



Adaptive recursive kernel conditional density estimators under censoring data

Yousri Slaoui and Salah Khardani

University of Poitiers
Laboratoire de Mathématiques et Applications
11 Boulevard Marie et Pierre Curie
86962 Futuroscope Chasseneuil.
E-mail address: Yousri.Slaoui@math.univ-poitiers.fr
URL: <http://www-math.univ-poitiers.fr/~yslaoui/>

University of Carthage
Ecole Nationale des Sciences et Technologies Avancées à Borj-Cédria
Laboratoire des Réseaux Intelligents et Nanotechnologie.
E-mail address: khardani_salah@yahoo.fr

Abstract. Let (T, X) be independent identically distributed pairs of random variables and denote $f(t|x)$ the conditional density of T given $X = x$, we consider that the random variable T is subject to random censoring by another random variable C . In this paper, we propose and investigate an adaptive recursive kernel conditional density estimation under censored data, which allows us to circumvent the weak performances of Kaplan-Meier estimator ([Kaplan and Meier, 1958](#)) in the right-tail of the distribution. The first aim of this paper is to study the properties of the proposed adaptive recursive estimators and compare it with the non-recursive estimator of $f(t|x)$. It turns out that, with an adequate selected bandwidth and a special stepsize, the proposed recursive estimators often provides better results compared to the non-recursive one in terms of estimation error and much better in terms of computational costs. We corroborated these theoretical results through some simulation study.

1. Introduction

The subject of this paper concerns the recursive kernel estimation of the conditional density function under censoring data when the covariates take values in finite dimensional \mathbb{R}^d . A general scheme of random censorship can be defined in the

Received by the editors May 19th, 2019; accepted March 15th, 2020.

2010 Mathematics Subject Classification. 62G05, 62G07, 62G08, 62G20, 62H12.

Key words and phrases. Asymptotic normality, censored data, conditional density, recursive kernel estimators, stochastic approximation algorithm, smoothing, survival data.

This work benefited from the financial support of the GDR 3477.

following way. Consider a triple (T, C, X) of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$. Here T is the variable of interest, C a censoring variable and X a concomitant variable. Throughout, we work with a sample $\{(T_i, C_i, X_i)_{1 \leq i \leq n}\}$ of independent and identically distributed replicæ of (T, C, X) , $n \geq 1$. In fact, in the right censorship model, the pairs (T_i, C_i) , $1 \leq i \leq n$, are not directly observed and the relevant information is given by $Y_i := \min\{T_i, C_i\}$ and $\delta_i := \mathbb{1}\{T_i \leq C_i\}$, $1 \leq i \leq n$. Survival data in clinical trials or failure time data in reliability studies, are often subject to such censoring. To be more specific, many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life which result in death. In the sequel, we impose the following assumptions upon the distribution of (X, T) . For $-\infty < t < \infty$, set $F(t) = \mathbb{P}(T \leq t)$, $G(t) = \mathbb{P}(C \leq t)$, and $L(t) = \mathbb{P}(Y \leq t)$, the right-continuous distribution functions of T , C and Y respectively. In order to estimate uniquely the distribution function F from the observed data (which ensures the identifiability of the model), we assume the independence between the T and C . Moreover, in this case the distribution function $L(t) = P(Y \leq t)$ satisfies $1 - L(t) = (1 - F(t))(1 - G(t))$. For any x , we can write the conditional density function $f(t|x)$ as $\frac{f(x,t)}{f(x)}$, where $f(\cdot, \cdot)$ is the joint probability density function of (X, T) and $f(\cdot)$ is the marginal density of X with respect to the Lebesgue measure. Studying the relationship between a response variable and an 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 nonparametric 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. Conditional density estimation of a scalar response given a scalar/multivariate covariate gained considerable interest in the statistical literature. For completely observed data, many authors are interested in the estimation of the conditional density and several nonparametric approaches have been proposed. For example, [Tjøstheim \(1994\)](#) and [Polonik and Yao \(2000\)](#) estimated the conditional density function indirectly. [Hyndman et al. \(1996\)](#) studied kernel estimator of the conditional density and its bias-corrected version. [Fan et al. \(1996\)](#) developed a direct estimation method via an innovative 'double-kernel' local linear approach. [Bashannyk and Hyndman \(2001\)](#) and [Hyndman and Yao \(2002\)](#) proposed several simple and useful rules for selecting bandwidths for the conditional density estimation. [Hall et al. \(2004\)](#) applied the cross-validation technique to estimate the conditional density. [Fan and Yim \(2004\)](#) proposed a consistent data-driven bandwidth selection procedure in estimating the conditional density functions. [Efro-movich \(2007\)](#) develops the theory of minimax estimation of the conditional density for regression settings with fixed and random designs of predictors, bounded and unbounded responses.

