Nonparametric conditional density estimation for censored data based on a recursive kernel

Salah Khardani[*] and Sihem Semmar

École Nationale d'Ingénieurs de Monastir (ENIM), Laboratoire de Physique-Mathématique, Fonctions Spéciales et Applications (P.M.F.S.A) Hammam-Sousse Université de Djilali Liabes. Fac. de Science
e-mail: khardani_salah@yahoo.fr; sihem.semmar@yahoo.fr

Abstract: Consider a regression model in which the response is subject to random right censoring. The main goal of this paper concerns the kernel estimation of the conditional density function in the case of censored interest variable. We employ a recursive version of the Nadaraya-Watson estimator in this context. The uniform strong consistency of the recursive kernel conditional density estimator is derived. Also, we prove the asymptotic normality of this estimator.

AMS 2000 subject classifications: 62G05, 62G07, 62G08, 62G20, 62H12.
Keywords and phrases: Asymptotic normality, censored data, conditional density, kernel estimator, recursive estimation, Kaplan–Meier estimator, uniform almost sure convergence.

Contents

1. Introduction

Studying the relationship between a response variable and a explanatory variable is one of the most important statistical analysis. Usually this relationship is modeled with the regression function. However, it is well known, this non-

---

[*]Corresponding author.

parametric model is not efficient in some pathological situations. For instance, the multi-modal densities case, the case where the expected value might be nowhere near a mode or for situations in which confidence intervals are preferred to point estimates. In all these case the conditional density is a pertinent model to explore this relationship. The main purpose of this paper is to study this nonparametric model when the response variable is subject to censoring, by using a kernel recursive estimation method.

Noting that the nonparametric modeling of censored data is intensively discussed in the recent statistical literature. It dates back to [5], who introduced a class of nonparametric regression estimators for the conditional survival function in the presence of right-censoring. [8, 9] studied the asymptotic properties of the distribution and quantiles functions estimators. [17] gave a simpler proof in the randomly right-censoring case for kernel, nearest neighbor, least squares and penalized least squares estimates. Further results was obtained by [15, 16]. Concerning the nonparametric conditional model, we cite for the conditional model in both (iid and mixing case) and for conditional quantiles function. In this vast variety of papers, the authors use the Nadaraya-Watson techniques as estimation method which is a particular case of the recursive kernel estimator considered in this paper. Moreover, this last has various advantages over the kernel method. We deal with recursive kernel estimators where, by recursive we mean that the estimator calculated from the first n observations, say f_{n+1}, is only a function of f_n and the $(n+1)^{-th}$ observation. As is well known, the recursive property is particularly interesting when the sample data are obtained by mean of some observational mechanism that allows an increase in the sample size over time. This situation is usual in many control and supervision problems and, above all, in time series analysis. In the above cases, the recursive estimates allow us to update the estimations as additional observations are obtained, unlike non-recursive methods where estimates must be completely recalculated when each additional item of data received. From a practical point of view, this iterative procedure provides an important saving in computational time and memory, since the updating of the estimates is independent of the previous sample size. It is not the case for the basic kernel estimator which has to be computed again on the whole sample. Recursive estimators show good theoretical properties, from the point of view of mean square error (small variance) and almost sure convergence.

The first recursive modifications of the Nadaraya-Watson estimate have been introduced by [1] and [12] say (AL) and (DW). In complete data, Kernel recursive estimators have been introduced by [27] and [28]. Next [10, 11, 23, 26] have independently studied the rates of the almost sure convergence of particular recursive density estimates.

The law of the iterated logarithm of the recursive density estimator was established by [25] and [23]. For other works on recursive density estimation, the reader is referred to the papers of [1] and [6]. Recently, in a context of α-mixing processes, [2] gave the exact asymptotic quadratic error of a general family of kernel estimator, whose (AL) and (DW) are particular cases. The asymptotic normality of the same family is obtained by [4].

The recursive regression estimator for identically independent distributed (i.i.d.) random variables has been studied by many authors among whom we quote [1, 18] and [24] for a nonparametric approach and [21] for semi-parametric models. In the strong mixing case, [22] derived the uniform almost sure convergence for (DW), while [23] showed its asymptotic normality. [20] studied some properties of local polynomial regression for dependent data.

Despite this great importance the recursive kernel estimation of censored nonparametric has not yet been fully explored. The present work is the first contribution that consider a recursive estimate in censored data. The main aim of this contribution is to study the asymptotic properties of the recursive kernel estimator of the conditional density and its derivatives, under random right censoring. Specifically, the asymptotic properties stated are the strong convergence and the asymptotic normality of these estimators. The paper is organized as follows. We present our model in Section 2. In Section 3 we introduce notations, assumptions and we state the main results. Finally, the proofs of the main results are relegated to Section 4 with some auxiliary results with their proofs.

2. Presentation of estimates

Consider n pairs of independent random variables (X_i, T_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, T) which is valued in $\mathbb{R}^d \times \mathbb{R}$. In this paper we consider the problem of nonparametric estimation of the conditional density of Y given $X = x$ when the response variable Y_i are rightly censored. Furthermore, we denote by $(C_i)_{i=1,\ldots n}$ the censoring random variables which are supposed independent and identically distributed with a common unknown continuous distribution function G. Thus, we construct our estimators by the observed variables $(X_i, Y_i, \delta_i)_{i=1,\ldots n}$, where $Y_i = T_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$, where $\mathbb{1}_A$ 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.

