Kernel estimation of the intensity of Cox processes

Nicolas Klutchnikoff [1] & Gaspar Massiot [2]

[1] *CREST (ENSAI) and IRMA (UMR 7501 Université de Strasbourg, CNRS), Campus de Ker-Lann, Rue Blaise Pascal, BP 37203, 35172 BRUZ cedex, France*
nicolas.klutchnikoff@ensai.fr
[2] *CREST (ENSAI) and IRMAR (UMR 6625 ENS Rennes, CNRS, UEB), Campus de Ker-Lann, Avenue Robert Schuman, 35170 BRUZ, France*
gaspar.massiot@ensai.fr

Résumé. Un processus de Cox d'intensité aléatoire $\lambda = (\lambda(t))_{t\in[0,1]}$ est un processus de comptage $N = (N_t)_{t\in[0,1]}$ tel que la loi conditionnelle de N sachant λ est un processus de Poisson d'intensité λ. Par abus, nous appellerons processus de Cox un processus de comptage $N = (N_t)_{t\in[0,1]}$ accompagné d'un co-processus $Z == (Z_t)_{t\in[0,1]}$ tel que, conditionnellement à Z, la loi de N est un processus de Poisson d'intensité $\theta(Z)$ avec θ une fonction déterministe. Idéalement, on voudrait estimer la fonction θ à partir d'un n-échantillon $(N^1, Z^1), \ldots, (N^n, Z^n)$ de copies de (N, Z). Cependant, un telle approche se heurte inévitablement au fléau de la dimension, car la convariable est à valeurs dans un espace de dimension infinie. En pratique, il n'est souvent pas nécessaire, ou tout du moins ce n'est pas strictement nécessaire pour la modélisation, d'observer toute la trajectoire du co-processus, mais seulement ses valeurs en des instants aléatoires. De la sorte, si le co-processus n'est observé qu'en un nombre fini d'instants aléatoires, on circonvient au fléau de la dimension. Nous construisons et étudions sous ce modèle les propriétés d'un estimateur de type noyau pour la fonction θ. Sa consistance, un théorème de la limite centrale ainsi qu'une vitesse de convergence pour l'erreur quadratique moyenne sont données.

Mots-clés. Processus de Cox, Estimateur à noyaux

Abstract. A Cox process $N = (N_t)_{t\in[0,1]}$ with random intensity $\lambda = (\lambda(t))_{t\in[0,1]}$ is formally defined as a counting process such that the conditional distribution of N given λ is a Poisson process with intensity λ. By a slight abuse we shall call Cox process a counting process $N = (N_t)_{t\in[0,1]}$ accompanied with a co-process $Z = (Z_t)_{t\in[0,1]}$ such that the conditional law of N given Z is a Poisson process with intensity $\theta(Z)$ where θ is a deterministic function. From a statistical point of view one of the major issue is to estimate the deterministic function θ using n independent copies $(N^1, Z^1), \ldots, (N^n, Z^n)$ of (N, Z). However, such an approach is subject to the curse of dimensionality as the covariate Z takes its values in an infinite dimension space. When dealing with practical problems it is often unnecessary or at least not strictly required for the modelling to observe the full trajectory of the co-process. One can instead observe the values taken by the co-process at some random times. The co-process is then observed at a finite number of random times thereby circumventing the curse of dimensionality. We construct and study the properties of a kernel estimator for the function θ. Its consistency, a central limit theorem and a rate of convergence of its mean square error are given.

Keywords. Cox process, Kernel estimator

1 Introduction

Counting processes have been used for many years to model a large variety of situations from neuroscience (Bialek *et al.* 1991; Brette 2008; Krumin and Shoham 2009) to seismic (Ogata 1988), financial (Merton 1976), insurance (Asmussen and Albrecher 2010) or biophysical data (Kou *et al.* 2005). The Poisson process has the simplest structure of any counting process. Cox processes are more flexible processes due to their random intensity. A Cox process $N = (N_t)_{t\in[0,1]}$ with random intensity $\lambda = (\lambda(t))_{t\in[0,1]}$ is formally defined as a counting process such that the conditional distribution of N given λ is a Poisson process with intensity λ.

Note that when Cox process data arise, the intensity of the process is mainly not directly observed but a co-process is observed instead. Returning to one of the previous example, in single-molecule experiments only the peaks inducing the counting process and an underlying process are observed (Kou *et al.* 2005). Another example is to be found in car insurance (Asmussen and Albrecher 2010) where the counting process models the occurrence of car crash that are subject to weather conditions. In these cases the counting process $N = (N_t)_{t\in[0,1]}$ that naturally raises is accompanied with a co-process $Z = (Z_t)_{t\in[0,1]}$, such that the conditional law of N given Z is a Poisson process with intensity $\theta(Z)$ where θ is a deterministic function. By a slight abuse we shall call Cox process such a counting process. From a statistical point of view one of the major issue is to estimate the deterministic function θ using n independent copies $(N^1, Z^1), \ldots, (N^n, Z^n)$ of (N, Z). However, such an approach is subject to the curse of dimensionality as the covariate Z takes its values in an infinite dimension space as seen in O'Sullivan (1993).

