we want our intensity estimates are amplitude scale invariant with respect to the scales of the copies of $\mathbb{R}$ where the wavelet coefficients belong to.

If we restrict to self similar location invariant priors we have $\sum_{j(\eta)=J}\sigma(\eta)^2 = 2^{|J|}\sigma(\eta)^2$ from which $\sigma(\eta) = \sqrt{c}.e^{-(\frac{k+\ln 2}{2})|J|}$ and given c and k we can obtain $\tilde{p}(c,k)$ the maximum probability estimate of p given c and k as in sections 5, 6 or 7 ($\tilde{p}$ will be p^*, $\hat{p}$ or p^T).

An exponentially invariant estimate for the intensity will be given by the mean maximum probability estimator

$$p = \lim_{n\to\infty} \frac{\left(\int_{\mathcal{K}_n} \tilde{p}(c,k).\frac{1}{c}dcdk\right)}{\int_{\mathcal{K}_n} \frac{1}{c}dcdk}$$

for a suitable choice of compact sets $\{\mathcal{K}_n\}_{n\in\mathbb{N}}$, $\mathcal{K}_n \subset \mathcal{K}_{n+1}$, $\cup_{n\in\mathbb{N}}\mathcal{K}_n = \mathbb{R}^2_+$.

9 Conclusion

We have presented a non parametric wavelet based maximum probability methodology of estimation of Poisson intensities. The procedure has the advantages of easily handling multidimensional processes and is developed for a very general setting of arbitrary priors on the wavelet coefficients. Even completely different from one another priors can be chosen for the wavelet coefficients in the same practical situation. This brings both flexibility in modelling and robustness in analysis. The estimator of the intensity have the desired properties of being nowhere negative and amplitude scale invariant while presenting minimization of energy and maximization of entropy features that are based on adapting priors conditionally to the observation of the process. A novel adaptive thresholding procedure, with a possible automatic choice of scales feature, produces a thresholded estimator using as "degrees of freedom", translations of wavelet coefficient priors which are allowed of forbidden by a built in adaptive hypothesis testing. The concept of exponentially decay invariance is introduced and applied to the self similar priors Poisson processes furnishing estimators that, in addition to the general case properties, are also exponentially decay invariant. Like other adaptive procedures, our methodology may be computationally demanding, although for special cases, like that of gaussian priors with heavy amount of occurrences on $\mathcal{O}$ relative to a small practical situation limited number of wavelet scale, this demand will be reduced. In case the observation region is not an $\mathbb{R}^m$-interval we refer to the solutions found in [11]. Our approach is a general purpose Poisson process estimation so finding plenty of practical applications.

Acknowledgements

The author thanks Our Lord and Saviour Jesus Christ.
This work was partially supported by FAPESP grant 03/10105-2.

References

[1] K. E. Timmermann, R. D. Novak, Multiscale Bayesian Estimation of Poisson Intensities, *IEEE*, 1998, 85-90.

[2] R. D. Novak, E. D. Kolaczyk, A Multiscale MAP Estimation Method for Poisson Inverse Problems, *IEEE*, 1998, 1682-1686.

[3] J. Heikkinen, E. Arjas, Non-Parametric Bayesian Estimation of a Spatial Poisson Intensity, *Scand J Statist*, 25 1998, 435-450.

[4] R. D. Novak, E. D. Kolaczyk, Multiscale Maximum Penalized Likelihood Estimators, *IEEE*, 2002, 156.

[5] P. Müller, B. Vidakovic, Bayesian Inference with Wavelets: Density Estimation *Jornal of computational and Graphical Statistics*, volume7, number4, 1998, 456-468.

[6] S. Barber, G. P. Nason, B. W. Silverman, Poste-

rior Probability Intervals for Wavelet Thresholding, *J. R. Statist. Soc. B*, 64, part2, 2002, 189-205.
[7] E. D. Kolaczyk, R. D. Novak, Multiscale Likelihood Analysis and Complexity Penalized Estimation *The Annals of Statistics*, Vol. 32, No. 2, 2004, 500-527.
[8] W. M. Lam, G. W. Wornell, Multiscale Representation and Estimation of Fractal Point Processes *IEEE Transactions on Signal Processing*, Vol. 43, No. 11, 1995, 2606-2617.
[9] A. Winter, H. Maître, N. Cambou, E. Legrand, Object Detection Using A Multiscale Probability Model *IEEE*, 1996, 269-272.
[10] M. A. T. Figueiredo, R. O. Novak, Wavelet-Based Image Estimation: An Empirical Bayes Approach Using Jeffrey's Noninformative Prior. *IEEE Transaction on Image Processing*. Vol. 10, No. 9. September, 2001, 1322–1331.
[11] J.C.S.de Miranda, P.A.Morettin, Estimation of the Density of Point Processes on $\mathbb{R}^m$ via Wavelets ,*Technical Report - Department of Mathematics - IME-USP* , No. 09, June, 2005.