The cumulative distribution function G, of the censoring random variables, is estimated by [14] estimator defined as follows

$$\bar{G}_n(t) = \begin{cases} \displaystyle\prod_{i=1}^{n} \left(1 - \frac{1 - \delta_{(i)}}{n - i + 1}\right)^{\mathbb{1}_{\{Y_{(i)} \leq t\}}} & \text{if} \quad t < Y_{(n)} \\ 0 & \text{otherwise} \end{cases},$$

which is known to be uniformly convergent to $\bar{G}$.

Given i.i.d. observations $(X_1, Y_1, \delta_1), \ldots, (X_n, Y_n, \delta_n)$ of (X, Y, δ), the kernel estimate of the conditional density $\phi(t|x)$ denoted $\breve{\phi}_n(t|x)$, is defined by

$$\forall x \in \mathbb{R}^d \text{ and } \forall y \in \mathbb{R} \qquad \breve{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_n^{-1} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) L\left(\frac{t - Y_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)},$$

where K, L are a kernels and h_n is a sequence of positive real numbers. Note that this last estimator has been recently used by [15].

A recursive version of the previous kernel estimator is defined by

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

where

$$\hat{g}_n(x,t) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x-X_i}{h_i}\right) L\left(\frac{t-Y_i}{h_i}\right), \qquad (2.1)$$

and

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

Remark 2.1. The Kaplan-Meir estimator is not recursive and the use of such estimator can slightly penalizes the efficiency of our estimator in term of computational time.

3. Assumptions and main results

Throughout the paper, we put $h_n^- = \inf_{i=1\ldots n} h_i$, $h_n^+ = \sup_{i=1\ldots n} h_i$ and all along the paper, when no confusion is possible, we denote by M and/or M' any generic positive constant. Further, we will denote by $F(\cdot)$ (resp. $G(\cdot)$) the distribution function of T (resp. of C) and by τ_F (resp. τ_G) the upper endpoints of the survival function $\bar{F}$ (resp. of $\bar{G}$). In the following we assume that $\tau_F < \infty$, $\bar{G}(\tau_F) > 0$ and C is independent to (X, T). We also assume that there exist a compact set $\mathcal{C} \subset \mathcal{C}_0 = \{x \in \mathbb{R}^d \ \ell(x) > 0\}$ where ℓ is the density of the explicative variable X, and Ω be a compact set such that $\Omega \subset (-\infty, \tau]$ where $\tau < \tau_F \wedge \tau_G$.

We introduce the following assumptions:

Assumption A1. The kernels K and L are Lipschitz continuous functions and compactly supported satisfy.

$$\int_{\mathbf{R^d}} u_l K(u) du = 0 \text{ for } l = 1, \ldots, d \text{ with } u = (u_1, \ldots, u_d)^T \ \ \text{and} \ \ \int_{\mathbf{R}} v L(v) dv = 0$$

Assumption A2.

(i) The marginal density $\ell(\cdot)$ is twice differentiable and satisfies a Lipschitz condition. Furthermore $\ell(x) > \Gamma$ for all $x \in \mathcal{C}$ and $\Gamma > 0$. Where $\mathcal{C}$ is a compact set of $\mathbb{R}$.

(ii) The joint density $g(\cdot, \cdot)$ of (X, T) is bounded function twice differentiable.

Remark 3.1. Assumptions A1 and A2 are usually used in non censoring kernel estimation method. The independence assumption between $(C_i)_i$ and $(X_i, T_i)_i$, may seem to be strong and one can think of replacing it by a classical conditional

independence assumption between $(C_i)_i$ and $(T_i)_i$ given $(X_i)_i$. However considering the latter demands an a priori work of deriving the rate of convergence of the censoring variable's conditional law (see [11]). Moreover our framework is classical and was considered by [7] and [17], among others.

3.1. Uniform strong consistency results with rate of convergence

In order to give the rate of the uniform almost sure convergence of our estimate we need the following additional assumptions:

Assumption C.

(i) The sequences h_n^+ and h_n^- satisfy $\lim_{n\to\infty} h_n^+ + \frac{\log n}{nh_n^{-d+1}} = 0 \to \infty$ as $n \to \infty$.

(ii) $\lim_{n\to\infty} n^\beta h_n^- = \infty$ for some $\beta > 0$.

Theorem 3.2. *Under Assumptions A1, A2 and C we have*

$$\sup_{x\in\mathcal{C}} \sup_{t\in\Omega} \left|\widehat{\phi}_n(t|x) - \phi(t|x)\right| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right), h_n^{+^2}\right)\right\} \quad a.s.\ as\ n \to \infty$$

(3.1)

Remark 3.3. Observe that, although the expression of the convergence rate is not classic in nonparametric statistic data analysis, this convergence rate is identifiable to the usual rate in the kernel method case where, for all i, we have $h_i = h_n = h_n^- = h_n^+$.

3.1.1. Proof of Theorem 3.2

Set

$$\widetilde{g}_n(x,t) := \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{(d+1)}}\delta_i \bar{G}^{-1}(Y_i)K_i(x)L_i(t)$$

with $K_i(x) = K(\frac{x-X_i}{h_i})$, $L_i(t) = L(\frac{t-Y_i}{h_i})$.

