A NONPARAMETRIC TEST FOR COX PROCESSES

Benoît CADRE [1], Gaspar MASSIOT[2] and Lionel TRUQUET[3]

[1] *IRMAR, ENS Rennes, CNRS, UEB*
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
benoit.cadre@ens-rennes.fr
[2] *IRMAR, ENS Rennes, CNRS, UEB*
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
gaspar.massiot@univ-rennes1.fr
[3] *IRMAR, Ensai, CNRS, UEB*
Campus de Ker Lann
Rue Blaise Pascal - BP 37203, 35172 Bruz Cedex, France
lionel.truquet@ensai.fr

Résumé. Nous proposons un test statistique qui permet de mettre en vidence la nature Poissonienne d'un processus de Cox. La méthodologie employée utilise une comparaison d'estimateurs des fonctions de moyenne et de variance basés sur n copies du processus de Cox. Nous étudions les propriétés asymptotiques de la statistique de test ainsi définie.

Mots-clés. Processus de Cox, Statistique de test, Statistiques non paramt́riques, Martingale.

Abstract. We propose a test statistic to highlight the Poisson nature of a Cox process. Our methodology involves a comparison of estimators of the mean and variance functions based on n copies of the Cox process. We investigate the asymptotic properties of the test statistic.

Keywords. Cox Process, Test Statistic, Nonparametric statistics, Martingale Theory.

1 Introduction

Count process formulation is commonly used to describe and analyse many kind of data in sciences and engineering. A special class of such processes that researchers across in different fields frequently encounter are the so-called Cox processes or doubly stochastic processes. Compared to the standard Poisson process, the key feature of a Cox process is that its arrival rate is stochastic, depending on some covariate. In other words, if we let $N = (N_t)_{t\in\mathbb{R}_+}$ the Cox process and $\Lambda = (\Lambda(t))_{t\in\mathbb{R}_+}$ the (stochastic) cumulative arrival rate then, conditioning on Λ, the distribution of N is that of a Poisson process with cumulative intensity Λ. The benefit of randomness in the cumulative intensity lies in the fact that the statistician can take into account auxiliary informations, thus leading to a better model. For general references, we refer the reader to the monographies by Cox and Isham (1980), Karr (1991) and Kingman (1993).

In actuarial sciences and risk theory for instance, the number of claims in the risk model may be represented by a Cox process. In this area, the central quantity is the ruin probability, that is the probability that the surplus of the insurer is below zero at some time (see e.g., Björk and Grandell, 1998; Grandell, 1991; Schmidili, 1996). Cox process also appears in biophysics and physical chemistry (see e.g., Kou et al., 2005; Kou, 2008; Zhang and Kou, 2010). In these fields, experimental data consist of photon arrival times with the arrival rate depending on the stochastic dynamics of the system under study (for example, the active and inactive states of an enzyme can have different photon emission intensities); by analyzing the photon arrival data, one aims to learn the system's biological properties. Cox process data arise in neuroscience, to analyse the form of neural spike trains, defined as a chain of action potentials emitted by a single neuron over a period of time (see e.g., Gerstner and Kistler, 2002; Reynaud-Bourret et al., 2013). Finally mention astrophysics as another area where Cox process data often occur (see e.g., Scargle, 1998; Carroll and Ostlie, 2007).

In general, it is tempting to associate to a model numerous covariates, and this possibly wrongly. With this kind of abuse, one may use a Cox process model though a Poisson process model is satisfactory. In this paper, we elaborate a nonparametric test statistic to highlight the Poisson nature of a Cox process. More precisely, based on i.i.d. copies of N, we construct a nonparametric test statistic for $\mathbf{H_0}$: N is a Poisson process *vs* $\mathbf{H_1}$: N is not a Poisson process. Among the various possibilities to elaborate a test statistic devoted to this problem, one could estimate both functions $t \mapsto \mathbb{E}[N_t|\Lambda]$ and $t \mapsto \mathbb{E}N_t$ and test whether these functions are equal. However, this approach suffers from two main drawbacks, that is curse of dimensionality (whenever Λ takes values in a high-dimensional space) and knowledge a priori on Λ. Another approach is to test whether time-jumps of N are Poisson time-jumps; in this direction, we refer the reader to the paper by Reynaud-Bourret et al. (2014), in which a modified Kolmogorov-Smirnov statistic is used.

