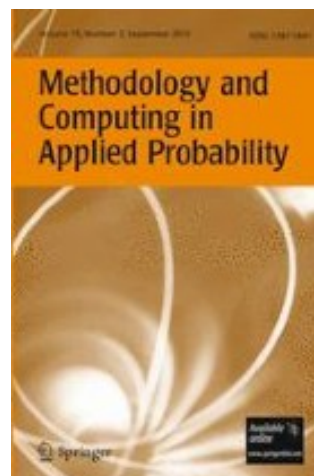


Functional Estimation of the Random Rate of a Cox Process

- Paula R. Bouzas; Ana M. Aguilera; Nuria Ruiz-Fuentes
- Functional Estimation of the Random Rate of a Cox Process
- *Methodology and Computing in Applied Probability* volume 14, pages 57–69 (2012)
- DOI: <https://doi.org/10.1007/s11009-010-9173-z>



Functional Estimation of the Random Rate of a Cox Process

Paula R. Bouzas · Ana M. Aguilera ·
Nuria Ruiz-Fuentes

Received: 16 November 2009 / Revised: 4 February 2010 /
Accepted: 9 March 2010 / Published online: 25 March 2010
© Springer Science+Business Media, LLC 2010

Abstract The intensity of a doubly stochastic Poisson process (DSPP) is also a stochastic process whose integral is the mean process of the DSPP. From a set of sample paths of the Cox process we propose a numerical method, preserving the monotone character of the mean, to estimate the intensity on the basis of the functional PCA. A validation of the estimation method is presented by means of a simulation as well as a comparison with an alternative estimation method.

Keywords Cox process · Monotone piecewise cubic interpolation ·
Functional principal component analysis · Functional data analysis

AMS 2000 Subject Classification 60G51 · 60G55 · 62H25 · 46N30

1 Introduction

The Cox process (CP) or doubly stochastic Poisson process is a generalization of the Poisson process whose intensity instead of being constant (homogeneous Poisson process) or a function of time (non-homogeneous Poisson process) is also a stochastic process influenced by another external one. Due to the stochastic nature of its intensity, the CP is more flexible and realistic in order to model real phenomena.

This work was partially supported by projects MTM2007-63793 of Dirección General de Investigación, Ministerio de Ciencia, P06-FQM-01470 from Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía and MTM2007-66791 of Plan Nacional I+D, Ministerio de Ciencia y Tecnología jointly by the FEDER and grant FQM-307 of Conserjería de Innovación de la Junta de Andalucía, all of them in Spain.

P. R. Bouzas (✉) · A. M. Aguilera
Department of Statistics and Operations Research, Univ. Granada, 18071, Granada, Spain
e-mail: paula@ugr.es

N. Ruiz-Fuentes
Department of Statistics and Operations Research, Univ. Jaén, 23071, Jaén, Spain

CP was first defined by Cox (1955) and it has been deeply studied for example by Snyder and Miller (1991) or Grigoriu (1995). From a martingale point of view, CP is studied by Daley and Vere-Jones (1988), Brèmaud (1981), Andersen et al. (1993), Last and Brandt (1995), among others. From all of this references we can observe that this counting process has been used in several fields like risk and insurance, medicine, signal processing, and many others.

CP is characterized by its intensity or its parametric function. The only restriction of the intensity process is that it is non negative. The parametric function of the CP is its mean and in case that it is absolutely continuous, the mean is the integral of the intensity process so that the mean is also stochastic. The mean process has non negative and nondecreasing sample paths.

Estimating the intensity process is a problem treated by many authors. For instance, Boel and Beneš (1980), Snyder and Miller (1991), Manton et al. (1999), Varini (2008), among many others, formulate several approaches using filtering methodology but it is always necessary to impose several assumptions on the intensity process. It is usual to assume that the first and second moments of the intensity are known or that it follows an explicit model. It is also more usual to approach the calculus of a linear estimation rather than non-linear even that for many applications the linearly constrained estimators are not accurate enough. In many cases, these assumptions are justified by real-world data but in many others, to establish this restrictions is just a practical way to deal with the complicated calculus of an analytic solution and computational requirements. For this reason, the development of suboptimal estimators has widely emerged.

Our work tries to deepen on how to estimate a CP without statistical assumptions on its intensity or mean processes. An attempt of relaxing the statistical assumptions was done in Bouzas et al. (2002) where it was proposed a methodology to forecast the sample paths of the CP by multivariate principal component regression in a future instant of time just from observed values.

Functional Data Analysis (FDA) models stochastic processes observed in discrete time points (as real processes can nearly always be observed) by reconstructing the functional form of their sample paths. See Ramsay and Silverman (1997) or Valderrama et al. (2000) for a deeper study. The main interest of FDA is that it is not necessary to impose a distribution to the process neither to have known moments. Bouzas et al. (2006a), proposed an estimation of the intensity process of a CP from the FDA point of view, just from observed sample paths of the CP. They were divided into subtrajectories and the intensity process was estimated in a finite set of points using a point estimator. Then, the stochastic structure of the intensity process was estimated by means of functional principal component analysis (FPCA) applied to the estimated values. In a posterior work (Bouzas et al. 2006b), it was proposed to apply FDA not to the process intensity but to the mean process of the CP and with the novelty of preserving the monotonicity property of its sample paths.

