Journal of Statistical Planning and Inference 198 (2019) 119–138

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Estimation of the average number of continuous crossings for non-stationary non-diffusion processes

Romain Azaïs [a], Alexandre Genadot [b,*]

[a] Inria Nancy –Grand Est, Team BIGS and Institut Élie Cartan de Lorraine, Nancy, France
[b] Université de Bordeaux, IMB, Institut Mathématiques de Bordeaux, INRIA Bordeaux Sud-Ouest, Team: CQFD, 351 cours de la Libération, 33405 Talence Cedex, France

A R T I C L E I N F O

Article history:
Received 31 March 2017
Received in revised form 12 April 2018
Accepted 13 April 2018
Available online 27 April 2018

MSC:
62M05
93C30
62G05

Keywords:
Piecewise deterministic process
Average number of crossings
Plug-in estimators

A B S T R A C T

Assume that you observe trajectories of a non-diffusive non-stationary process and that you are interested in the average number of times where the process crosses some threshold (in dimension $d = 1$) or hypersurface (in dimension $d \geq 2$). Of course, you can actually estimate this quantity by its empirical version counting the number of observed crossings. But is there a better way? In this paper, for a wide class of piecewise smooth processes, we propose estimators of the average number of continuous crossings of a hypersurface based on Kac–Rice formulae. We revisit these formulae in the uni- and multivariate framework in order to be able to handle non-stationary processes. Our statistical method is tested on both simulated and real data.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

We consider a random but non-diffusive trajectory X that models some physical or biological phenomenon. In order to set the ideas down, X may be issued from a piecewise deterministic Markov process (PDMP) (Davis, 1993) but our theoretical results will be established for a more general class of stochastic models. In this paper, we are not interested in the estimation of the parameters that govern the underlying model that has generated the trajectory X but in a functional of this trajectory. Indeed, in numerous situations, the main quantity of interest is related to the average number of crossings of some level (in dimension 1) or hypersurface (in dimension $d \geq 2$) within a given time window. For instance, if X models the cumulative exposure to some food contaminant (Bertail et al., 2008), the toxic effects depend on the time spent beyond a critical threshold. In reliability, X may describe the size of a crack in a certain material (Abdessalem et al., 2016). Exceeding a dangerosity threshold may lead to rupture and thus to a dramatic event. Unfortunately, the trajectory X is often observed on a discrete temporal grid which time step size is imposed by the measuring devices. In this context, some crossings may be missed in such a way that the crude Monte Carlo-type estimator that only counts the number of observed crossings generally underestimates the theoretical quantity of interest. However, modern datasets often contain more information than the location at the measure time: they may provide the instantaneous velocity. This is typically the case when one studies spatial trajectories captured by a GPS device like the terrestrial and marine movements of lesser black-backed gulls investigated in Garthe et al. (2016a). This additional information is crucial because it may give a clue on the likelihood of an

* Corresponding author.
E-mail address: alexandre.genadot@u-bordeaux.fr (A. Genadot).

https://doi.org/10.1016/j.jspi.2018.04.002
0378-3758/© 2018 Elsevier B.V. All rights reserved.

unobserved crossing between two successive measure times. The main objective of this article is to show that one can take into account the velocity measurement to better estimate the average number of crossings.

In the present article, we deal with piecewise smooth processes (PSP's) as defined in Borovkov and Last (2012). They form a very general class of non-diffusion stochastic processes composed of deterministic trajectories following some differential equation punctuated by random jumps at random times. PDMP's are a particular case of PSP's because of the particular link between the inter-jumping times and the deterministic path that ensures the Markov property to hold. It should be noted that we do not impose any kind of Markov assumption in this paper. Kac–Rice formulae give a concise relation between the average number $C_S(H)$ of crossings of X with a hypersurface S within the time window $[0, H]$ and some features of the underlying stochastic model. They have already been stated for non-stationary one-dimensional PSP's (Dalmao and Mordecki, 2015) and for stationary multidimensional PSP's (Borovkov and Last, 2012). These papers (Borovkov and Last, 2012; Dalmao and Mordecki, 2015) and our approach are different and complementary. In the applications (typically spatial trajectories captured by GPS), the stochastic process of interest X is often both multidimensional and non-stationary. In Theorem 2.10, we establish under mild conditions the following Kac–Rice formula for multidimensional non-stationary PSP's,