[Efro-movich \(2010\)](#) write a conditional density as a sum of orthogonal projection on all possible subspaces of reduced dimensionality and then estimating each projection via a shrinkage method. [Bertin et al. \(2016\)](#) consider the problem of estimating the conditional density in the case of independent sample distributed in the multivariate setting.

However, all these papers assume that the observations are complete. Noting that the nonparametric modeling of censored data is intensively discussed in the recent statistical literature. It dates back to [Beran \(1981\)](#), who introduced a class of nonparametric regression estimators for the conditional survival function in the presence of right-censoring. [Dąbrowska \(1987, 1989\)](#) studied the asymptotic properties of the distribution and quantiles functions estimators. [Kohler et al. \(2002\)](#) gave a simpler proof in the randomly right-censoring case for the kernel, nearest neighbor, least squares and penalized least squares estimates. [Brunel et al. \(2007\)](#) provided an adaptive nonparametric strategy to estimate the conditional density in the presence of censoring. [Bouaziz and Lopez \(2010\)](#) introduced a semi parametric procedure to estimate the conditional density under censoring response. [Liang and Liu \(2013\)](#) defined a kernel estimator of the conditional density for a left-truncated and right-censored model based on the generalized product-limit estimator of the conditional distributed function.

In this vast variety of papers, the authors use the Nadaraya-Watson techniques as an estimation method which is a particular case of the recursive kernel estimator considered in this work.

In recent years, data streams have become an increasingly important area of research. Internet packet data, stock market activity, credit card transactions, and Internet and phone usage. In those situations, the data arrive so rapidly that it is impossible for the user to store them all in the disk (as a traditional database), and then interact with them at the time of our choosing.

In such a large sample data context, building a recursive estimators 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. In this way, the recursive estimator can be updated with each new observation. This iterative scheme saves computer time in a practical case, whereas a non-recursive estimator needs to be recalculated completely when a new data set is observed. This propriety is particularly important in the conditional density estimation since the number of points at which the function is estimated is usually very large.

The Robbins-Monro procedure was originally proposed by [Robbins and Monro \(1951\)](#) and further developed and investigated as well as applied in many different situations (see, among many others, [Blum \(1954\)](#), [Fabian \(1967\)](#); [Kushner and Clark \(1978\)](#); [Hall and Heyde \(1980\)](#); [Ruppert \(1982\)](#); [Chen \(1988\)](#); [Duflo \(1996\)](#); [Dippon and Renz \(1997\)](#); [Révész \(1973, 1977\)](#); [Dippon \(2003\)](#), [Mokkadem and Pelletier \(2007\)](#)).

The first recursive version of Rosenblatt's kernel density estimator and the most famous one was introduced by [Wolverton and Wagner \(1969\)](#), and was widely studied; see among many others [Yamato \(1970/71\)](#), [Davies \(1972/73\)](#), [Devroye \(1979\)](#), [Wegman and Davies \(1979\)](#) and [Roussas and Tran \(1992\)](#). Competing recursive estimators, which may be regarded as weighted versions of Wolverton and Wagner's estimator, were introduced and studied by [Deheuvels \(1973\)](#), [Wegman and Davies \(1979\)](#) and [Duflo \(1997\)](#). [Hall and Patil \(1994\)](#) defined a large class of weighted recursive estimators, including all the previous recursive estimators. Recently, [Slaoui \(2013, 2014a, 2014b, 2015, 2016\)](#), studied the problem of the bandwidth selection and derived some properties of the recursive kernel estimator of the density function, distribution function and then of the regression function. [Khardani](#)

and Slaoui (2019) defined and studied a new estimator of the regression function when the response random variable is subject to random right-censoring.

