SEMI-RECURSIVE KERNEL CONDITIONAL DENSITY ESTIMATORS UNDER RANDOM CENSORSHIP AND DEPENDENT DATA

Sihem Semmar,Ali Laksaci, Salah Khardani

Djilali Liabes University .
Fac. of Science, Algeria , doc.semmar@gmail.com
King Khalid University, Abha, Kingdom of Saudi Arabia.
Ecole Nationale des Sciences et Technologies , Tunisia

ABSTRACT

In this work, we extend to the case of the strong mixing data the results of Khardani and Semmar. A kernel−type recursive estimator of the conditional density function is introduced. We study the properties of these estimators and compare them with Rosemblatt's nonrecursive estimator. Then, a strong consistency rate as well as the asymptotic distribution of the estimator are established under an α-mixing condition. A simulation study is considered to show the performance of the proposed estimator.
Keywords : Asymptotic normality, censored data, conditional density, mixing sequences, recursive kernel estimators, survival data, strong consistency.

1. INTRODUCTION

It is well known that censored data are, usually, used to describe many practical applications in a wide variety of fields, including economics, medicine, biology, and biostatistics. For example, let T be the survival time of individuals who are involved in a clinical study or a possible monotone transformation of the survival time. Consider a random covariable X, such as the age, the dose of a drug or the cholesterol level. As often occurs in practice, the response T is subject to random right censoring. In other words, instead of observing T, one observes the pair of variables (Y,δ), where $Y = \min(T,C), \delta = 1_{T\leq C}$, and C represents the censoring time, which is expected to be conditionally independent of T given X.
In this article we focus on the nonparametric estimation of the conditional density function of response censored variable T given a vectorial variable X. We estimate this model by a recursive kernel approach and we derive its asymptotic properties when the observations are dependent. We note that this subject of the conditional density estimation plays an important role not only in the prediction of response variable given a regressor but it is a fundamental tool to explore all relationships between responses and covariates.

The main purpose of this contribution is to study the recursive estimator of the conditional density when the response variable is subject to random right censoring and the observations are $\alpha-$ mixing process. We establish the almost complete consistency as well as the asymptotic normality of the constructed estimator. All these asymptotic results are obtained with precision of the convergence rate and are established under some standard conditions. It should be noted that the present contribution generalize to the dependent case the results of Khardani and Semmar (2014) in the iid case. Recall that this generalization to time series case is not a simple extension but it requires some alternative and additional tools. We point out the density estimation in censored time series data has been studied by Khardani, Lemdani, and Ould Saïd (2010, 2011) by using

the classical kernel method. Alternatively in the present work, we use the recursive approach and we treat the conditional case. The benefit of this consideration is double : First, the recursivity offers, rapid, accurate and robust estimator. Second, conditioning allows to provide predictors with more specification. From practical point of view, our approach is very useful in big data analysis which has a lot of interest in the last years. In particular the Internet network provide massive data arriving in streams (including Twitter activity, the Facebook news stream, Internet packet data, stock market activity, credit card transactions and Internet and phone usage), and if they are not processed immediately or stored, then they are lost forever. Thus, data streams have become an increasingly important filed of research of statistical analysis. In this context, building a semi-recursive estimator which does not require to store all the data in memory and can be updated easily in order to deal with online data is of great interest. Indeed, the recursive estimator can be updated with each new observationX_{n+1} which permits to save computational time and storage memory, unlike to the non-recursive case where the estimator needs to be recalculated completely when a new data set is observed. This gain is more important in the conditional density estimation, since the number of points at which the function is estimated is usually very large.

2. RECURSIVE KERNEL CONDITIONAL DENSITY UNDER A RANDOM CENSORSHIP

Let $(X_i,T_i)1$ be a $\mathbb{R}^d \times \mathbb{R}$ valued stationary strongly mixing or α-mixing process defined on probability space (Ω,A,P) : The object of this article is to study the conditional density of T_i given $X_i = x$ which is obtained by

$$\phi(t|x) = \frac{g(x,t)}{\ell(x)} \tag{1}$$

where $g(.,.)$ denotes the joint density of (T_i,X_i) and $\ell(.)$ denotes the marginal densityof X_i. Typically, we consider the case of random right censorship. Formally we suppose that the $(T_i)_{1\leq i\leq n}$ generated form a stationary $\alpha-$ mixing sequence with unknown continuous distribution function (df) F and density g. In this situation of censored lifetimes model we assume that there exist a sample of independent and identically distributed (iid) censoring random variables (rvs) $(C_i)_{1\leq i\leq n}$ with continuous df G, such that T_i is observed only if $T_i \leq C_i$: Then we proceed, with the n pairs Y_i, δ_i with

$$Y_i = T_i \wedge C_i \quad and \quad \delta_i = 1_{T_i\leq C_i\}} \quad 1 \leq i \leq n. \tag{2}$$

where 1_A denotes the indicator function of the set A. To follow the convention in biomedical studies, we assume that $(C_i)_{1\leq i\leq n}$ and $(T_i,X_i)_{1\leq i\leq n}$ are independent; this condition is plausible whenever the censoring is independent of the modality of the patients.
From the observations $(X_1,Y_1,\delta_1),\cdots,(X_n,Y_n,\delta_n) \quad of \quad (X,Y,\delta)$, the kernel estimate of the conditional density $\phi(t|x)$ denoted $\bar{\phi}_n(t|x)$ is defined by

$$\bar{\phi}_n(t|x) = \frac{\sum_{i=1}^{n} h_n^{-1}\delta_i \bar{G}_n^{-1}(Y_i)K\left(\frac{x-X_i}{h_n}\right)L\left(\frac{t-Y_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)} = \frac{\bar{g}_n(x,t)}{\ell_n(x)} \tag{3}$$

where K,L are kernels and h_n is a sequence of positive real numbers.