The present paper gives a new methodology of estimating the intensity process from observed trajectories of the CP, just making use of the nondecreasing monotonicity of the mean. The sketch of the present work is the following. In Section 2, once the sample paths of the CP have been observed, we obtain point estimators of the mean process at the knots of a partition of the time interval and then the functional forms of the mean sample paths are reconstructed by means of a monotone piece wise cubic interpolation (Fritsch and Carlson 1980) because of

these sample paths must be monotone increasing (Bouzas et al. 2006b). In Section 3, differentiating and expressing them in an adequate way, the functional form of sample paths of the intensity process are obtained and finally, in Section 4 its stochastic structure is estimated by FPCA.

Finally, Section 5 shows an application of the method proposed to a simulation of CP in order to validate it, as well as a comparison with the intensity modelling proposed in Bouzas et al. (2006a).

2 Reconstruction of the Mean Sample Paths

A CP $\{N(t) : t \geq t_0\}$ with intensity stochastic process $\{\lambda(t, x(t)) : t \geq t_0\}$ is defined as a conditioned PP with intensity the process $\{\lambda(t, x(t)) : t \geq t_0\}$ given the information process $\{x(t) : t \geq t_0\}$.

Therefore, the probability density function (pdf), using the conditioning method from Snyder and Miller (1991), is given by

$$\begin{aligned}
 P[N(t) = n] &= E\{P[N(t) = n/x(\sigma) : t_0 \leq \sigma < t]\} \\
 &= E\left\{\frac{1}{n!} \left(\int_{t_0}^t \lambda(\sigma, x(\sigma)) d\sigma\right)^n \exp\left(-\int_{t_0}^t \lambda(\sigma, x(\sigma)) d\sigma\right)\right\}
 \end{aligned}$$

for $n = 0, 1, 2, \dots$. Let $\Lambda(t, x(t)) = \int_{t_0}^t \lambda(\sigma, x(\sigma)) d\sigma$ be the parametric function of the CP and its mean as it is well known. Then, it is clearly a process itself influenced by the information process.

Not only the pdf depends on the intensity process (or on the mean process) but all the characteristics of the CP. That is the reason why it is so important to estimate the intensity of a CP. Our purpose is to estimate $\lambda(t, x(t))$, having no previous knowledge about its structure, from several trajectories of $N(t)$ observed on the interval $[t_0, t_0 + rT)$. From now on, the mean and the intensity processes will be denoted by $\Lambda(t)$ and $\lambda(t)$ respectively, in order not to complicate the notation.

Let us now start from the initial situation of having observed k independent trajectories of the CP $\{N(t); t \geq t_0\}$ with intensity process $\{\lambda(t); t \geq t_0\}$ in the interval $[t_0, t_0 + rT)$ denoted by

$$\{N_\omega(t) : \omega = 1, \dots, k\}$$

From each one of the observed trajectories of $N(t)$, Bouzas et al. (2006b) shows how to estimate values of the mean sample path, $\Lambda_\omega(t)$, in a finite set of time points of the shorter interval $[t_0, t_p = t_0 + T)$. It is based on splitting up each initial trajectory in r independent shorter ones all of them in $[t_0, t_p = t_0 + T)$. Therefore, we work with r subtrajectories of the Poisson process in order to estimate one of the mean process. Let us denote the estimated values by $\hat{\Lambda}_\omega(t_j)$, ($j = 0, \dots, p; \omega = 1, \dots, k$).

In order to use FPCA later on for the stochastic estimation of the intensity process, first it is necessary to reconstruct the functional form of the mean process sample paths by approximating them in a finite space generated by a basis of functions. In literature about FDA, different methods have been considered for obtaining the basis coefficients of sample paths, as for example, least squares approximation on a space generated by trigonometric functions (Aguilera et al. 1995), wavelets (Ocaña et al. 1998) or cubic B-splines (Escabias et al. 2004a) and cubic spline

interpolation (Aguilera et al. 1996; Escabias et al. 2004b). For a general review of how to reconstruct the true functional form of functional data, the interested reader can be referred to Ramsay and Silverman (1997). In order to reconstruct the functional form of each mean trajectory and knowing that they are nondecreasing monotone, Bouzas et al. (2006b) also proposed to interpolate each sample path by monotone piecewise cubic interpolation to preserve the monotonicity (Fritsch and Carlson 1980).

Let us take a fixed ω , the \mathcal{C}^1 and following Bouzas et al. (2006b), the monotone piecewise cubic interpolating function for the set of monotone data $\{(t_0, \hat{\Lambda}_\omega(t_0)), \dots, (t_p, \hat{\Lambda}_\omega(t_p))\}$ is the following

$$p_{\omega j}(t) = \hat{\Lambda}_\omega(t_j)H_1(t) + \hat{\Lambda}_\omega(t_{j+1})H_2(t) + d_{\omega j}H_3(t) + d_{\omega j+1}H_4(t), \quad t \in [t_j, t_{j+1})$$

$$j = 0, \dots, p-1 \quad (1)$$

where $d_{\omega j} = \left. \frac{dp_{\omega j}(t)}{dt} \right|_{t=t_j}$, $d_{\omega j+1} = \left. \frac{dp_{\omega j}(t)}{dt} \right|_{t=t_{j+1}}$ and $H_s(t)$ are the usual Hermite basis functions for the interval $[t_j, t_{j+1}]$ given by

$$H_1(t) = \phi\left(\frac{t_{j+1}-t}{h_j}\right), \quad H_2(t) = \phi\left(\frac{t-t_j}{h_j}\right)$$

$$H_3(t) = -h_j\psi\left(\frac{t_{j+1}-t}{h_j}\right), \quad H_4(t) = h_j\psi\left(\frac{t-t_j}{h_j}\right) \quad (2)$$

with $h_j = t_{j+1} - t_j$, $\phi(x) = 3x^2 - 2x^3$ and $\psi(x) = x^3 - x^2$. For every subinterval $[t_j, t_{j+1}]$, the derivatives d_j and d_{j+1} are previously calculated by an algorithm initiated with the values given by the standard three-point difference formula to be satisfactory for preserving the monotonicity (Fritsch and Carlson 1980). Then, we have obtained the functional reconstruction of each mean sample paths; let us denote them as $\hat{\Lambda}_\omega(t)$, $t \in [t_0, t_p = t_0 + T)$, $\omega = 1, \dots, k$.