The first aim of this paper is to propose an adaptive recursive kernel conditional density estimation under censored data, and the second aim is to give the conditions under which the recursive estimators $\hat{f}(t|x)$ (defined below) is consistent. The paper is organized as follows. In Section 2 we recall our kernel conditional density estimator in the censorship model with some notations. In Section 3 the assumptions and main results are given. Section 4 is devoted to our simulation experiments. The proofs of the main results are relegated to Section 5 with some auxiliary results with their proofs. We concluded the paper in Section 6.

2. Presentation of estimates

Given independent and identically distributed observations $(X_1, Y_1, \delta_1), \dots, (X_n, Y_n, \delta_n)$ of (X, Y, δ) , the kernel estimate of the conditional density $f(t|x)$ denoted $\bar{f}_n(t|x)$, is defined for any $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$ by

$$\bar{f}_n(t|x) = \frac{\sum_{i=1}^n h_n^{-1} \delta_i \bar{G}^{-1}(Y_i) K_0\left(\frac{x-X_i}{h_n}\right) K_1\left(\frac{t-Y_i}{h_n}\right)}{\sum_{i=1}^n K_0\left(\frac{x-X_i}{h_n}\right)}, \quad (2.1)$$

where K_0, K_1 are kernels and h_n is a sequence of positive real numbers. This estimator was introduced in Ould-Saïd (2006). In practice, the survival function $\bar{G}(\cdot)$ is unknown, hence it is not possible to use (2.1) as an estimator. One way to overcome this difficulty is to replace $\bar{G}(\cdot)$ by a modified Kaplan and Meier (1958) estimate $\bar{G}_n(\cdot)$ (see Lo et al., 1989) defined as

$$\bar{G}_n(t) = \prod_{Y_{(i)} \leq t} \left(\frac{n-i+1}{n-i+2} \right)^{1-\delta_{(i)}}.$$

As mentioned in Brunel et al. (2007), \bar{G}_n was proposed to satisfy the following property: $\bar{G}_n(t) \geq 1/(n+1)$, for all t , and which it convergence to \bar{G} was proven in Brunel and Comte (2005). $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ are the order statistics of $(Y_{(i)})_{1 \leq i \leq n}$ and $\delta_{(i)}$ is the concomitant of $Y_{(i)}$.

Then, our proposal estimators to estimate recursively the conditional density $f(t|x)$, can be viewed as an adaptive recursive version of the previous kernel estimator and defined as follows

$$\hat{f}_n(t|x) = \begin{cases} \frac{\hat{f}_n(x,t)}{\hat{f}_n(x)} & \text{if } \hat{f}_n(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (2.2)$$

where for any $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$,

$$\begin{aligned} \hat{f}_n(x,t) : &= (1 - \beta_n) \hat{f}_{n-1}(x,t) \\ &+ \beta_n h_n^{-d-1} \delta_n \bar{G}_n^{-1}(Y_n) K_0\left(\frac{x-X_n}{h_n}\right) K_1\left(\frac{t-Y_n}{h_n}\right), \end{aligned} \quad (2.3)$$

and for any $x \in \mathbb{R}^d$,

$$\hat{f}_n(x) := (1 - \gamma_n) \hat{f}_{n-1}(x) + \gamma_n h_n^{-d} K_0\left(\frac{x-X_n}{h_n}\right), \quad (2.4)$$

with (β_k) and (γ_k) sequences of positive real numbers there go to zero. We let $(Q_n, \Pi_n) = (\prod_{i=1}^n (1 - \beta_i), \prod_{i=1}^n (1 - \gamma_i))$ and $(\widehat{f}_0(x, t), \widehat{f}_0(x)) = (0, 0)$. Then, it follows from (2.3), that for any $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$,

$$\widehat{f}_n(x, t) := Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \delta_k \overline{G}_n^{-1}(Y_k) K_0\left(\frac{x - X_k}{h_k}\right) K_1\left(\frac{t - Y_k}{h_k}\right), \tag{2.5}$$

and from (2.4), that for any $x \in \mathbb{R}^d$,

$$\widehat{f}_n(x) := \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-d} K_0\left(\frac{x - X_k}{h_k}\right). \tag{2.6}$$