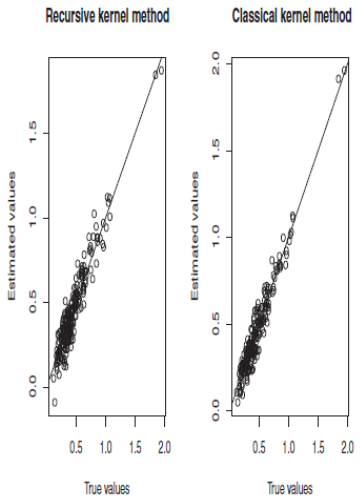

FIGURE 1 – .Recursive and non-recursive estimators for a standard normal distribution case.

A recursive version of the previous kernel estimator is defined by

$$\begin{aligned}\widehat{\phi}_n(t|x) &= \frac{\sum_{i=1}^{n} h_i^{-(d+1)} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x-X_i}{h_i}\right) L\left(\frac{t-Y_i}{h_i}\right)}{\sum_{i=1}^{n} h_i^{-d} K\left(\frac{x-X_i}{h_i}\right)} \\ &= \frac{\hat{g}_n(x,t)}{\ell_n(x)}\end{aligned}$$

where

$$\hat{g}_n(x,t) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x-X_i}{h_i}\right) L\left(\frac{t-Y_i}{h_i}\right),$$

and

$$\ell_n(x) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d}} K\left(\frac{x-X_i}{h_i}\right), \ \forall x \in \mathbb{R}^d.$$

3. SIMULATIONS RESULTS

In this section, we discuss the feasibility and the performance of the recursive nonparametric method through an empirical study. Precisely, our main purpose is to compare the ratio (efficiency/rapidity) between recursive and classical kernel conditional density estimation in the time series case.

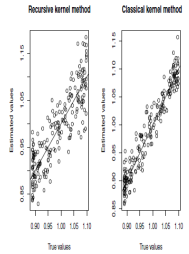

FIGURE 2 – .Recursive and non-recursive estimators for a skewed unimodal distribution case.

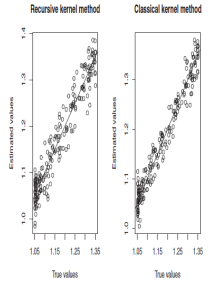

FIGURE 3 – .Recursive and non-recursive estimators for a skewed bimodal distribution .

4. THE BIBLIOGRAPHY

Ahmad, I., and P. E. Lin. 1976. Nonparametric sequential estimation of a multiple regression function.*Bulletin of Mathematical Statistics* 17 :63–75.

Amiri, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, *arXiv :1211.5767v2.*

Beran, R. (1981). Nonparametric regression with randomly censored survival data, *Technical university of Clifornia, Berkeley*.

Bollerslev, T. 1986. General autoregressive conditional heteroskedasticity.*Journal of Econometrics* 31 (3) :307–27.

Bosq, D. 1998. Nonparametric statistics for stochastic processes. *In Lecture notes in statistics*, Vol. 149. New York : Springer.

Bradley, R. C. 2007. Introduction to strong mixing conditions, Vol. I–III, Utah : Kendrick Press. Carbonez, A., L. Gyorfi, and E. C. Van der Meulen. 1995. Partitioning estimates of a regression function under random censoring.*Statistics Decisions* 13 :21–37.

Khardani, S., Lemdani, M., Ould Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship, *J. of the Korean Statistical Society*, 39, 455–469.

Khardani, S., Lemdani, M., Ould Saïd, E.(2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series, *J. Stat. Plann. Inference*, 141, 3426–3436.

Khardani, S., and S. Semmar. 2014. Nonparametric conditional density estimation for censored data based on a recursive kernel.*Electronic Journal of Statistics* 8 (2) :2541–56.

Khardani, S., and Y. Slaoui. 2019. Nonparametric relative regression under random censorship model.*Statistics and Probability Letters* 151 :116–22.

Khardani, S., and Y. Slaoui. 2019. Recursive kernel density estimation and optimal bandwidth selection under a- mixing data.*Journal of Statistical Theory and Practice* 13 (2) :13–36.

Khardani, S., M. Lemdani, and E. Ould Saïd. 2010.Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship.*Journal of the Korean Statistical Society* 39

Khardani, S., M. Lemdani, and E. Ould Saïd. 2011.Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series. *Journal of Statistical Planning and Inference* 141.

Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions,*Stochastic Process. Appl*, 36 (1), 107–116.

Walk, H. 2010. Strong laws of large numbers and nonparametric estimation. In *Recent developments in applied probability and statistics*, eds. L. Devroye, B. Karas€ozen, M. Kohler and R. Korn, 183–214. Heidelberg : Springer Physica,.

Wang, L., and H. Y. Liang. 2004. Strong uniform convergence of the recursive regression estimator under ϕ mixing conditions.*Metrika* 59 (3) :245–61.

Wegman, J., and I. Davies. 1979. Remarks on some recursive estimators of a probability density. *The Annals of Statistics* 7 (2) :316–27.

Wegman, J. Davies, I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Statist*, 7, 316–327.

Wolverton, C. and Wagner, T.J. (1969). Asymptotically optimal discriminant functions for pattern classifcation, *IEEE Trans. Inform. Theory*, 15, 258–265.