Now, the proof of this Theorem is based on the following decomposition

$$\sup_{x\in\mathcal{C}} \sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x) - \phi(t|x)\right| \leq \sup_{x\in\mathcal{C}} \sup_{t\in\Omega}\left|\frac{\widehat{g}_n(x,t)}{\ell_n(x)} - \frac{\widetilde{g}_n(x,t)}{\ell_n(x)}\right| + \left|\frac{\widetilde{g}_n(x,t)}{\ell_n(x)} - \frac{\mathbb{E}\widetilde{g}_n(x,t)}{\ell_n(x)}\right|$$

$$+ \left|\frac{\mathbb{E}\widetilde{g}_n(x,t)}{\ell_n(x)} - \frac{g(x,t)}{\ell_n(x)}\right| + \left|\frac{g(x,t)}{\ell_n(x)} - \frac{g(x,t)}{\ell(x)}\right|$$

$$\leq \frac{1}{\inf_{x\in\mathcal{C}}\ell_n(x)}\left\{\sup_{x\in\mathcal{C}} \sup_{t\in\Omega}|\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)|\right.$$

$$+ \sup_{x\in\mathcal{C}} \sup_{t\in\Omega}|\widetilde{g}_n(x,t) - g(x,t)|$$

$$\left. + \sup_{x\in\mathcal{C}} \sup_{t\in\Omega}|\phi(t|x)|\sup_{x\in\mathcal{C}}|\ell(x) - \ell_n(x)|\right\} \quad (3.2)$$

So, the proof of this Theorem is a direct consequence of Lemmas 3.4–3.6.

Lemma 3.4. *Under Assumptions C, A1 and A2(ii), we have*

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\tilde{g}_n(x,t) - g(x,t)| = O\left\{ \max\left(\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}} \right), h_n^{+^2} \right) \right\} \quad a.s. \ as \ n \to \infty.$$

Lemma 3.5. *Under Assumptions C, A1 and A2(i), we have*

$$\sup_{x \in \mathcal{C}} |\ell(x) - \ell_n(x)| = O\left\{ \max\left(\left(\sqrt{\frac{\log n}{nh_n^{-d}}} \right), h_n^{+^2} \right) \right\} \quad a.s. \ as \ n \to \infty.$$

Lemma 3.6. *Under Assumptions C, A1 and A2(ii), we have*

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| = O\left\{ \left(\sqrt{\frac{\log \log n}{n}} \right) \right\} \quad a.s. \ as \ n \to \infty.$$

3.2. Asymptotic normality

Now, we study the asymptotic normality of our estimate. To do that, we replace condition C by the following assumption:

Assumption N.

(i) $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (\frac{h_n}{h_i})^k = \theta_k$.
(ii) $h_n^{(d+1)} \log \log n = o(1)$, $\lim_{n \to \infty} nh_n^{(d+1)} h_n^{+^4} = 0$ and $\lim_{n \to \infty} nh_n^{(d+1)} = \infty$.

Theorem 3.7. *Under Assumptions A1, A2 and N, we have, for any $(x,t) \in \mathcal{A}$,*

$$\sqrt{nh_n^{(d+1)}} \left(\hat{\phi}_n(t|x) - \phi(t|x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(x,t) \right)$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution,

$$\sigma^2(x,t) = \theta_{d+1} \frac{\phi(t|x)}{\ell(x)\bar{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K^2(z)L^2(y)dzdy$$

and $\mathcal{A} = \{(x,t) \quad \sigma^2(x,t) \neq 0\}$.

Corollary 3.8. *Based on $\overline{G}_n(\cdot), \hat{\phi}_n(\cdot|x)$ and $\ell_n(x)$ we easily get a plug-in estimator $\hat{\sigma}_n^2(x,t)$ for $\sigma^2(x,t)$ which, under the assumptions of Theorem 3.7, gives a confidence interval of asymptotic level $1 - \alpha$ for $\phi(t|x)$*

$$\left[\hat{\phi}_n(t|x) - \frac{u_{1-\alpha/2}\hat{\sigma}_n(x,t)}{\sqrt{nh^{d+1}}}, \hat{\phi}_n(t|x) + \frac{u_{1-\alpha/2}\hat{\sigma}_n(x,t)}{\sqrt{nh^{d+1}}} \right]$$

where $u_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution.

3.2.1. Proof of Theorem 3.7

It is clear that

$$\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x) - \phi(t|x)\right) = \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right]$$

$$+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right]$$

$$+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)\right]$$

$$+ \sqrt{nh_n^{(d+1)}}\frac{g(x,t)}{\ell_n(x)\ell(x)}\left[\ell(x) - \ell_n(x)\right]. \qquad (3.3)$$

Thus, The proof of Theorem 3.7 can be deduced directly from the following Lemmas

Lemma 3.9. *Under the Hypotheses of Theorem 3.7, we have*

$$\sqrt{nh_n^{(d+1)}}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right] \to 0 \quad a.s. \ as \ n \to \infty, \qquad (3.4)$$

$$\sqrt{nh_n^{(d+1)}}\left[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)\right] \to 0 \qquad as \ n \to \infty \qquad (3.5)$$

and

$$\sqrt{nh_n^{(d+1)}}\left(\ell_n(x) - \ell(x)\right) \to 0 \quad in \ probability \qquad as \ n \to \infty. \qquad (3.6)$$