TRABALHOS DO DEPARTAMENTO DE MATEMÁTICA

TÍTULOS PUBLICADOS

2005-01 PEREIRA, A. L. and PEREIRA, M. C. Generic simplicity for the solutions of a nonlinear plate equation. 22p.

2005-02 ARAGONA, J. Colombeau generalized functions on quasi-regular sets. 25p.

2005-03 PEREIRA. A. L. Global attractor and nonhomogeneous equilibria for a non local evolution equation in an unbounded domain. 26p.

2005-04 SHESTAKOV, I. and ZHUKAVETS, N. Malcev Poisson superalgebra of free Malcev superalgebra on one odd generator. 19p.

2005-05 SHESTAKOV, I. and ZHUKAVETS, N. Speciality of Malcev superalgebras on one odd generator. 15p.

2005-06 BILLIG, Y., FUTORNY, V. and MOLEV, A. Verma modules for Yangians. 18p.

2005-07 GORODSKI, C. Jacobi fields on normal homogeneous Riemannian manifolds. 6p.

2005-08 GRISHKOV, A. Notes on simple Lie algebras over a field of characteristic 2. 7p.

2005-09 MIRANDA, J. C. S. and MORETTIN, P. A. Estimation of the density of point processes on $\mathbb{R}^m$ via wavelets. 15p.

2005-10 KASHUBA, I., OVSIENKO, S. and SHESTAKOV, I. Jordan algebras versus associative algebras. 35p.

2005-11 IWAKI, E., JESPERS, E. and JURIAANS, S. O. The hypercentre of the unit group of an integral group ring. 11p.

2005-12 FUTORNY, V. and OVSIENKO, S. Theory of Galois algebras. 54p.

2005-13 MIRANDA, J. C. S. Characterization of the intensity for a class of point processes. 16p.

2005-14 MIRANDA, J. C. S. Neglecting higher order infinitesimals and the study of some measures. 16p.

2005-15 MIRANDA, J. C. S. Some wide classes of point processes interrelationships and three working properties. 15p.

2005-16 MIRANDA, J. C. S. Sure inference analysis. 27p.

2005-17 GRISHKOV, A. N. and ZAVARNITSINE, A. V. Sylow's theorem for Moufang loops I. 14p.

2005-18 LOURENÇO, M. L. and TOCHA, N. N. Zeros of complex homogeneous polynomials. 7p.

2005-19 ALEXANDRINO, M. M. Proofs of conjectures about singular riemannian foliations. 19p.

2005-20 MELO, S. T. and SILVA, C. C. K-Theory of pseudodifferential operators with semi-periodic symbols. 13p

2005-21 COX, B. L. and FUTORNY, V. Structure of intermediate Wakimoto modules. 20p.

2005-22 BAHTURIN, Y. A., SHESTAKOV, I. P. and ZAICEV, M. V. Group Gradings on Jordan Algebras $M_n^{(+)}$.30p.

2005-23 BILIOTTI, L. and GORODSKI, C. Polar actions on compact irreducible symmetric spaces are taut. 7p.

2005-24 LOURENÇO, M. L. and PELLEGRINI, L. Interpolation by analytic functions on preduals of Lorentz sequence spaces. 9p.

2005-25 LOURENÇO, M. L. and ACOSTA, M. D. Šilov boundary for holomorphic functions on some classical Banach spaces.13p.

2006-01 MIRANDA, J. C. S. Adaptive maximum probability estimation of multidimensional Poisson processes intensity function. 9p.

Nota: Os títulos publicados nos Relatórios Técnicos dos anos de 1980 a 2005 estão à disposição no Departamento de Matemática do IME-USP.

Cidade Universitária "Armando de Salles Oliveira"
Rua do Matão, 1010 - Cidade Universitária
Caixa Postal 66281 - CEP 05315-970