This recursive property is particularly useful when the observation number increases since $\widehat{f}_n(x, t)$ [resp. $\widehat{f}_n(x)$] can be easily updated with each additional observation. From a practical point of view, this arrangement provides important savings in computational time which is a consequence of the fact that the estimate updating is independent of the history of the data. The main drawback of the non-recursive kernel estimator is the use of all data at each step of estimation. From a theoretical point of view, the main advantage of the investigation of such recursive estimators is that we can prove consistency with the same rate as the non-recursive one.

Remark 2.1.

- For simplification, the bandwidth parameter that controls the smoothness of the estimation is chosen to be the same in the both considered kernels K_0 and K_1 .
- To limit the effect of censoring of the interest variable T in equation (2.1) and (2.5) and to have a less biased estimator, we consider the function \overline{G} . Note that our proposed estimator is a generalization of the estimator considered by [Khardani and Semmar \(2014\)](#), which is a special case of our proposed estimator with a stepsize $(\beta_n) = (n^{-1})$.
- The joint probability density function $f(\cdot)$ is not affected by censoring and is therefore consistently estimated by $\widehat{f}_n(\cdot)$.
- Many authors investigated the problems caused by the bad behavior of Kaplan-Meier estimators in the tail of the distribution, we refer the readers to [Marron and Padgett \(1987\)](#), [Brunel and Comte \(2005\)](#), [Brunel et al. \(2007\)](#) and [Bouaziz and Lopez \(2010\)](#).

3. Assumptions and main results

We define the following class of regularly varying sequences.

Definition 3.1. Let $\gamma \in \mathbb{R}$ and $(v_n)_{n \geq 1}$ be a nonrandom positive sequence. We say that $(v_n) \in \mathcal{GS}(\gamma)$ if

$$\lim_{n \rightarrow +\infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \tag{3.1}$$

For any distribution function L let $\tau_L = \sup\{t, L(t) < 1\}$ be its support's right endpoint.

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 \overline{F} (resp. of \overline{G}).

To formulate our results, we make use of the following assumptions gathered here for easy reference.

A1: The joint density $f(\cdot, \cdot)$ of (X, T) is bounded and continuously differentiable up to order 2.

A2: The sequence (X, T) and C are independent.

A3: The density $f(\cdot)$ is twice continuously differentiable.

A4: (i) $K_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}^d} K_0(z) dz = 1$, and, for all $j \in \{1, \dots, d\}$, $\int_{\mathbb{R}} z_j K_0(z) dz_j = 0$ and $\int_{\mathbb{R}^d} z_j^2 |K_0(z)| dz < \infty$, where $z = (z_1, \dots, z_d) \in \mathbb{R}^d$.

(ii) $K_1 : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K_1(z) dz = 1$, $\int_{\mathbb{R}} z K_1(z) dz = 0$ and $\int_{\mathbb{R}} z^2 |K_1(z)| dz < \infty$.

A5: (i) $(h_n) \in \mathcal{GS}(-a)$ with $a \in]0, 1]$.

(ii) $(\beta_n) \in \mathcal{GS}(-\beta)$ with $\beta \in]1/2, 1]$.

(iii) $(\gamma_n) \in \mathcal{GS}(-\alpha)$ with $\alpha \in]1/2, 1]$, such that $\lim_{n \rightarrow +\infty} \frac{\gamma_n h_n}{\beta_n} = 0$.

(iv) $\lim_{n \rightarrow \infty} (n\beta_n) \in]\min\{2a, \frac{\beta-(d+1)a}{2}\}, \infty]$.

(v) $\lim_{n \rightarrow \infty} (n\gamma_n) \in]\min\{2a, \frac{\alpha-ad}{2}\}, \infty]$.