Lemma 3.10. *Under Assumptions A1, A2 and N(i), we have*

$$\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma'^2\right)$$

where $\sigma'^2(x,t) = \theta\frac{g(x,t)}{G(t)}\int_{\mathbb{R}^d}\int_{\mathbb{R}}K_i^2(z)L_i^2(y)dzdy.$

4. Numerical study

In this short section we compare the finite-sample performance of the recursive kernel method and the classical kernel via a Monte Carlo study. For this comparison study, we consider the same models used in [16] that is

$$\begin{array}{lll} M1 & Y = X^2 + 1 + \epsilon & \text{parabolic case,} \\ M2 & Y = sin(1.5X) + \epsilon & \text{sinus case,} \\ M3 & Y = exp(X - 0.2) + \epsilon & \text{exponential case} \end{array}$$

where the random variables X and ϵ are i.i.d. and follow respectively the normal distribution $N(0,1)$ and $N(0,\sigma)$.

It is clear that the conditional density expression is closely related to the distribution of ϵ. Thus, the conditional densities are respectively

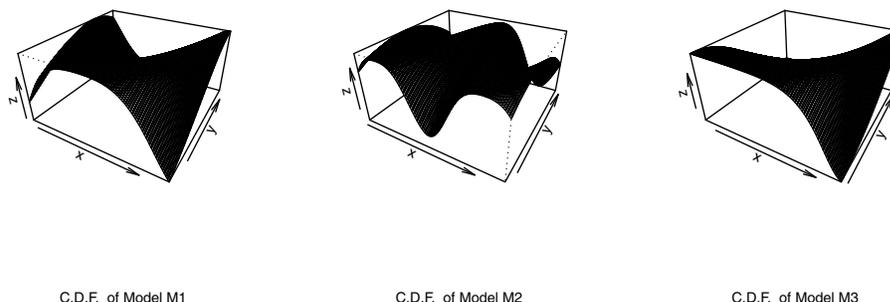C.D.F. of Model M1   C.D.F. of Model M2   C.D.F. of Model M3

FIG 1. *The C.D.f.*

In order to control the effect of the censoring in the efficiency of both estimators we variate the percentage of censoring for each models by considering a various censoring distributions. Precisely, we generate the censoring variables C by an exponential distribution $\mathcal{C}(\lambda_1)$ shifted by λ_2 (for the exponential model), by a normal distribution $N(0, \sigma_1)$ (for sinus case) and by $N(0, \sigma_2)$ (for parabolic case). Thus, the behavior of both estimators is evaluated over a several parameters, such as the sample size n, the percentage of censoring τ controlled by $(\lambda_1, \lambda_2, \sigma_1, \sigma_2)$, the dimension of the regressors d and the type of model M. For sake of shortness, we consider the unidimensional case, we fixe the sample size $n = 200$, we took $\sigma = 0.2$, we consider three censoring type ($\tau = 10$, $\tau = 40$ and $\tau = 70$). The test of the performance of both estimators is described by the following averaged squared errors

$$MSE(KERNEL) = \frac{1}{n} \sum_{i=1}^{n} (\bar{\phi}_n(Y_i|X_i) - \phi(Y_i|X_i))^2$$

and recursive

$$MSE(RECURSIVE) = \frac{1}{n} \sum_{i=1}^{n} (\widehat{\phi}_n(Y_i|X_i) - \phi(Y_i|X_i))^2$$

Now, for our practical study, we use the Gaussian kernel and we consider the well-known smoothing parameter defined by $h_{n,i} = \sigma_n^2 i^{-1/5}$ where

$$\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The obtained results are given in Table 1. It is clear from Table 1 that the recursive method is slightly better than the classical kernel method. However, the main advantage of the recursive method is that considerably faster than the classical one for the three models. In particular, it reduces sensibly the computation time in function of the sample size and the kind of models. Overall, both methods give a satisfactory level of accuracy and the latter is strongly dependent to the censoring rate.

Model	τ	MSE(KERNEL)	MSE(RECURSIVE)
M1	10	0.41	0.22
	40	0.64	0.55
	70	1.74	1.97
M2	10	0.59	0.36
	40	0.33	0.30
	70	1.80	1.84
M3	10	0.79	0.29
	40	1.32	1.18
	70	2.17	2.65

5. Proofs of the intermediates results

Proof of Lemma 3.4. We write

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\tilde{g}_n(x,t) - g(x,t)| \leq \sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\tilde{g}_n(x,t) - \mathbb{E}\tilde{g}_n(x,t)|$$
$$+ \sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\mathbb{E}\tilde{g}_n(x,t) - g(x,t)|.$$

For the quantity $\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\mathbb{E}\tilde{g}_n(x,t) - g(x,t)|$, we use the fact that, for all measurable function φ and for all $i = 1, \ldots, n$.

$$\mathbb{1}_{\{T_1 \leq C_1\}} \varphi(Y_1) = \mathbb{1}_{\{T_1 \leq C_1\}} \varphi(T_1).$$

