Statistics and Probability Letters 138 (2018) 9–19

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Optimal bandwidth selection in kernel density estimation for continuous time dependent processes

Khadijetou El Heda [a], Djamal Louani [b,*]

[a] LMPA, Université du Littoral Côte d'Opale, Calais, France
[b] UFR Sciences Exactes et Naturelles, Université de Reims, France

ARTICLE INFO

Article history:
Received 29 May 2017
Received in revised form 27 December 2017
Accepted 2 February 2018
Available online 21 February 2018

MSC:
60-G10
62-G07
47-A35

Keywords:
Stationary process
Continuous time process
Density estimation
Ergodicity
Kernel estimator
Bandwidth

ABSTRACT

The choice of the smoothing parameter in nonparametric function estimation is of major concern. The estimation accuracy highly depends on how such a choice is performed. In this paper, we construct a bandwidth selection procedure pertaining to the kernel density estimation when a continuous time dependent and stationary process is considered.

1. Introduction

Let $(X_t, \quad 0 \leq t \leq T), T \in \mathbb{R}^+$ be a continuous time stationary and ergodic process with a marginal density function f. Consider the kernel estimator of f defined, for any $x \in \mathbb{R}$, by

$$f_{T,h}(x) = \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) dt,$$

where h is the smoothing parameter belonging to $H_T := [a_T, b_T]$ where $[a_T, b_T] \subset \mathbb{R}_+$. Here, K stands as a positive measurable function integrating to one. It is well-known that large values of the bandwidth parameter oversmooth the density estimators while small values undersmooth the curves. In this paper, we aim at investigating the optimal smoothing parameter choice using the well-known usual cross-validation criterion.

The optimal window width choice in nonparametric estimation framework has motivated a number of studies throughout the literature. We refer first to the work of Stone (1984) who considered the topic in the discrete time case. He described the window selection rule which yields to an asymptotic optimal choice under the assumption that the marginal density is bounded whenever the cross-validation is used. Notice that this criterion has been introduced by Rudemo (1982) and Bowman (1984). For the same needs, the regression function estimation is considered by Härdle and Marron (1985) with

* Corresponding author.
E-mail address: djamal.louani@upmc.fr (D. Louani).

https://doi.org/10.1016/j.spl.2018.02.001
0167-7152/

the same criterion. There exist several automatic data driven selection rules to handle the problem of the optimal smoothing parameter choice. The so-called plug-in methods minimize both the mean integrated square error (MISE) and the mean square error (MSE). See Silverman (1986), Heidenreich et al. (2013) and Hall and Marron (1987) for more details. Chacón et al. (2007) showed that there exists a minimizer of the exact MISE of the kernel density estimator as a function of the bandwidth and gave the limit properties of the optimal parameter. A further work due to Chacón and Tenreiro (2012) gives properties of the exact optimal bandwidth which minimizes the MSE of the kernel density estimator. Tenreiro (2017) presents in his work a modified version of least square cross-validation by introducing some weights (WLSCV). Simulation results show that the WLSCV method performs better than the standard one for both "easy-to-estimate" and "hard-to-estimate" density cases. At this time, all the works referred to consider independent random samples.

In the discrete time dependent data case, Hart and Vieu (1990) established under a set of more restrictive conditions than the Stone's single one, the asymptotic optimality of the window width pertaining to the cross-validation criterion using the ISE. Kim and Denis (1997) established the asymptotic optimal properties under nearly mild conditions as in Stone's case. The hazard rate function kernel estimator has been investigated by Youndjé et al. (1996). They highlighted the role of the bandwidth in the ISE and established the optimal asymptotic properties in the window selection procedure using the cross-validation criterion. In the functional data framework, Rachdi and Vieu (2007) considered the regression function estimate to construct an asymptotic optimal smoothing parameter for the cross-validation criterion.

Commonly, the criterion used to measure loss is the MISE. The window h_{opt} which minimizes the MISE is taken to be nonrandom. Since the value of the "optimal" bandwidth must in practice be a random variable, it is more natural to minimize the ISE instead of the MISE. In order to be more precise on the matter, consider the mean integrated square error and the integrated square error associated to $f_{T,h}$ defined respectively by

$$M_h = \mathbb{E}\int [f_{Th}(x) - f(x)]^2 dx$$

and

$$L_{T,h} = \int [f_{Th}(x) - f(x)]^2 dx = \int f_{Th}^2(x)dx - 2\int f_{Th}(x)f(x)dx + \int f^2(x)dx.$$

Notice that the minimization of $L_{T,h}$ with respect to h is obtained while minimizing the quantity

$$L_{T,h} - \int f^2(x)dx = \int f_{Th}^2(x)dx - 2\int f_{Th}(x)f(x)dx.$$

Since $L_{T,h} - \int f^2$ depends on the unknown function f, the optimal bandwidth obtained by such a procedure still depends on f. We have then to build a procedure avoiding to deal with unknown quantities in practice. In this respect, we estimate the function f and introduce the cross-validation criterion defined as

$$M_{T,h} = \int f_{Th}^2(x)dx - \frac{2}{T^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int_{T_{i-1}}^{T_i}\int_{T_{j-1}}^{T_j} K_h(X_s - X_t)dsdt,$$

where, for some $n \in \mathbb{N}$, $\delta = \frac{T}{n}$ and $T_j = j\delta$, $1 \leq j \leq n$. The optimal window selection rule is performed then by minimizing $M_{T,h}$ with respect to h.

Remark 1. Similarly to the discrete case, introduce the quantity

$$f_{T,-i}(x) := \frac{1}{(n-1)\delta}\sum_{\substack{j=1\\ j\neq i}}^{n}\int_{T_{j-1}}^{T_j} K_h(x - X_t)dt,$$

that stands as the kernel estimate of f while the part $\{X_t : t \in [T_{i-1}, T_i]\}$ of the process is leaved out. An alternative asymptotic equivalent cross-validation criterion is given by

$$\frac{1}{n}\sum_{i=1}^{n}\int f_{T,-i}^2(x)dx - \frac{2}{T}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{T,-i}(X_s)ds.$$

2. Results

Introduce now the assumptions necessary to establish our results gathered together here for easy reference. Let $K^{(2)}$ be the convolution product of K with itself. Notice that $K^{(2)}$ has the same properties as K. Whenever f_{X_s,X_t} stands as the join density of the vector (X_s, X_t) and f_{X_s} is the marginal density of X_s, set from now on

$$g_{s,t} := f_{X_s,X_t} - f_{X_s}f_{X_t}, \quad g_{0,u} := g_u, \quad \|.\|_\infty := \sup_{(y,z)\in\mathbb{R}^2} |.|.$$