Journal of Statistical Planning and Inference 184 (2017) 48–61

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Nonparametric tests for Cox processes

Benoît Cadre [a,*], Gaspar Massiot [b], Lionel Truquet [b]

[a] IRMAR, ENS Rennes, CNRS, Campus de Ker Lann, Avenue Robert Schuman, 35170 Bruz, France
[b] IRMAR, Ensai, CNRS, Campus de Ker Lann, Avenue Robert Schuman, 35170 Bruz, France

ARTICLE INFO

Article history:
Received 7 March 2016
Received in revised form 27 September 2016
Accepted 4 December 2016
Available online 13 December 2016

MSC:
62C12
62M07
60G44

Keywords:
Functional statistic
Cox process
Test statistic
Local alternative
Nonparametric statistics

ABSTRACT

In a functional setting, we elaborate and study two test statistics to highlight the Poisson nature of a Cox process when n copies of the process are available. Our approach involves a comparison of the empirical mean and the empirical variance of the functional data and can be seen as an extended version of a classical overdispersion test for count data. The limiting distributions of our statistics are derived using a functional central limit theorem for càdlàg martingales. Our procedures are easily implementable and do not require any knowledge on the covariate. We address a theoretical comparison of the asymptotic power of our tests under some local alternatives. A numerical study reveals the good performances of the method. We also present two applications of our tests to real datasets.

1. Introduction

Count process formulation is commonly used to describe and analyze many kinds of data in sciences and engineering. A widely used class of such processes is the family of the so-called Cox processes or doubly stochastic Poisson 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 $T > 0$ denotes the observation period, $N = (N_t)_{t\in[0,T]}$ the Cox process and $\Lambda = (\Lambda(t))_{t\in[0,T]}$ the (stochastic) cumulative arrival rate then, conditioning on Λ, the distribution of N is that of a Poisson process with cumulative intensity Λ. Using randomness in the cumulative intensity, the statistician can take into account some auxiliary information, thus leading to a more realistic model. For general references, we refer the reader to the books by Cox and Isham (1980), Karr (1991), Kingman (1993) or Snyder and Miller (1991).

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, i.e. the probability that the surplus of the insurer is negative at some time (see e.g. Björk and Grandell, 1988; 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 analyze the form of

* Corresponding author.
E-mail addresses: benoit.cadre@ens-rennes.fr (B. Cadre), massiot@ensai.fr (G. Massiot), truquet@ensai.fr (L. Truquet).

http://dx.doi.org/10.1016/j.jspi.2016.12.001
0378-3758/

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., 2014). Finally, let us 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 incorporate abusively numerous covariates in the statistical model, though a Poisson process model might be satisfactory. In this paper, we elaborate two nonparametric test statistics to highlight the Poisson nature of a Cox process. More precisely, based on i.i.d. copies of N, we construct two nonparametric test statistics for $\mathbf{H_0}$: N is a Poisson process vs $\mathbf{H_1}$: N is not a Poisson process. This setting of i.i.d. copies of N is justified by the fact that in many situations, the duration of observation is limited but the number of observed paths is large.

Among the various possibilities for constructing a test statistic devoted to this problem, a naive approach consists in first estimating both functions $t \mapsto \mathbb{E}[N_t|\Lambda]$ and $t \mapsto \mathbb{E}N_t$ and then testing whether these functions are equal. However, this approach suffers from two main drawbacks: the curse of dimensionality (whenever Λ takes values in a high-dimensional space) and the knowledge a priori on Λ. Another approach is to test whether the 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.

In this paper, we elaborate and study two test statistics, both based on the observation that a Cox process is a Poisson process if, and only if its mean and variance functions are equal. As we shall see, this approach leads to very simple and easily implementable tests.

The paper is organized as follows. In Section 2, we first present the test statistics, then we establish their asymptotic properties. We also compare the asymptotic properties of the tests by using a local alternative. Section 3 is devoted to a simulation study. An application to real data is presented in Section 4. The proofs of our results are postponed to the three last sections of the paper.

2. Tests for Cox processes

2.1. Principle of the tests

Throughout the paper, $T > 0$ is the (deterministic) duration of observation, and $N = (N_t)_{t\in[0,T]}$ is a Cox process with cumulative intensity process $\Lambda = (\Lambda(t))_{t\in[0,T]}$, such that the fourth moment of N_T is finite, i.e. $\mathbb{E}N_T^4 < \infty$, and $\mathbb{E}N_t \neq 0$ for some $t \in]0, T[$. Note that the mean function $t \mapsto \mathbb{E}N_t = \mathbb{E}\Lambda(t)$ might not be absolutely continuous, so it is not necessarily the integral of an intensity.

We let m and σ^2 the mean and variance functions of N, i.e. for all $t \in [0, T]$:

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

Recall that for all $t \in [0, T]$ (see p. 66 in the book by Kingman, 1993):

$$\sigma^2(t) = m(t) + \text{var}(\mathbb{E}[N_t|\Lambda]) = m(t) + \text{var}\big(\Lambda(t)\big). \tag{2.1}$$

Hence, $\sigma^2(t) \geq m(t)$ that is, each N_t is overdispersed. Moreover, if $m = \sigma^2$, then $\mathbb{E}[N_t|\Lambda] = \mathbb{E}N_t$ for all $t \in [0, T]$, thus 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 feature for the construction of our test statistics. With this respect, the problem can be written as follows:

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

From now on, we let the data $N^{(1)}, \ldots, N^{(n)}$ to be independent copies of N. By above, natural test statistics are based on the process $\hat{\sigma}^2 - \hat{m} = (\hat{\sigma}^2(t) - \hat{m}(t))_{t\in[0,T]}$, where $\hat{m}$ and $\hat{\sigma}^2$ are the empirical counterparts of m and σ^2:

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

In this paper, 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).

Our first main result gives the asymptotic distribution of the process $\hat{\sigma}^2 - \hat{m}$ under $\mathbf{H_0}$ (see Section 5 for the proof).

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*

$$\sqrt{n}\,\big(\hat{\sigma}^2 - \hat{m}\big) \xrightarrow{\text{(law)}} \big(B_{2m(t)^2}\big)_{t\leq T}. \quad \square$$

As far as we know, the martingale property for $\hat{\sigma}^2 - \hat{m}$ has not been observed yet. This property, which is interesting by itself, plays a crucial role in the derivation of the asymptotic result.