Then,

$$\mathbb{E}\tilde{g}_n(x,t)$$
$$= n^{-1} \sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \mathbb{E}\left\{ K\left(\frac{x - X_1}{h_i}\right) \delta_i \bar{G}^{-1}(T_i) L\left(\frac{t - T_1}{h_i}\right) \right\}$$
$$= n^{-1} \sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \mathbb{E}\left\{ K\left(\frac{x - X_1}{h_i}\right) \bar{G}^{-1}(T_i) L\left(\frac{t - T_1}{h_i}\right) \mathbb{E}\left[\mathbb{1}_{\{T_i \leq C_i\}} | X_i, T_i\right] \right\}$$
$$= n^{-1} \sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \mathbb{E}\left\{ K\left(\frac{x - X_1}{h_1}\right) L\left(\frac{t - T_1}{h_1}\right) \right\}$$

Therefore,

$$|\mathbb{E}\tilde{g}_n(x,t) - g(x,t)|$$
$$\leq n^{-1} \sum_{i=1}^{n} \left| \int_{\mathbf{R}^d} \int_{\mathbf{R}} K(u)L(v)[g(x - h_i u, t - h_i v) - g(x,t)]dudv \right|$$
$$\leq M n^{-1} \sum_{i=1}^{n} h_i^2 \leq M h_n^{+^2}.$$

Therefore,

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\mathbb{E}\tilde{g}_n(x,t) - g(x,t)| = O(h_n^{+^2}).$$

Now, concerning the quantity $\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\tilde{g}_n(x,t) - \mathbb{E}\tilde{g}_n(x,t)|$ we use the compactness property of the sets $\mathcal{C}$ and Ω to write that, for some $(x_k)_{1 \leq k \leq \lambda_n}$ and $(t_j)_{1 \leq j \leq \kappa_n}$,

$$\mathcal{C} \subset \bigcup_{k=1}^{\lambda_n} \mathcal{B}(x_k, a_n) \qquad \text{and} \qquad \Omega \subset \bigcup_{j=1}^{\kappa_n} B(t_j, b_n)$$

where $\lambda_n \sim a_n^{-d}$ and $\kappa_n \sim b_n^{-1}$ with $a_n = b_n = n^{-(d+1)\beta - 1/2}$.

Now, for any $x \in \mathcal{C}$ and $t \in \Omega$, we set by $\tilde{k}(x) = \arg\min_k \|x_k - x\|$ and $\tilde{j}(t) = \arg\min_j |t - t_j|$ Then, for any $(x,t) \in \mathcal{C} \times \Omega$, we can write

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} |\tilde{g}_n(x,t) - \mathbb{E}\tilde{g}_n(x,t)| \leq \sup_{x \in \mathcal{C}} \sup_{t \in \Omega} \left| \tilde{g}_n(x,t) - \tilde{g}_n(x, t_{\tilde{j}}) \right|$$

$$+ \sup_{x \in \mathcal{S}} \sup_{t \in \Omega} \left| \mathbb{E}\tilde{g}_n(x, t_{\tilde{j}}) - \mathbb{E}\tilde{g}_n(x,t) \right|$$

$$+ \max_j \sup_{x \in \mathcal{C}} \left| \tilde{g}_n(x, t_j) - \tilde{g}_n(x_{\tilde{k}}, t_j) \right|$$

$$+ \max_j \sup_{x \in \mathcal{C}} \left| \mathbb{E}\tilde{g}_n(x_{\tilde{k}}, t_j) - \mathbb{E}\tilde{g}_n(x, t_j) \right|$$

$$+ \max_k \max_j |\tilde{g}_n(x_k, t_j) - \mathbb{E}\tilde{g}_n(x_k, t_j)|$$

$$=: \mathcal{T}_{1,n} + \mathcal{T}_{2,n} + \mathcal{T}_{3,n} + \mathcal{T}_{4,n} + \mathcal{T}_{5,n}. \qquad (5.1)$$

Concerning $(\mathcal{T}_{1,n})$: We use the Lipschitizian condition on L to get

$$\sup_{x \in \mathcal{C}} \sup_{t \in \Omega} \left| \tilde{g}_n(x,t) - \tilde{g}_n(x, t_{\tilde{j}}) \right|$$

$$\leq \sup_{x \in \mathcal{C}} \sup_{t \in \Omega} \frac{1}{n} \sum_{i=1}^n h_i^{-(d+1)} \delta_i \bar{G}^{-1}(Y_i) K_i(x) \left| L_i(t) - L_i(t_{\tilde{j}}) \right|$$

$$\leq \sup_{x \in \mathcal{C}} \sup_{t \in \Omega} C \left| t - t_{\tilde{j}} \right| \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} h_i^{-(d+1)} \delta_i \bar{G}^{-1}(Y_i) K_i(x)$$

$$\leq M b_n \frac{1}{n} \sum_{i=1}^n h_i^{-(d+2)}$$

$$\leq M \frac{b_n}{h_n^{-d+2}}. \qquad (5.2)$$

So, under Assumption C(ii), we have

$$(\mathcal{T}_{1,n}) = O\left(\sqrt{\frac{\log n}{n h_n^{-(d+1)}}} \right).$$

By using the same arguments as those used $\mathcal{T}_{1,n}$ we obtain

$$(\mathcal{T}_{2,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right), \quad (\mathcal{T}_{3,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right)$$

$$\text{and} \quad (\mathcal{T}_{4,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right).$$