3 Functional Reconstruction of the Intensity Sample Paths

This section shows how to obtain a functional expression of the intensity sample paths of the CP corresponding to the mean sample paths $\hat{\Lambda}_\omega(t)$.

As $\Lambda(t) = \int_{t_0}^t \lambda(\sigma) d\sigma$, we will obtain a \mathcal{C}^0 functional estimation of the intensity sample paths by differentiating in Eq. 1. Choosing a fixed ω , the piecewise interpolated function is

$$p'_{\omega j}(t) = \hat{\Lambda}_\omega(t_j)H'_1(t) + \hat{\Lambda}_\omega(t_{j+1})H'_2(t) + d_{\omega j}H'_3(t) + d_{\omega j+1}H'_4(t), \quad t \in [t_j, t_{j+1})$$

$$j = 0, \dots, p-1 \quad (3)$$

This polynomial has the inconvenient that it is not expressed in terms of a basis and not in the whole interval $[t_0, t_p)$, nevertheless we need it for its further use in FPCA.

The piecewise polynomial defined in each subinterval by Eq. 3 is quadratic and must be \mathcal{C}^0 because it is the differentiation of a cubic \mathcal{C}^1 interpolation polynomial. It is known that the \mathcal{C}^0 quadratic Lagrange basis in the interval $[t_j, t_{j+1}]$ is $\langle F_1, F_2, F_3 \rangle$

where

$$F_1 = \frac{(t - t_{j+1})(2t - t_j - t_{j+1})}{h_j^2}, \quad F_2 = \frac{-4(t - t_j)(t - t_{j+1})}{h_j^2}$$

$$\text{and } F_3 = \frac{(t - t_j)(2t - t_j - t_{j+1})}{h_j^2}$$

therefore, it is necessary to express each $H'_1(t)$, $H'_2(t)$, $H'_3(t)$ and $H'_4(t)$ as a linear combination of the basis functions, it is

$$\begin{aligned}
 & aF_1 + bF_2 + cF_3 \\
 &= \left(\frac{-4b + 2a + 2c}{h_j^2} \right) t^2 \\
 &+ \left(\frac{-2at_{j+1} + a(-t_j - t_{j+1}) + 4bt_j + 4bt_{j+1} + c(-t_j - t_{j+1}) - 2ct_j}{h_j^2} \right) t \\
 &+ \frac{-at_{j+1}(-t_j - t_{j+1}) + ct_j(t_j + t_{j+1}) - 4bt_jt_{j+1}}{h_j^2} \tag{4}
 \end{aligned}$$

Let us start deriving the linear combination $H'_1(t) = aF_1 + bF_2 + cF_3$. As mentioned above, $H'_1(t) = \phi' \left(\frac{t_{j+1}-t}{h_j} \right)$ so that

$$H'_1(t) = 6 \frac{t^2}{h_j^3} + \left(\frac{6}{h_j^2} - 12 \frac{t_{j+1}}{h_j^3} \right) t - 6 \frac{t_{j+1}}{h_j^2} + 6 \frac{t_{j+1}}{h_j^3} \tag{5}$$

Making Eq. 5 equal to Eq. 4, we have the three following equations

$$\begin{aligned}
 -4b + 2a + 2c &= \frac{6}{h_j} \\
 -2at_{j+1} + a(-t_j - t_{j+1}) + 4bt_j + 4bt_{j+1} - 2ct_j + c(-t_j - t_{j+1}) &= 6 - 12 \frac{t_{j+1}}{h_j} \\
 -at_{j+1}(-t_j - t_{j+1}) - ct_j(-t_j - t_{j+1}) - 4bt_jt_{j+1} &= -6t_{j+1} + 6 \frac{t_{j+1}^2}{h_j}
 \end{aligned}$$

so we have that $H'_1(t) = aF_1 + bF_2 + cF_3$ where

$$\begin{aligned}
 a &= -6 \frac{t_j - t_{j+1} + h_j}{h_j(-t_j + t_{j+1})} = 0 \\
 b &= -\frac{3}{2} \frac{t_j - t_{j+1} + 2h_j}{h_j(-t_j + t_{j+1})} = -\frac{3}{2} \frac{h_j}{h_j(-t_j + t_{j+1})} = \frac{-3}{2h_j} \\
 c &= 0
 \end{aligned}$$

Let us now find the linear combination $H'_2(t) = aF_1 + bF_2 + cF_3$. Because of $H'_2(t) = \phi' \left(\frac{t-t_j}{h_j} \right)$ we have

$$H'_2(t) = -6\frac{t^2}{h_j^3} + \left(\frac{6}{h_j^2} + 12\frac{t_j}{h_j^3} \right) t - 6\frac{t_j}{h_j^2} - 6\frac{t_j}{h_j^3} \quad (6)$$

Making Eq. 6 equal to Eq. 4, we have that

$$\begin{aligned} \frac{1}{h_j^2} (-4b + 2a + 2c) t^2 &= -6\frac{t^2}{h_j^3} \\ \frac{1}{h_j^2} (-2at_{j+1} + a(-t_j - t_{j+1}) + 4bt_j + 4bt_{j+1} - 2ct_j + c(-t_j - t_{j+1})) t \\ &= \left(\frac{6}{h_j^2} + 12\frac{t_j}{h_j^3} \right) t \\ \frac{1}{h_j^2} (-at_{j+1}(-t_j - t_{j+1}) - ct_j(-t_j - t_{j+1}) - 4bt_j t_{j+1}) &= - \left(6\frac{t_j}{h_j^2} + 6\frac{t_j^2}{h_j^3} \right) \end{aligned}$$

then, $a = 0$, $b = \frac{3}{2h_j}$, $c = 0$.