When dealing with practical problems it is often unnecessary or at least not strictly required for the modelling to observe the full trajectory of the co-process. One can instead observe the values taken by the co-process at some random times. The co-process is then observed at a finite number of random times thereby circumventing the curse of dimensionality.

We consider the following model: $N = (N_t)_{t\in[0,1]}$ is the counting process; $Z = (Z_t)_{t\in[0,1]}$ is the co-process such that for all $t \in [0, 1]$, Z_t is a $\mathbb{R}^d$ valued random variable. The random times are $S_1 < S_2 < \ldots$. We assume that given $S = (S_1, S_2, \ldots)$, N is a Cox process with intensity function

$$\begin{aligned}\lambda : [0,1] &\to \mathbb{R}_+\\ t &\mapsto \theta_S\big(t, \vec{Z}_S(t)\big),\end{aligned}$$

where $\vec{Z}_S(t) = (Z_{S_1}, \dots, Z_{S_{M_S(t)}})$, $M_S(t) = \max(l > 0 : S_l \leq t)$ and for all $t \in [0,1]$, $\theta_S(t, \cdot)$ is a real valued function defined in $\mathbb{R}^{dM_S(t)}$.

In the sequel we consider that given S, $(N^1, Z^1), \dots, (N^n, Z^n)$ are *i.i.d.* copies of (N, Z). We propose the following kernel estimator for the function θ_S defined for all $t \in [0,1]$ and $z \in \mathbb{R}^{dM_S(t)}$:

$$\hat{\theta}_{h,\eta}(t,z) = \frac{\frac{1}{n}\sum_{k=1}^n H_\eta\left(z - \vec{Z}_S^k(t)\right)\sum_{i=1}^{N_t^k} K_h(t - T_i^k)}{\frac{1}{n}\sum_{l=1}^n H_\eta\left(z - \vec{Z}_S^l(t)\right) \vee a_n},$$

where a_n is a decreasing non-negative sequence with $a_n \underset{n\to+\infty}{\to} 0$, $H = \mathcal{H}^{\otimes dM_S(t)}$, with $\mathcal{H} : \mathbb{R} \to \mathbb{R}$ a kernel, $H_\eta(\cdot) = \frac{1}{\eta^{dM_S(t)}} H\left(\frac{\cdot}{\eta}\right)$, η a bandwidth, $K : \mathbb{R}_+ \to \mathbb{R}$ a kernel, $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ and h a bandwidth.

Note that $M_S(t)$ increases as t increases, which represents the fact that as t increases we get more information. As a consequence, it is to be expected that as t increases, the convergence rate decreases.

2 Results

In this section we give some regularity results on the suggested estimator to compare it with the usual regularity properties of kernel based non parametric estimators.

To this end suppose that given S, $\vec{Z}_S(t)$ admits a density $f_{\vec{Z}_S(t)|S}$. We need the following assumptions: given S, for fixed $t \in [0,1]$,

(H1) $\theta_S :]S_{M_S(t)}, S_{M_S(t)+1}[\times \mathbb{R}^{dM_S(t)} \to \mathbb{R}_+$ and $f_{\vec{Z}_S(t)|S}$ are continuous, positive functions;

(H2) $\theta_S :]S_{M_S(t)}, S_{M_S(t)+1}[\times \mathbb{R}^{dM_S(t)} \to \mathbb{R}_+$ and $f_{\vec{Z}_S(t)|S}$ are twice differentiable and have bounded continuous partial derivatives.

(H3) $\|\theta_S\|_\infty < \infty$ and $0 < F_0 \leq f \leq F_\infty < \infty$.

Remark that we make local regularity assumptions on θ_S here. Indeed t is almost surely not an arrival time of the process N and for all $u \in]S_{M_S(t)}, S_{M_S(t)+1}[$, $\theta_S(u, \cdot)$ is a real valued function defined in $\mathbb{R}^{dM_S(t)}$.

The following usual assumptions are also made on the two kernels

(H4) $\text{supp}\mathcal{H} = [-1,1]$, $\mathcal{H} \in \mathbb{L}^3(\mathbb{R})$, $\|\mathcal{H}\|_2 < \infty$, $\|\mathcal{H}\|_\infty < \infty$, $\int_{\mathbb{R}} \mathcal{H}(u)\mathrm{d}u = 1$ and for all $k \in \{1,2\}$, $\int_{\mathbb{R}} u^k \mathcal{H}(u) = 0$.

(H5) $\text{supp}K = [0,1]$, $K \in \mathbb{L}^3(\mathbb{R})$, $\|K\|_2 < \infty$, $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$ and for all $k \in \{1,2\}$, $\int_{\mathbb{R}} u^k K(u) = 0$.