We elaborate and study a test statistic based on the observation that a Cox process is a Poisson process if, and only if its mean and variance function are equal (see beginning of Section 2). As we shall see, this approach leads to a very simple test, easily implementable and does not require to have any information on a covariate. Moreover, our test can be used to test if a counting process is Poissonian versus an overdispersed counting process (see Theorem 2.1, the power of our test is computed under a local alternative which does not use a Cox representation of the process). But as Cox processes form a natural class of overdispersed counting processes and the non dispersion of the countings characterizes the Poisson processes among the Cox processes, we keep this alternative hypothesis to formulate our test.

Our test can be seen as a functional version of the classical test of overdispersion used for testing the Poisson distribution of a sequence of countings. See for instance Rao and Chakravarti (1956) or Bhning (1994) for some overdispersion tests based on a comparison between the sample mean and the sample variance of counting sequences. Such overdispersion tests are widely used in actuarial science in the study of claims counts and as a goodness of fit test for Poisson regression models. See for instance Denuit et al. (2007)

2 Test statistic

Let $T > 0$ and $N = (N_t)_{t\in[0,T]}$ be a Cox process with stochastic cumulative intensity $\Lambda = (\Lambda(t))_{t\in[0,T]}$, and such that $\mathbb{E}N_T^4 < \infty$. We let m and σ^2 the mean and variance functions of N, i.e. for all $t \in [0,T]$:

$$m(t) = \mathbb{E}N_t \text{ and } \sigma^2(t) = \mathrm{var}(N_t).$$

Recall that for all $t \in [0,T]$ (see p. 66 in the book by Kingman, 2002) :

$$\sigma^2(t) = m(t) + \mathrm{var}(\mathbb{E}[N_t|\Lambda]).$$

Hence, $\sigma^2(t) \geq m(t)$ that is, each N_t is over-dispersed. Moreover, if $m = \sigma^2$, then $\mathbb{E}[N_t|\Lambda] = \mathbb{E}N_t$ for all $t \in [0,T]$, so that N is a Poisson process. As a consequence, N is a Poisson process if, and only if $m = \sigma^2$. This observation is the key for the construction of the test statistic, insofar the problem can be written as follows :

$$\mathbf{H_0} : \sigma^2 = m \text{ vs } \mathbf{H_1} : \exists t \leq T \text{ with } \sigma^2(t) > m(t).$$

From now on, we let the data $N^{(1)}, \cdots, N^{(n)}$ to be independent copies of N. By above, natural test statistics are based on the stochastic process $\hat{\sigma}^2 - \hat{m} = (\hat{\sigma}^2(t) - \hat{m}(t))_{t\leq T}$, where $\hat{m}$ and $\hat{\sigma}^2$ are the empirical counterparts of m and σ^2 :

$$\hat{m}(t) = \frac{1}{n}\sum_{i=1}^{n} N_t^{(i)} \text{ and } \hat{\sigma}^2(t) = \frac{1}{n-1}\sum_{i=1}^{n}\left(N_t^{(i)} - \hat{m}(t)\right)^2.$$

In the next result, convergence in distribution of stochastic processes is intended with respect to the Skorokhod topology (see Chapter VI in the book by Jacod and Shiryaev, 2003).

Theorem 2.1. *Let $B = (B_t)_{t\in\mathbb{R}_+}$ be a standard brownian motion on the real line. Under $\mathbf{H_0}$, $\hat{\sigma}^2 - \hat{m}$ is a martingale and*

$$\sqrt{n}\left(\hat{\sigma}^2 - \hat{m}\right) \xrightarrow{\text{(law)}} \left(B_{2m(t)^2}\right)_{t\leq T}.$$

Among the various possibilities of test statistics induced by $\hat{\sigma}^2 - \hat{m}$, we concentrate on

$$\hat{S}_1 = \sup_{t\leq T}\left(\hat{\sigma}^2(t) - \hat{m}(t)\right), \text{ and } \hat{S}_2 = \int_0^T \left(\hat{\sigma}^2(t) - \hat{m}(t)\right)\mathrm{d}t.$$

In particular, they are chosen so as to respect the unilateral nature of the problem (due to the fact that the alternative hypothesis may be written $\mathbf{H_1} : \sigma^2(t) > m(t)$ for some $t \leq T$).