$$C_S(H) = \int_S |(r(x), \nu(x))| \int_0^H p_{X(s)}(x)\mathrm{d}s\, \sigma_{d-1}(\mathrm{d}x), \tag{1}$$

where r is the velocity of the deterministic motion, ν is a field of unit normals of S, $p_{X(t)}$ denotes the density of $X(t)$ and σ_{d-1} stands for the Hausdorff measure (see Remark 2.2). The strategy developed to state (1) also gives a fresh look at Kac–Rice formula for one-dimensional processes. Consequently, we also provide a new proof of this formula for one-dimensional PSP's in Corollary 2.7. In addition, we investigate in Corollaries 2.8 and 2.12 the Euclidean-mode setting often used in applications of hybrid Markov models (for instance in Abdessalem et al., 2016), which has never been studied from that perspective in the literature to the best of our knowledge.

The distribution $p_{X(t)}$ appearing in the Kac–Rice formula (1) is generally unknown but can be estimated (for example by kernel methods) from a dataset of trajectories observed within the time window $[0, H]$. In a wide range of applications, the deterministic motion is assumed to be known because it has been postulated by scientific laws, in particular in physical or biological models. This allows us to propose a new strategy for estimating the average number of crossings $C_S(H)$. If $\widehat{p}_{X(t)}$ denotes an estimate of $p_{X(t)}$, the number of crossings $C_S(H)$ can be approximated by the plug-in estimator

$$\widehat{C}_S(H) = \int_S |(r(x), \nu(x))| \int_0^H \widehat{p}_{X(s)}(x)\mathrm{d}s\, \sigma_{d-1}(\mathrm{d}x).$$

In the simulation study presented in Section 4, we show that $\widehat{C}_S(H)$ better estimates $C_S(H)$ than the Monte Carlo estimator that only returns the empirical mean of observed crossings within the interval $[0, H]$, in particular when the time step size is large. In the real data application given in Section 5, we investigate the spatial trajectories of lesser black-backed gulls studied in Garthe et al. (2016a). We do not assume any model for the velocity r but we directly estimate the scalar product $|(r(x), \nu(x))|$ appearing in (1) from instantaneous velocity measurements provided in the dataset (Garthe et al., 2016b). Estimated Kac–Rice formulae allow us to approximate the average depth of marine and terrestrial trips of the birds within a one-day window, which helps to describe their daily habits.

The paper is organized as follows. Section 2 is devoted to Kac–Rice formulae for non-stationary PSP's. The theoretical framework is presented in Section 2.1, while results for one-dimensional (multidimensional, respectively) processes are given in Section 2.2 (Section 2.3, respectively). The proofs of the results stated in Section 2 have been deferred until Appendix A. Section 3 is dedicated to the statistical inference procedures. The proofs of the estimation results established in this section have been deferred until Appendix B. A simulation study on PDMP's is provided in Section 4 through three examples: stationary one-dimensional telegraph process in Section 4.1, piecewise deterministic simulated annealing in Section 4.2 and non-stationary two-dimensional telegraph process in Section 4.3. Finally, real data experiments are investigated in Section 5.

2. Kac–Rice formulae for piecewise smooth processes

2.1. Definitions and background material

2.1.1. A class of piecewise smooth processes

A piecewise smooth process (PSP) X on an open domain $\mathcal{X}$ of $\mathbb{R}^d$, endowed with the Euclidean norm $\|\cdot\|$ and associated scalar product $(\cdot,\cdot)$, involves continuous-time deterministic motions punctuated by random jumps at random times. Its dynamic is governed by a marked point process $(T_n, M_n)_{n\geq 0}$ on $\mathbb{R}_+ \times \mathcal{X}$ and a vector field $r : \mathcal{X} \to \mathbb{R}^d$, the T_n's being the jump times of the continuous-time trajectory. The sequence of the jump times is assumed to be almost-surely increasing, i.e., $\mathbb{P}(T_0 = 0 < T_1 \leq T_2 \leq \cdots) = 1$, and almost-surely going to infinity, i.e., $\mathbb{P}(\lim_{n\to\infty} T_n = \infty) = 1$. It should be noticed that the distribution of the T_i's can depend on the marks. The following assumption on the jump times is quite classical in the context of hybrid models (see Davis, 1993, Standard conditions (24.8) and Remark (24.9)).

Assumption 2.1. The expected number of jump times over any bounded interval is finite.

Download English Version:

https://daneshyari.com/en/article/7547046

Download Persian Version:

https://daneshyari.com/article/7547046

Daneshyari.com