A6: $\tau_F < \infty$, $\overline{G}(\tau_F) > 0$ and $\tau < \tau_F < \tau_G$.

Discussion of the assumptions.

- **A2** is specific to the model: the independence between $(C_n)_n$ and $(X_n, Y_n)_n$ may seem to be strong and one can think of replacing it by a classical conditional independence assumption between $(C_n)_n$ and $(Y_n)_n$ given $(X_n)_n$.
- Assumptions **A1**, **A3** and **A4** are classical in nonparametric estimation.
- Assumptions **A5(i) – A5(iii)** gives conditions for the bandwidth which allows getting the bias and variance of our estimators.
- Assumptions **A5(iv)** and **A5(v)** on the limit of $(n\beta_n)$ and on the limit of $(n\gamma_n)$ as n goes to infinity are usual in the framework of stochastic approximation algorithms. They implies in particular that the limit of $([n\beta_n]^{-1})$ and $([n\gamma_n]^{-1})$ are finite.
- Condition (3.1) was introduced by [Galambos and Seneta \(1973\)](#) to define regularly varying sequences (see also [Bojanic and Seneta, 1973](#)) and by [Mokkadem and Pelletier, 2007](#)) in the context of stochastic approximation algorithms. Noting that the acronym \mathcal{GS} stand for (Galambos and Seneta). Typical sequences in $\mathcal{GS}(\gamma)$ are, for $b \in \mathbb{R}$, $n^\gamma (\log n)^b$, $n^\gamma (\log \log n)^b$, and so on.

Throughout this paper we will use the following notations:

$$\epsilon_1 = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1} \quad (3.2)$$

$$\epsilon_2 = \lim_{n \rightarrow \infty} (n\beta_n)^{-1} \quad (3.3)$$

$$\begin{aligned} \mu_i^2 &= \int_{\mathbb{R}^d} x_i^2 K_0(x) dx, \quad f_{j,j}^{(2)}(x, t) = \frac{\partial^2 f}{\partial x_j \partial x_j}(x, t), \quad f^{(2)}(x, t) = \frac{\partial^2 f}{\partial t \partial t}(x, t), \\ I_2 &= \int_{\mathbb{R}} t^2 K_1(t) dt, \quad S_1 = \sum_{j=1}^d \mu_j^2 f_{j,j}^{(2)}(x, t), \quad S_2 = I_2 f^{(2)}(x, t), \\ S_3 &= \sum_{j=1}^d \mu_j^2 f_{j,j}^{(2)}(x), \quad E_1 = \left[\frac{S_1 + S_2}{(1 - 2a\epsilon_2)} - \frac{f(t|x) S_3}{(1 - 2a\epsilon_1)} \right]^2, \\ E_2 &= \frac{1}{(2 - (\beta - (d + 1)a)\epsilon_2)} \frac{f(x, t)}{G(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy. \end{aligned}$$

Our first result is the following proposition, which gives the bias and variance of $\widehat{f}_n(t|x)$.

Proposition 3.2. *(Bias and variance of $\widehat{f}_n(t|x)$)*

Under Assumptions **A1** – **A6**, for $t \in A \subsetneq [0, \tau]$ and assuming that $f^{(2)}(x, t)$ is continuous at (x, t) and that $f^{(2)}(x)$ is continuous at x . If $a \in [0, \beta/(d + 5)]$, then

$$\mathbb{E} \left[\widehat{f}_n(t|x) \right] - f(t|x) = \frac{1}{2f(x)} E_1^{1/2} h_n^2 + o(h_n^2). \tag{3.4}$$

If $a \in]\beta/(d + 5), 1[$, then

$$\mathbb{E} \left[\widehat{f}_n(t|x) \right] - f(t|x) = o \left(\sqrt{\beta_n/h_n^{d+1}} \right). \tag{3.5}$$