Finally, in order to study $\mathcal{T}_{5,n}$ we use Bernstein's inequality. To do that, we put, for $1 \le i \le n$, $1 \le k \le \lambda_n$, and $1 \le j \le \kappa_n$

$$U_i = U_i(x_k, t_j)$$
$$:= \left\{ h_i^{-(d+1)} \delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) - \mathrm{E}\left[h_i^{-(d+1)} \delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) \right] \right\}.$$

Using the fact that the kernels K and L are bounded, we get

$$|U_i| \le Ch_i^{-(d+1)} \le Ch_n^{-(d+1)} = M.$$

Moreover, by a similar ideas to those used in the first part of this Lemma, we show that

$$Var(U_i) \le C \int_{\mathbf{R^d}} \int_{\mathbf{R}} \frac{1}{h_i^{-(d+1)}} K^2(u) L^2(v) g(x_k - rh, t_j - sh) du dv$$

$$\le Ch_i^{-(d+1)} \le Ch_n^{-(d+1)} := \sigma.$$

Hence, by Bernstein's inequality (see [13]), it follows that for all $\epsilon > 0$:

$$\mathrm{P}\left\{ \left| n^{-1} \sum_{i=1}^n U_i \right| > \epsilon \right\} \le 2 \exp\left\{ -\left(\frac{n\epsilon}{M}\right) h\left(\frac{M\epsilon}{\sigma^2}\right) \right\} \tag{5.3}$$

where $h(u) = 3u/(6 + 2u)$ for all $u > 0$.

Now, taking $\epsilon = \epsilon_0 \left(\frac{\log n}{nh_n^{-(d+1)}}\right)^{1/2}$, we have for any (k, j), we obtain

$$\mathrm{P}\left\{ \left| n^{-1} \sum_{i=1}^n U_i \right| > \epsilon \right\} \le 2 \exp\left\{ -\frac{3\epsilon_0^2 \log n}{3c + c_0} \right\}$$

$$\le 2n^{-C\epsilon_0^2}$$

Thus,

$$\mathbb{P}\left\{ \max_{k=1,\dots,\lambda_n} \max_{j=1,\dots,\kappa_n} \left| \sum_{i=1}^n U_i(x_k, t_j) \right| > \epsilon \right\} \le \mathcal{C}\lambda_n \kappa_n n^{-C\epsilon_0}. \tag{5.4}$$

Consequently, Borel-Cantelli's lemma and an appropriate choice of ϵ_0 allows us to write that:

$$\mathcal{T}_{5,n} = O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right). \tag{5.5}$$

Proof of Lemma 3.5. Firstly, we write

$$
\begin{aligned}
\sup_{x \in \mathcal{C}} |\ell_n(x) - \ell(x)| \quad &\leq \quad \sup_{x \in \mathcal{C}} |\ell_n(x) - \mathbb{E}[\ell_n(x)]| + \sup_{x \in \mathcal{C}} |\mathbb{E}[\ell_n(x)] - \ell(x)| \\
&=: \quad \mathcal{L}_{1n} + \mathcal{L}_{2n}.
\end{aligned}
$$

The first term $\mathcal{L}_{1n}$ is very close to the last part of Lemma 3.4. So, by a standard analytical argument we get,

$$
\mathcal{L}_{2n} = O(h_n^{+^2}). \tag{5.6}
$$

While the proof of the second term For $\mathcal{L}_{2n}$ follows the same lines as in Lemma 3.4. Therefore, we get

$$
\mathcal{L}_{1n} = O_{a.s.} \left(\frac{\log n}{n h_n^{-d}} \right)^{1/2}
$$

which completes the proof of this Lemma.

Proof of Lemma 3.6. It is clear that

$$
\begin{aligned}
|\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| \quad &\leq \quad \sum_{i=1}^{n} \left| \frac{1}{n h_i^{(d+1)}} \delta_i K_i(x) L_i(t) \left(\frac{1}{\bar{G}(Y_i)} - \frac{1}{\bar{G}_n(Y_i)} \right) \right| \\
&\leq \quad \frac{\sup_{t \leq \tau_F} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau)} \widetilde{g}_n(x,t) \tag{5.7}
\end{aligned}
$$

Since $\bar{G}_n(\tau) > 0$, in conjunction with the SLLN and the LIL on the censoring law (see formula (4.28) in [11]), the result is an immediate consequence of Lemma 3.4.

Proof of Lemma 3.9.

- **Proof of 3.4.** Similarly to the previous Lemma, we have

$$
\sqrt{n h_n^{(d+1)}} |\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| \leq \sqrt{n h_n^{(d+1)}} \frac{\sup_{t \leq \tau_F} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau)} \widetilde{g}_n(x,t).
$$

Further, as

$$
\sup_{t \leq \tau_F} |\bar{G}_n(t) - \bar{G}(t)| = O_{a.s.} \left(\sqrt{\frac{\log \log n}{n}} \right)
$$

then

$$
\sqrt{n h_n^{(d+1)}} \left(\frac{\sup_{t \leq \tau_F} \bar{G}_n(t) - \bar{G}(t)}{\bar{G}_n(\tau_F) \bar{G}(\tau_F)} \right) = O_{a.s.} \left(\sqrt{\log \log n h_n^{(d+1)}} \right).
$$

From N(ii) we obtain that

$$
\sqrt{n h_n^{(d+1)}} \frac{\sup_{t \leq \tau_F} \bar{G}_n(t) - \bar{G}(t)}{\bar{G}_n(\tau_F) \bar{G}(\tau_F)} = o_{a.s.}(1)
$$

The latter combined with the results of Lemma 3.4 allows us to complete the proof of 3.4.