We now present the asymptotic properties of $\hat{S}_1$ and $\hat{S}_2$.

Corollary 2.2. *Let* $\hat{M}^2(T) = \int_0^T (T-t)\hat{m}(t)^2 \mathrm{d}t$.

(i) Under $\mathbf{H_0}$*,*

$$\sqrt{n}\frac{\hat{S}_1}{\hat{m}(T)} \overset{\text{(law)}}{\longrightarrow} |\mathcal{N}(0,2)|, \textit{ and } \sqrt{n}\frac{\hat{S}_2}{\hat{M}(T)} \overset{\text{(law)}}{\longrightarrow} \mathcal{N}(0,4).$$

(ii) Under $\mathbf{H_1}$*,*

$$\sqrt{n}\frac{\hat{S}_1}{\hat{m}(T)} \overset{\text{prob.}}{\longrightarrow} +\infty, \textit{ and } \sqrt{n}\frac{\hat{S}_2}{\hat{M}(T)} \overset{\text{prob.}}{\longrightarrow} +\infty$$

Hence, the test statistics $\hat{S}_1/\hat{m}(T)$ and $\hat{S}_2/\hat{M}(T)$ define asymptotic tests with maximal power and the rejection regions for tests of level $\alpha \in]0,1[$ are

$$\left\{\frac{\hat{S}_1}{\hat{m}(T)} \geq \sqrt{\frac{2}{n}} q_{1-\alpha/2}\right\} \text{ and } \left\{\frac{\hat{S}_2}{\hat{M}(T)} \geq \frac{2}{\sqrt{n}} q_{1-\alpha}\right\}, \tag{2.1}$$

where for each $\beta \in]0,1[$, q_β is the $\mathcal{N}(0,1)$-quantile of order β.

Bibliographie

[1] Björk, T. and Grandell, J. (1988). Exponential inequalities for ruin probabilities in the Cox case, *Scandinavian Actuarial Journal*, 77-111.

[2] Carroll, B. and Ostlie, D. (2007). *An Introduction to Modern Astrophysics*, 2nd ed. Benjamin Cummings, Reading, MA.

[3] Cox, D.R. and Isham, V. (1980). *Point Processes*, Chapman and Hall, London.

[4] Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J. F. (2007). *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems.* John Wiley & Sons.

[5] Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge Univ. Press, Cambridge.

[6] Grandell, J. (1991). *Aspects of Risk Theory*, Springer-Verlag, New-York.

[7] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes, 2nd Ed.*, Springer, New-York.

[8] Karr, A.F. (1991). *Point Processes and their Statistical Applications, 2nd Ed.*, Marcel Dekker, New-York.

[9] Kingman, J.F.C. (1993). *Poisson Processes*, Oxford Studies in Probability, Oxford.

[10] Kou, S. C. (2008). Stochastic networks in nanoscale biophysics: Modeling enzymatic reaction of a single protein, *J. Amer. Statist. Assoc.*, 961?5.

[11] Kou, S. C., Xie, X. S. and Liu, J. S. (2005). Bayesian analysis of single-molecule experimental data (with discussion), *J. Roy. Statist. Soc. Ser. C*, 469?6.

[12] Rao, C. R., and Chakravarti, I. M. (1956). Some small sample tests of significance for a Poisson distribution. *Biometrics, 12*(3), 264-282.

[13] Reynaud-Bourret, P., Rivoirard, V., Grammont , F., Tuleau-Malot, C. (2014). Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis, *Journal of Mathematical Neuroscience*, 4:3.
[14] Scargle, J. D. (1998). Studies in astronomical time series analysis. V. Bayesian blocks, a new method to analyze structure in photon counting data. *Astrophys. J.*, 405?8.
[15] Schmidili, H. (1996). Lundberg inequalities for a Cox model with a piecewise constant intensity, *J. Applied Probability*, 196-210.
[16] Zhang, T. and Kou, S.C. (2010). Nonparametric inference of doubly stochastic Poisson process data via the kernel method, *Ann. Applied Statist.*, 1913-1941.