If $a \in [0, \beta/(d + 5)[$, then

$$\text{Var} \left[\widehat{f}_n(t|x) \right] = o(h_n^4). \tag{3.6}$$

If $a \in [\beta/(d + 5), 1[$, then

$$\text{Var} \left[\widehat{f}_n(t|x) \right] = \frac{1}{f^2(x)} E_2 \frac{\beta_n}{h_n^{d+1}} + o \left(\frac{\beta_n}{h_n^{d+1}} \right). \tag{3.7}$$

Let us now state the following theorem, which gives the weak convergence rate of $\widehat{f}_n(t|x)$ defined in (2.2). Below, we write $Z \stackrel{\mathcal{D}}{=} \mathcal{N}(\mu, \sigma^2)$ whenever the random variable Z follows a normal law with expectation μ and variance σ^2 , $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution and $\xrightarrow{\mathbb{P}}$ the convergence in probability.

Theorem 3.3 (Weak pointwise convergence rate). *Under Assumptions **A1** – **A6**, for $t \in A \subsetneq [0, \tau]$ and assuming that $f^{(2)}(x, t)$ is continuous at (x, t) and that $f^{(2)}(x)$ is continuous at x .*

(1) *If there exists $c \geq 0$ such that $\beta_n^{-1} h_n^{d+5} \rightarrow c$, then*

$$\sqrt{\beta_n^{-1} h_n^{d+1}} \left(\widehat{f}_n(t|x) - f(t|x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\sqrt{c}}{2f(x)} E_1^{1/2}, \frac{E_2}{f^2(x)} \right).$$

(2) *If $\beta_n^{-1} h_n^{d+5} \rightarrow \infty$, then*

$$\frac{1}{h_n^2} \left(\widehat{f}_n(t|x) - f(t|x) \right) \xrightarrow{\mathbb{P}} \frac{E_1^{1/2}}{2f(x)}.$$

Remark 3.4. The variance of the considered estimator is of order $O\left(\frac{\beta_n}{h_n^{d+1}}\right)$ and not $O\left(\frac{\beta_n}{h_n^d}\right)$ and consequently the rate of convergence is of order $O\left(\sqrt{\beta_n^{-1}h_n^{d+1}}\right)$ and not $O\left(\sqrt{\beta_n^{-1}h_n^d}\right)$ as usual obtained in density estimation and regression estimation. It is due to the presence of the kernel K_0 .

In order to measure the quality of our recursive estimator (2.2), we use the Mean Squared Error (*MSE*).

Proposition 3.5 (*MSE of $\hat{f}_n(t|x)$*). *Under Assumptions A1 – A6, for $t \in A \subsetneq [0, \tau]$ and assuming that $f^{(2)}(x, t)$ is continuous at (x, t) and that $f^{(2)}(x)$ is continuous at x .*

(1) *If $a \in (0, \beta / (d + 5))$, then*

$$MSE \left[\hat{f}_n(t|x) \right] = \frac{E_1}{4f^2(x)} h_n^4 + o(h_n^4).$$

(2) *If $a = \beta / (d + 5)$, then*

$$MSE \left[\hat{f}_n(t|x) \right] = \frac{E_1}{4f^2(x)} h_n^4 + \frac{E_2}{f^2(x)} \frac{\beta_n}{h_n^{d+1}} + o\left(h_n^4 + \frac{\beta_n}{h_n^{d+1}}\right).$$

(3) *If $a \in (\beta / (d + 5), 1)$, then*

$$MSE \left[\hat{f}_n(t|x) \right] = \frac{E_2}{f^2(x)} \frac{\beta_n}{h_n^{d+1}} + o\left(\frac{\beta_n}{h_n^{d+1}}\right).$$

The following Corollary ensures that the bandwidth which minimizes the *MSE* depend on the stepsizes (γ_n) and (β_n) and then the corresponding *MSE* depend also on the stepsizes (γ_n) and (β_n) .