- **Proof of 3.5.** It is shown in the first part of Lemma 3.4, that

$$[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)] = O(h_n^{+^2}).$$

Thus,

$$\sqrt{nh_n^{(d+1)}}\,[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)] = O(\sqrt{nh_n^{d+1}(h_n^{+^4})})$$

which goes to zero under the second part of Assumption N(ii).

- **Proof of 3.6.** By a simple analytical arguments we show that

$$Var\,(\ell_n(x) - \ell(x)) = O\left(n^{-1}\sum_{i=1}^{n} h_i^d\right) \text{ and } \mathbb{E}\,[\ell_n(x) - \ell(x)] = O(h_n^{+^2}).$$

Now, Assumption N(ii) gives $nh_n^{d+1}(h_n^{+^4} \to 0$ and Assumption N(i) implies that

$$nh_n^{d+1}Var\,(\ell_n(x) - \ell(x)) = o(1).$$

So,

$$\sqrt{nh_n^{d+1}}\,(\ell_n(x) - \ell(x)) \to 0 \quad \text{in probability} \quad as \quad n \to \infty.$$

Proof of Lemma 3.10. The proof of this Lemma is based on the version of the central limit Theorem given in ([19], p. 275) where the main point is to calculate the following limit

$$nh_n^{d+1}Var\,[\tilde{g}_n(x,t)] \to \sigma'^2(x). \tag{5.8}$$

Indeed, we have

$$
nh_n^{d+1}Var\,[\tilde{g}_n(x,t)]
$$

$$
= \frac{nh_n^{(d+1)}}{n^2}Var\left[\sum_{i=1}^{n} h_i^{-(d+1)}\bar{G}^{-1}(Y)K_i(x)L_i(t)\mathbb{1}_{\{T_1 \le C_1\}}\right]
$$

$$
= \frac{h_n^{(d+1)}}{n}\sum_{i=1}^{n} h_i^{-2(d+1)}\mathbb{E}\left[\bar{G}^{-2}(T)K_i^2(x)L_i^2(t)\mathbb{E}\,[\mathbb{1}_{T_1 \le C_1}|X_1,T_1]\right]
$$

$$
- \frac{h_n^{(d+1)}}{n}\sum_{i=1}^{n} h_i^{-2(d+1)}\left[\mathbb{E}\left\{\bar{G}^{-1}(T)K_i(x)L_i(t)\mathbb{E}\,[\mathbb{1}_{T_1 \le C_1}|X_1,T_1]\right\}\right]^2
$$

$$
:= \nabla_n^1 + \nabla_n^2.
$$

Observe that

$$\nabla_n^2 = h_n^{(d+1)}\mathbb{E}^2\,[\tilde{g}_n(x,t)].$$

Once again, we use the result of Lemma 3.4 to show that $\nabla_n^2 = o(1)$.

Now, concerning the first term ∇_n^1, we have

$$\nabla_n^1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right)^{(d+1)}\int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{K^2(z)L^2(y)}{\bar{G}(t - yh_i)}g(x - zh_i, t - yh_i)dzdy.$$

The continuity of the functions $\bar{G}$ and g permit to write

$$\nabla_n^1 = \frac{g(x,t)}{\bar{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K^2(z) L^2(y) \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{h_n}{h_i} \right)^{(d+1)} \right) + o \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{h_n}{h_i} \right)^{(d+1)} \right)$$

From Assumption A1(ii), we obtain the claimed result (5.8).

Let's now prove our asymptotic result. To do that we put

$$\left(nh_n^{(d+1)} \right)^{\frac{1}{2}} [\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))] = \sum_{i=1}^n w_{i,n}(x)$$

where

$$w_{i,n}(x) = \frac{\left(nh_n^{(d+1)} \right)^{\frac{1}{2}}}{nh_i^{d+1}} \left\{ \delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) - \mathbb{E}\left(\delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) \right) \right\}$$

and we prove that for some $\beta > 2$

$$\frac{\sum_{i=1}^n \mathbb{E}\left[|w_{i,n}(x)|^\beta \right]}{\left(Var \left(\sum_{i=1}^n w_{i,n}(x) \right) \right)^{(\beta)/2}} \to 0.$$

Indeed, set $\psi_{i,n}^\beta(x) = \mathbb{E}|w_{i,n}(x)|^\beta$. Applying the Cr-inequality (see [19], p. 155)

$$\psi_{i,n}^\beta(x) = \mathbb{E} \left| \frac{\left(nh_n^{(d+1)} \right)^{\frac{1}{2}}}{nh_i^{d+1}} \left[\delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) - \mathbb{E}\left(\delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) \right) \right] \right|^\beta$$

$$\leq 2^{\beta-1} \frac{\left(nh_n^{(d+1)} \right)^{\frac{\beta}{2}}}{n^\beta (h_i^{d+1})^\beta} \mathbb{E}\left(\left| \delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) \right|^\beta \right)$$

$$+ 2^{\beta-1} \frac{\left(nh_n^{(d+1)} \right)^{\frac{\beta}{2}}}{n^\beta (h_i^{d+1})^\beta} \left| \mathbb{E}\left(\delta_i \bar{G}^{-1}(Y_i) K_i(x) L_i(t) \right) \right|^\beta. \tag{5.9}$$