In the following we take $a_n = (n\eta^{dM_S(t)})^{\varepsilon-1}$ for $\varepsilon \in (0, 1/2)$. We study the point-wise mean squared error function of our estimator denoted for fixed $t \in [0,1]$, $z \in \mathbb{R}^\infty$, $h > 0$ and $\eta > 0$ by

$$\text{MSE}(t,z) = \mathbb{E}\left[\hat{\theta}_{h,\eta}\big(t, z_S(t)\big) - \theta_S\big(t, z_S(t)\big)\right]^2,$$

where $z_S(t)$ is the projection on $\mathbb{R}^{dM_S(t)}$ of z.

From now on the projection on $\mathbb{R}^{dM_S(t)}$ of any vector will be denoted $\cdot_S(t)$.

Theorem 1. *Assume that* **(H1)–(H5)** *are satisfied.*
For fixed $t \in [0,1]$, if $h \to 0$, $\eta \to 0$, $nh\eta^{dM_S(t)} \to +\infty$ and $n\eta^{dM_S(t)+4} \to 0$ a.s. *as $n \to +\infty$ then for all $z \in \mathbb{R}^\infty$, the point-wise mean square error writes*

$$\begin{aligned}\text{MSE}(t,z) \le C \Bigg[& \mathbb{E}\frac{\|\mathcal{H}\|_2^{2dM_S(t)}}{n\eta^{dM_S(t)}} + h^4 + \eta^4 + h^2\eta^2 + \mathbb{E}\frac{1}{nh\eta^{dM_S(t)}} \\ & + o\left(\mathbb{E}\frac{\|\mathcal{H}\|_2^{2dM_S(t)}}{n\eta^{dM_S(t)}}\right) + o(h^4) + o(\eta^4) + o(h^2\eta^2) + o\left(\mathbb{E}\frac{1}{nh\eta^{dM_S(t)}}\right)\Bigg]\end{aligned} \tag{2.1}$$

where C is a deterministic constant.

Note that we can get the consistency of our estimator under weaker assumptions as shown in the following proposition

Proposition 2. *Assume that* **(H1),(H3),(H4),(H5)** *are satisfied.*
For fixed $t \in [0,1]$, if $h \to 0$, $\eta \to 0$ and $nh\eta^{dM_S(t)} \to +\infty$ a.s. *as $n \to +\infty$ then for all $z \in \mathbb{R}^\infty$*

$$\hat{\theta}_{h,\eta}\left(t, z_S(t)\right) \xrightarrow{\mathbb{P}} \theta_S\left(t, z_S(t)\right).$$

Theorem 3. *Assume that* **(H1)–(H5)** *are satisfied.*
For fixed $t \in [0,1]$, if $nh\eta^{dM_S(t)} \to +\infty$, $n\eta^{dM_S(t)+2} \to 0$, and $nh^3\eta^{dM_S(t)} \to 0$ a.s. *as $n \to +\infty$ then for all $z \in \mathbb{R}^\infty$ such that $\theta_S\big(t, z_S(t)\big) \neq 0$*

$$\left(nh\eta^{dM_S(t)}\right)^{1/2} \frac{\hat{\theta}_{h,\eta}\big(t, z_S(t)\big) - \theta_S(t,z)}{\left[\hat{\theta}_{h,\eta}\big(t, z_S(t)\big) \int K^2 H^2 / \hat{f}_\eta\big(z_S(t)\big)\right]^{1/2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

where $\hat{f}_\eta\big(t, z_S(t)\big) = \dfrac{1}{n}\displaystyle\sum_{l=1}^n H_\eta\left(z - \vec{Z}_S^l(t)\right) \vee a_n.$

Bibliographie

[1] Asmussen, S. and H. Albrecher (2010). *Ruin probabilities.* Vol. 14. World Scientific Publishing Company.
[2] Bialek, W., F. Rieke, R. De Ruyter van Steveninck, and D. Warland (1991). *Reading a neural code.* Science 252.5014, pp. 1854–1857.
[3] Brette, Romain (2008). *Generation of correlated spike trains.* Neural computation 21.1, pp. 188–215.
[4] Kou, SC, X Sunney Xie, and Jun S Liu (2005). *Bayesian analysis of single-molecule experimental data.* Journal of the Royal Statistical Society: Series C (Applied Statistics) 54.3, pp. 469–506.
[5] Krumin, M. and S. Shoham (2009). *Generation of spike trains with controlled auto-and cross-correlation functions.* Neural Computation 21.6, pp. 1642–1664.
[6] Merton, R. C. (1976). *Option pricing when underlying stock returns are discontinuous.* Journal of financial economics 3.1, pp. 125–144.
[7] Ogata, Y. (1988). *Statistical models for earthquakes occurences and residual analysis for point processes.* Journal of the Royal Statistical Society. B 44, pp. 102–107.
[8] O'Sullivan, Finbarr (1993). *Nonparametric estimation in the Cox model.* The Annals of Statistics, pp. 124–145.