Analogously, for $H'_3(t) = -h_j\psi' \left(\frac{t_{j+1}-t}{h_j} \right)$ we find out that $a = 1$, $b = -\frac{1}{4}$ and $c = 0$.

And for $H'_4(t) = h_j\psi' \left(\frac{t-t_j}{h_j} \right)$ the solution is $a = 0$, $b = -\frac{1}{4}$ and $c = 1$.

In conclusion, we obtain that

$$\begin{aligned} H'_1(t) &= -\frac{3}{2h_j} F_2(t), & H'_2(t) &= \frac{3}{2h_j} F_2(t) \\ H'_3(t) &= F_1(t) - \frac{1}{4} F_2(t), & H'_4(t) &= F_3(t) - \frac{1}{4} F_2(t) \end{aligned}$$

Therefore, for every $j = 0, \dots, p-1$, the piecewise functional estimation of the intensity sample path becomes

$$\begin{aligned} p'_{\omega j}(t) &= \hat{\Lambda}_{\omega}(t_j) \frac{-3}{2h_j} F_2(t) + \hat{\Lambda}_{\omega}(t_{j+1}) \frac{3}{2h_j} F_2(t) + d_{\omega j} \left(F_1(t) - \frac{1}{4} F_2(t) \right) \\ &+ d_{\omega j+1} \left(F_3(t) - \frac{1}{4} F_2(t) \right), \quad t \in [t_j, t_{j+1}) \end{aligned}$$

Considering

$$\begin{aligned} F_{1j} &= \begin{cases} F_1, & t \in [t_j, t_{j+1}) \\ 0, & \text{otherwise} \end{cases}, & F_{2j} &= \begin{cases} F_2, & t \in [t_j, t_{j+1}) \\ 0, & \text{otherwise} \end{cases} \\ \text{and } F_{3j} &= \begin{cases} F_3, & t \in [t_j, t_{j+1}) \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

instead of F_1, F_2 and F_3 , we could rewrite $p'_{\omega j}(t)$ with these functions and then derive a functional expression for the intensity sample path in the whole interval $[t_0, t_p]$ as

$$p'_{\omega}(t) = \sum_{j=0}^p p'_{\omega j}(t), \quad t \in [t_0, t_p)$$

and after some manipulations we can write it the following way

$$p'_{\omega}(t) = \sum_{j=0}^{p-1} F_{2j}(t) \left[\frac{3}{2h_j} \left(\hat{\Lambda}_{\omega}(t_{j+1}) - \hat{\Lambda}_{\omega}(t_j) \right) - \frac{d_{\omega j} + d_{\omega j+1}}{4} \right] + \sum_{j=0}^{p-2} \left(F_{3j}(t) + F_{1j+1}(t) \right) d_{\omega j+1} + F_{10}(t)d_{\omega 0} + F_{3p-1}(t)d_{\omega p} \tag{7}$$

As mentioned above, $p_{\omega}(t)$ was a piecewise cubic spline interpolation in C^1 so $p'_{\omega}(t)$ is a quadratic polynomial in C^0 . Therefore, its dimension must be $2p + 1$. Clearly, the functions $F_{2j}(t), j = 0, \dots, p - 1; (F_{3j}(t) + F_{1j+1}(t)), j = 0, \dots, p - 2; F_{10}(t)$ and $F_{3p-1}(t)$ are $2p + 1$ independent quadratic functions in C^0 so $p'_{\omega}(t), \omega = 0, \dots, k$ are now expressed in terms of a basis on the whole interval $[t_0, t_p)$.

From now on, we will denote this basis by $\{B_l(t)\}_{l=1, \dots, 2p+1}$ and its coefficients in Eq. 7 by $a_{\omega l}, l = 1, \dots, 2p + 1$ and $\omega = 0, \dots, k$ in order to unify the notation. The expression of Eq. 7 for all the sample paths can be rewritten in a matrix form like this

$$p'(t) = A B(t), \quad t \in [t_0, t_p)$$

where

$$p'(t) = (p'_1(t), \dots, p'_k(t))^T$$

$$A = (a_{\omega l})_{\omega=1, \dots, k; l=1, \dots, 2p+1}$$

$$B(t) = (B_1, \dots, B_{2p+1})^T$$

4 Stochastic Estimation of the Intensity Process

This section estimates the continuous stochastic structure of the intensity process. Having reconstructed its functional sample paths in Section 2, now FPCA will be applied to reduce dimension and derive an orthogonal expansion for the intensity in terms of uncorrelated random variables.

By analogy with the multivariate case, the functional principal components of a second order and quadratic mean stochastic process with sample paths in the space $L^2[t_0, t_p]$ of square integrable functions are defined as uncorrelated generalized linear combinations of the process variables whose weight functions (principal factors) are obtained as the eigenfunctions of the sample covariance kernel (Ramsay and Silverman 1997). In order to obtain the principal components (p.c.'s) of a stochastic process with sample paths in a finite dimension space generated by a set of linearly independent functions, Ocaña et al. (1999) proved the equivalence between FPCA with respect to the usual inner product in $L^2[t_0, t_p]$ and standard multivariate PCA in \mathbb{R}^{2p+1} . In this paper, we have adapted this result to the case of functional

sample paths expressed in terms of the basis $\{B_l(t)\}_{l=1,\dots,2p+1}$ found in the last section obtaining the following equivalence between multivariate and functional PCA's.