Furthermore, by a standard arguments, we show that

$$\mathbb{E}\left(\left| \delta_i \bar{G}^{-1}(Y) K_i(x) L_i(t) \right|^\beta \right) = O(h_i^{(d+1)})$$

and

$$\left| \mathbb{E}\left(\delta_i \bar{G}^{-1}(Y) K_i(x) L_i(t) \right) \right|^\beta = O(h_i^{\beta(d+1)}).$$

Therefore,

$$\psi_n^\beta(x) = \sum_{i=1} \psi_{i,n}^\beta(x) = O\left(n^{1-\beta/2} h_n^{(d+1)(1-\beta/2)} \right).$$

Because of $1 - \beta/2 < 0$ we have $\psi_n^\beta(x) \to 0$ which implies that

$$\lim_{n \to \infty} \frac{\sum_{i=1}^n \mathbb{E} \, |\psi_{i,n}(x)|^\beta}{\left(Var \left(\sum_{i=1}^n w_{i,n}(x) \right) \right)^{(\beta)/2}} \to 0.$$

The proof of this Lemma is now complete.

Acknowledgements

References

[1] AHMAD, I., LIN, P.E. (1976). Nonparametric sequential estimation of a multiple regression function, *Bull. Math. Statist.*, 17, 63–75. MR0436439

[2] AMIRI, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortemement mélangeant, *C. R. Acad. Sci. Paris, Ser I*, 347, 309–314. MR2537542

[3] AMIRI, A. (2012). Recursive regression estimators with application to nonparametric prediction, *J. Nonparam. Statist.*, 24(1), 169–186. MR2885832

[4] AMIRI, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, arXiv:1211.5767v2. MR3071881

[5] BERAN, R. (1981). Nonparametric regression with randomly censored survival data, Technical university of California, Berkeley.

[6] CARROLL, J. (1976). On sequential density estimation, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 36, 137–151. MR0423710

[7] CARBONEZ, A., GYÖRFI, L., VANDER MEULIN, E.C. (1995). Partitioning estimates of a regression function under random censoring, *Statist. & Decisions*, 13, 21–37. MR1334076

[8] DABROWSKA, D.M. (1987). Nonparametric regression with censored survival time data, *Scandi. J. Statist.*, 14, 181–197. MR0932943

[9] DABROWSKA, D.M. (1989). Uniform consistency of the kernel conditional Kaplan Meier estimate, *Ann. Statist*, 17, 1157–1167. MR1015143

[10] DAVIES, I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist.*, 15, 49–54. MR0334390

[11] DEHEUVELS, P., EINMAHL, J.H.J. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications, *Ann. Probab.*, 28, 1301–1335. MR1797314

[12] DEVROYE, L., WAGNER, T.J. (1980). On the L^1 convergence of kernel estimators of regression functions with application in discrimination, *Z. Wahrschein. Verw. Get.*, 51, 15–25. MR0566104

[13] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, 58, 13–30. MR0144363

[14] KAPLAN, E.L., MEIER, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, 53, 457–481.

[15] 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. MR2780217

[16] 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. MR2817352

[17] KOHLER, M., MÁTHÉ, K., PINTER, M. (2002). Prediction from randomly Right Censored Data, *J. Multivariate Anal.*, 80, 73–100. MR1889835

[18] KRZŸZAK, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation, *IEEE Trans. Inform. Theory*, 38, 1323–1338. MR1168751

[19] LOÈVE, M. (1963). *Probability Theory*, New York: Springer-Verlag. MR0123342

[20] MASRY, E., FAN, J. (1997). Local polynomial estimation of recursive function for mixing processes, *Scandinave Journal of Statistics*, 24, 165–179. MR1455865

[21] NGUYEN, T., SARACCO, J. (2010). Estimation récursive en régression inverse par tranches, *Journal de la société française de statistique*, 151(2), 19–46. MR2652790

[22] ROUSSAS, G.G. (1990). Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl.*, 36(1), 107–116. MR1075604

[23] ROUSSAS, G.G., TRAN, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions, *Annals of Statist.*, 20(1), 98–120. MR1150336

[24] WALK, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Ann. Inst. Statist. Math.*, 53(4), 691–707. MR1879605

[25] WEGMAN, J., DAVIES, I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Statist.*, 7, 316–327. MR0520242

[26] WERTZ, W. (1985). Sequential and recursive estimators of the probability density, *Statistics*, 16, 277–295. MR0789071

[27] WOLVERTON, C., WAGNER, T.J. (1969). Asymptotically optimal discriminant functions for pattern classifcation, *IEEE Trans. Inform. Theory*, 15, 258–265. MR0275576

[28] YAMATO, H. (1972). Sequential estimation of a continuous probability density function and mode, *Bull. Math. Statist.* 14, 1–12. MR0381187