Let us consider the centered estimated intensity process

$$\bar{p}'(t) = p'(t) - \mu_{p'}(t) = (A - \bar{A}) B(t)$$

where $\bar{A} = (\bar{a}_{\omega l})$ with elements $\bar{a}_{\omega l} = \frac{1}{k} \sum_{\omega=1}^k a_{\omega l}$ ($l = 1, \dots, 2p+1$, $\omega = 1, \dots, k$) and $\mu_{p'}(t) = \bar{A} B(t)$.

Let $(A - \bar{A})$ be the matrix whose rows are the coefficients of each trajectory of the centered process with respect to the basis $\langle B_l(t) \rangle_{l=1,\dots,2p+1}$ and \mathbb{P} the matrix whose components are the usual inner products between basis functions given by $\langle B_i, B_j \rangle_{\mu} = \int_{t_0}^{t_p} B_i(t) B_j(t) dt$. Then, FPCA of $\bar{p}'_{\omega}(t)$ in the space generated by the basis $\{B_l(t)\}_{l=1,\dots,2p+1}$ with respect to the usual metric in $L^2[t_0, t_p]$ is equivalent to multivariate PCA of the data matrix $(A - \bar{A}) \mathbb{P}^{1/2}$ with respect to the usual inner product in \mathbb{R}^{2p+1} . Let us observe that in the case of an orthonormal basis ($\mathbb{P} = \mathbb{I}$) FPCA is equivalent to multivariate PCA of the basis coefficients centered data matrix.

Once the eigenvectors g_j of the covariance matrix of $(A - \bar{A}) \mathbb{P}^{1/2}$ have been computed, the sample paths of the interpolated process $\bar{p}'(t)$ can be represented as follows in term of its p.c.'s

$$\bar{p}'_{\omega}(t) = \sum_{j=1}^{2p+1} \zeta_{\omega j} f_j(t), \quad \omega = 1, \dots, k$$

where $f_j(t)$ are eigenfunctions of the sample covariance matrix of the centered process given by

$$f_j(t) = \sum_{l=1}^{2p+1} f_{lj} B_l(t)$$

with the vector of coefficients being $f_j = \mathbb{P}^{-1/2} g_j$ and

$$\zeta_{\omega j} = \int_{t_0}^{t_p} \bar{p}'_{\omega}(t) f_j(t) dt = (A_{\omega} - \bar{A}_{\omega}) \mathbb{P}^{1/2} g_j$$

where $(A_{\omega} - \bar{A}_{\omega})$ is the ω th row of $(A - \bar{A})$.

Then, the intensity process $\lambda(t)$ can be approximated by the following truncated orthogonal p.c.'s decomposition of its estimation denoted by $\lambda^q(t)$

$$\lambda^q(t) = p'^q(t) = \mu_{p'}(t) + \sum_{j=1}^q \zeta_j f_j(t) \quad (8)$$

This way, the dimension $2p+1$ is reduced to q so that we get to explain a high proportion of the total variance of the mean process (as next to 1 as possible) given by $\frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^{2p+1} \lambda_j}$ where λ_j is the variance of the j th p.c. ζ_j given by the j th eigenvalue of the covariance matrix of $(A - \bar{A}) \mathbb{P}^{1/2}$ associated to the j th eigenvector g_j .

5 Simulation

This section presents how to apply the methodology of estimation of the intensity process presented in the previous sections to a simulation of a CP. Firstly, we have simulated 150 sample paths $(N_1(t), \dots, N_{150}(t))$ of a Poisson process, $N(t)$, whose random intensity is a random variable with uniform distribution in $[0, 1]$ each instant of time of the time interval $[0, 300]$. The simulation method of a CP is an adaptation of a former one published in Grigoriu (1995).

The sample paths of $\Lambda(t)$ have been estimated as explained in Bouzas et al. (2006b). Each long trajectory of $N(t)$ in $[0, 300]$ has been split up into 30 shorter ones of equal amplitude 10. Therefore, from each original sample path $N_\omega(t)$ we obtain 30 subtrajectories $N_{\omega i}(t)$, $t \in [0, 10)$ with $i = 1, \dots, 30$ and $\omega = 1, \dots, 150$. Clearly, we have fixed $k = 150$, $t_0 = 0$, $r = 30$, $t_0 + rT = 300$ and so $t_0 + T = 10$ (see Section 2). For each ω , we have estimated the mean process in $t_j = j \times \frac{1}{2}$ ($j = 0, \dots, 20$) by means of a point estimator. That is, for each ω , we have estimated a sample path of the mean in 21 equally spaced time points in the interval $[0, 10)$. Then, the functional form of the 150 sample paths of the mean are reconstructed by monotone piece wise cubic interpolation as explained in the mentioned article.

Afterwards, we have differentiated the interpolated mean trajectories to derive an estimation of the functional intensity trajectories as mentioned in Section 3. Finally, we have applied FPCA to the estimated intensity trajectories (see Section 4). As the p.c.'s ζ_j , the eigenfunctions $f_j(t)$ and the eigenvalues λ_j have been computed, we have found out that the first $q = 15$ p.c.'s accumulate 85.76% of the total variance. Then, the intensity process has been modelled in terms of the first 15 p.c.'s by means of the orthogonal expansion given by Eq. 8 with $q = 15$. Several simulated subtrajectories, the functional reconstruction of the intensity sample paths and stochastically estimated sample paths are represented in Figs. 1 and 2.

In order to validate the method, we try to evaluate its accuracy. The usual way of doing it is to compare the estimated parameter with the real one used to obtain the simulation. In our case, as we explained in Section 2, we have estimated a sample path from r simulated ones so we compare it with all of them. Therefore, the mean square error of estimation of $\lambda(t)$ is defined by

$$\varepsilon^2 = \frac{1}{k} \sum_{\omega=1}^k \varepsilon_\omega^2 \quad \text{where} \quad \varepsilon_\omega^2 = \frac{1}{p} \sum_{j=0}^p [\lambda_\omega^q(t_j) - \lambda_\omega(t_j)]^2; \quad \omega = 1, \dots, k$$

The error ε^2 has become 0.1014 which is quite small and the error variability between sample paths has become 0.0148, so this suggest that the method is good and stable.

Other simulations have been developed to compare the results. It can be observed that besides that the behavior of the errors is similar for all the simulations, this methodology has also the property of producing small errors even with few subtrajectories which would be important in case that a CP can not be observed in a very long time interval so it were not possible to split it in a lot of subtrajectories. Table 1 shows the estimation errors and their standard deviation when applying the methodology to several examples of simulations with different number of trajectories and subtrajectories. Then, we can conclude the goodness and stability of the estimation.

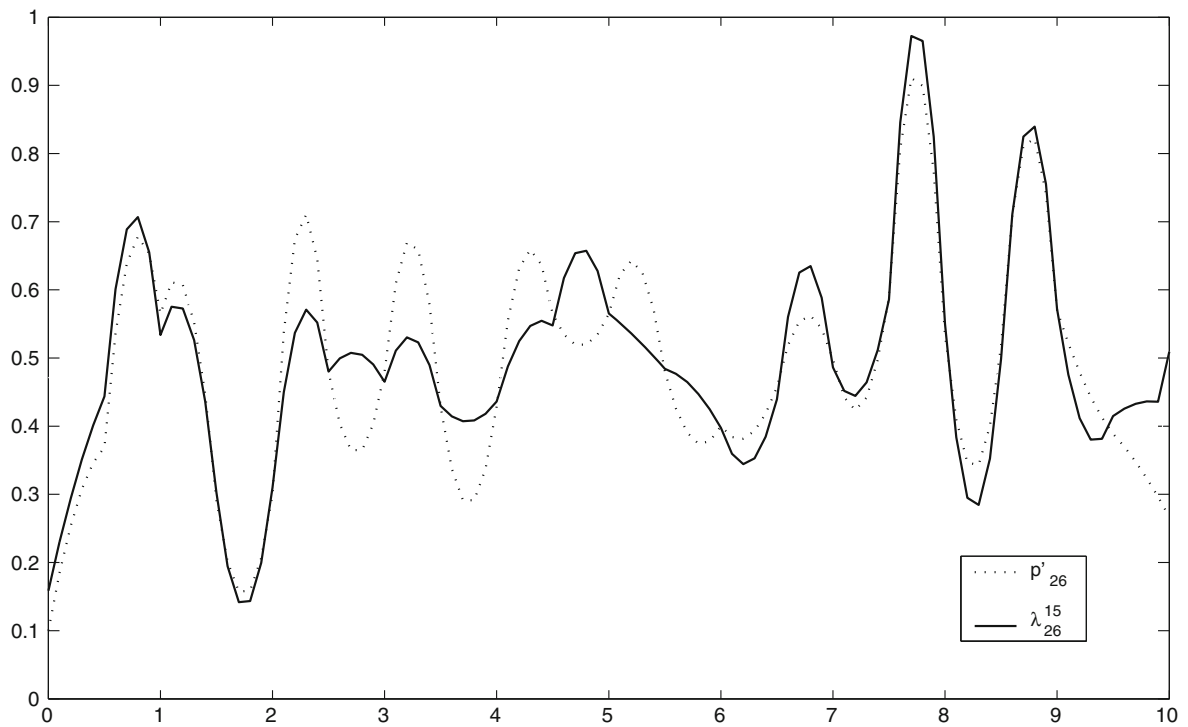


Fig. 1 Functional reconstruction and estimated sample path for the intensity process, $\omega = 26$

5.1 Comparison with a Former Methodology

An estimation of the intensity process of a CP without previous statistic information about the process has already been approached in Bouzas et al. (2006a). From

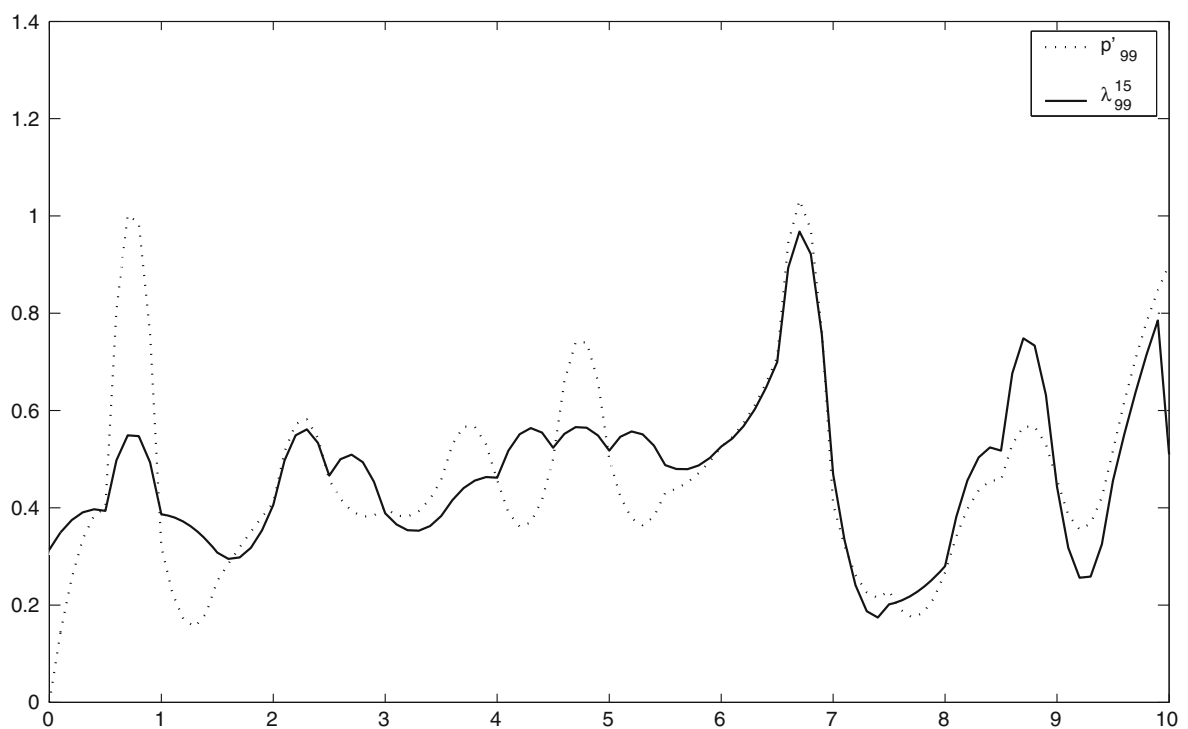


Fig. 2 Functional reconstruction and estimated sample path for the intensity process, $\omega = 99$

Table 1 Estimation error and standard deviation with different number of trajectories and subtrajectories

k	r	ε^2	$\sigma(\varepsilon_\omega^2)$
150	30	0.1014	0.0148
	10	0.1023	0.0265
	5	0.0974	0.0238
40	30	0.1014	0.025
	10	0.1	0.0259
	5	0.0953	0.0235
25	30	0.1025	0.0255
	10	0.1031	0.0267
	5	0.0986	0.0217

observed sample paths of the CP, the intensity process was estimated in a finite set of time points by means of a point estimator which was proved to be unbiased and consistent. The functional form of the intensity sample paths were reconstructed by an usual cubic spline interpolation. Then, FPCA applied to them derived the stochastic estimation of the intensity process.

As explained in the text, in the present paper we develop a new functional reconstruction of the intensity sample paths by means of a previous reconstruction of the mean sample paths preserving their analytical property of monotonicity. After differentiation of the mean sample paths, the adapted FPCA to the new framework using the proposed basis function becomes an “ad hoc” estimation method for the intensity process of a CP. Therefore, we are comparing two methods of modelling the intensity process of a CP by means of FDA. The difference between them is that the former one is easier to compute but does not take into account any theoretical property of the mean or intensity and the present one has more steps but preserves

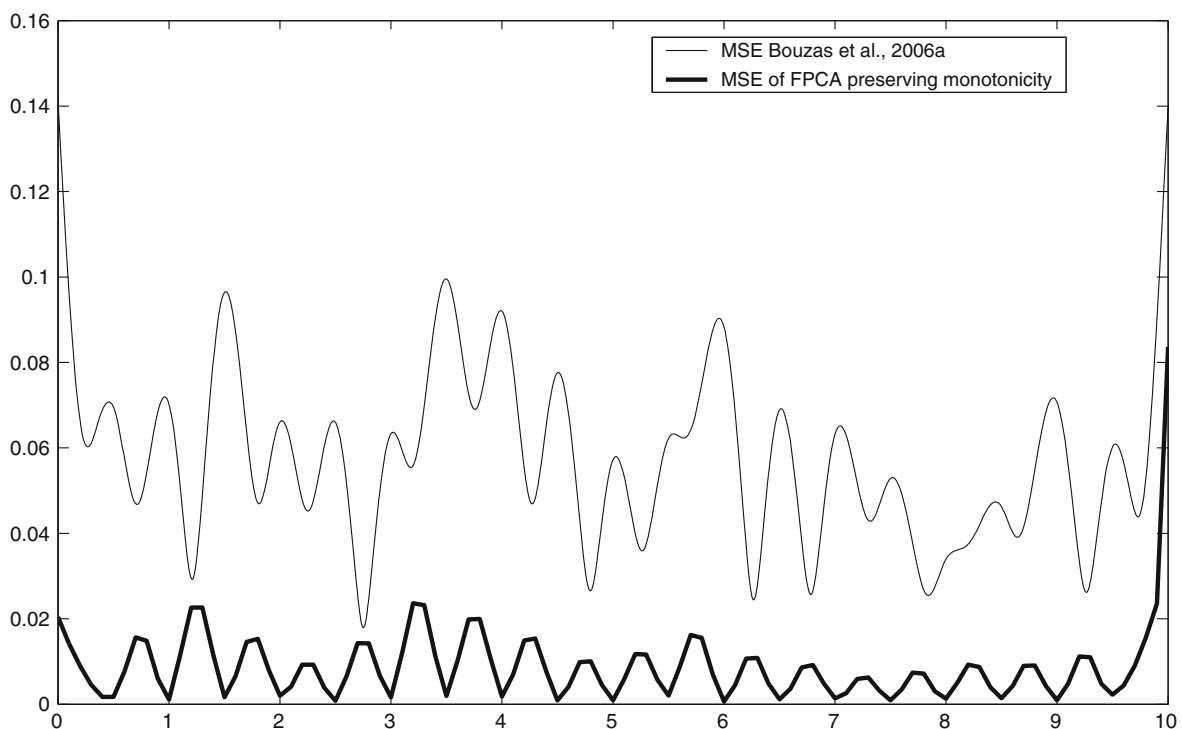


Fig. 3 FPCA errors in both methods

the monotonicity of the mean sample paths so it is more accurate from the theoretical point of view. Besides, even the method we propose in this paper is more complex, the errors made are smaller. In fact, Fig. 3 shows the FPCA errors computed applying the former method and the smaller present ones based in a simulation treated with both methods. The simulation of the CP is the one used as example in Section 5 ($k = 150$, $r = 30$, $t_0 = 0$ and $t_0 + T = 10$). Then, even though the former method of estimation provided good estimations, the new method proposed in this paper gives better results.

6 Conclusions

The intensity process of a CP characterize it so it is very important to develop an accurate estimation of it. Due to the fact that in many real point phenomena, it is not known the theoretical structure of its intensity or any of its moments, we work in the framework of FDA. The intensity has been former estimated applying usual methods in this field deriving good results. Even though, we deepen in this estimation trying to give a finest one.

This paper proposes a FPCA method adapted to preserve the monotonicity of the mean process of a CP. This method is based on the point estimation of the mean thanks to the independence of increments of the CP, a functional estimation of its sample paths using a basis of functions that preserve their known monotonicity, the relation between mean and intensity processes to obtain the estimated sample paths of the intensity and a further application of FPCA with the mentioned basis.

This new method is more theoretically complicated than the former one also in the field of FDA but more accurate as it is thought specially for the characteristics of a CP. A common problem in observing this kind of processes is that long sample paths are not always available so it is important that this method also gives good estimations when using few subtrajectories. The complicated calculations for estimating the functional mean sample paths and finally the stochastic intensity model as well as the simulation technique have been implemented in MatLab so as an interest in the short term it is to organize all the programs involved in a MatLab package. Also, the proposed estimation of the intensity can be used in generalizations a CP such as marked or filtered CP and it can be a good basis in order to forecast a CP in a future instant of time.

References

- Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York
- Aguilera AM, Gutiérrez R, Ocaña FA, Valderrama MJ (1995) Computational approaches to estimation in the principal component analysis of a stochastic process. *Appl Stoch Models Data Anal* 11(4):279–299
- Aguilera AM, Gutiérrez R, Valderrama MJ (1996) Approximation of estimators in the PCA of a stochastic process using B-splines. *Commun Stat (Simulation)* 25(3):671–690
- Boel R, Beneš V (1980) Recursive nonlinear estimation of a diffusion acting as the rate of an observed Poisson process. *IEEE Trans Inf Theory* 26(5):561–575
- Bouzas PR, Aguilera AM, Valderrama MJ (2002) Forecasting a class of doubly stochastic Poisson processes. *Stat Pap* 43:507–523

- Bouzas PR, Aguilera AM, Valderrama MJ, Ruiz-Fuentes N (2006a) On the structure of the stochastic process of mortgages in Spain. *Comput Stat* 21:73–89
- Bouzas PR, Valderrama MJ, Aguilera AM, Ruiz-Fuentes N (2006b) Modelling the mean of a doubly stochastic Poisson process by functional data analysis. *Comput Stat Data Anal* 50:2655–2667
- Brèmaud P (1981) *Point processes and queues: Martingale dynamics*. Springer, New York
- Cox DR (1955) Some statistical methods connected with series of events. *J R Stat Soc B* 17:129–164
- Daley DJ, Vere-Jones D (1988) *An introduction to the theory of point processes*. Springer, New York
- Escabias M, Aguilera AM, Valderrama MJ (2004a) Principal component estimation of functional logistic regression: discussion of two different approaches. *J Nonparametr Stat* 16(3–4):365–384
- Escabias M, Aguilera AM, Valderrama MJ (2004b) Modelling environmental data by functional principal component logistic regression. *Environmetrics* 16(1):95–107
- Fritsch FN, Carlson RE (1980) Monotone piecewise cubic interpolation. *SIAM J Numer Anal* 17(2):238–246
- Grigoriu M (1995) *Applied non-Gaussian processes, examples theory, simulation, linear random vibration and MATLAB solutions*. PTR Prentice Hall, Upper Saddle River
- Last G, Brandt A (1995) *Marked point processes on the real line (The dynamic approach)*. Springer, New York
- Manton JH, Krishnamurty V, Elliot RJ (1999) Discrete time filters for doubly stochastic Poisson processes and other exponential noise models. *Int J Adapt Control Signal Process* 13:393–416
- Ocaña FA, Aguilera AM, Valenzuela O (1998) A wavelet approach to functional principal component analysis. In: Payne R, Green P (eds) *Proceedings of COMPSTAT 98*. Physica, Heidelberg, pp 413–418
- Ocaña FA, Aguilera AM, Valderrama JM (1999) Functional principal components analysis by choice of norm. *J Multivar Anal* 15:262–276
- Ramsay JO, Silverman BM (1997) *Functional data analysis*. Springer, New York
- Snyder DL, Miller MI (1991) *Random point processes in time and space*, 2nd edn. Springer, New York
- Valderrama MJ, Aguilera AM, Ocaña FA (2000) *Predicción dinámica mediante análisis de datos funcionales*. La Muralla, Madrid
- Varini E (2008) A Monte Carlo method for filtering a marked doubly stochastic Poisson process. *Stat Methods Appl* 17:183–193