Corollary 3.6. *Let Assumptions A1–A5 hold. To minimize the MSE of $\hat{f}_n(t|x)$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$ and the stepsize (γ_n) must be chosen in $\mathcal{GS}(-1)$, the bandwidth (h_n) must equal*

$$\left(\left\{ (d + 1) \frac{E_2}{E_1} \right\}^{1/(d+5)} \beta_n^{1/(d+5)} \right).$$

Then, we have

$$MSE \left[\hat{f}_n(t|x) \right] = \frac{1}{4f^2(x)} \frac{(d + 5)}{(d + 1)^{(d+1)/(d+5)}} E_1^{(d+1)/(d+5)} E_2^{4/(d+5)} \beta_n^{4/(d+5)} + o\left(\beta_n^{4/(d+5)}\right).$$

The following Corollary shows that a special choice of the stepsizes $((\beta_n), (\gamma_n)) = ((\gamma_0 n^{-1}), (\beta_0 n^{-1}))$ ensures that the bandwidth which minimizes the *MSE* depend on the stepsizes (β_n) and (γ_n) , which fulfilled that $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$, $\lim_{n \rightarrow \infty} n\beta_n = \beta_0$ and that $(\gamma_n) \in \mathcal{GS}(-1)$ and $(\beta_n) \in \mathcal{GS}(-1)$, then the optimal value for h_n depend on γ_0 and β_0 and the corresponding *MSE* depend also on γ_0 and β_0 .

Corollary 3.7. *Let Assumptions A1–A5 hold. To minimize the MSE of $\hat{f}_n(t|x)$, the stepsize (β_n) must be chosen in $\mathcal{GS}(-1)$, $\lim_{n \rightarrow \infty} n\beta_n = \beta_0$ and the stepsize*

(γ_n) must be chosen in $\mathcal{GS}(-1)$, $\lim_{n \rightarrow \infty} n\gamma_n = \gamma_0$ the bandwidth (h_n) must equal

$$\left(\left\{ (d+1) \frac{E'_2}{E'_1} \right\}^{1/(d+5)} \beta_0^{1/(d+5)} n^{-1/(d+5)} \right). \tag{3.8}$$

Then, we have

$$\begin{aligned} \text{MSE} \left[\widehat{f}_n(t|x) \right] &= \frac{1}{4f^2(x)} \frac{(d+5)}{(d+1)^{(d+1)/(d+5)}} E_1^{(d+1)/(d+5)} E_2^{4/(d+5)} \beta_0^{4/(d+5)} n^{-4/(d+5)} \\ &+ o \left(n^{-4/(d+5)} \right), \end{aligned}$$

with

$$\begin{aligned} E'_1 &= \left[\beta_0 \frac{S_1 + S_2}{(\beta_0 - 2a)} - \gamma_0 \frac{f(t|x) S_3}{(\gamma_0 - 2a)} \right]^2, \\ E'_2 &= \frac{\beta_0}{(2\beta_0 - (\beta - (d+1)a))} \frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dy dz. \end{aligned}$$

3.1. Computational cost. The advantage of recursive estimators over their nonrecursive counterparts is that their update, from a sample of size n to one of size $n + 1$, requires fewer computations. This property can be generalized, if we suppose that we receive separately two sets of data, the first one of cardinal n_1 smaller or equal to $n - 1$ and the second set of cardinal $n_2 = n - n_1$. First, we approximate the sequence (\widehat{f}_n) by the unobservable sequence (\widetilde{f}_n) recursively defined by

$$\begin{aligned} \widetilde{f}_n(x, t) &= (1 - \beta_n) \widetilde{f}_{n-1}(x, t) \\ &+ \beta_n h_n^{-d-1} \delta_n \overline{G}^{-1}(Y_n) K_0 \left(\frac{x - X_n}{h_n} \right) K_1 \left(\frac{t - Y_n}{h_n} \right). \end{aligned} \tag{3.9}$$

We can check that it follows from (3.9) that

$$\begin{aligned} \widetilde{f}_n(x, t) &= \prod_{j=n_1+1}^n (1 - \beta_j) \widetilde{f}_{n_1}(x, t) \\ &+ \sum_{k=n_1}^{n-1} \left[\prod_{j=k+1}^n (1 - \beta_j) \right] \frac{\beta_k}{h_k^{d+1}} \delta_k \overline{G}^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \\ &+ \frac{\beta_n}{h_n^{d+1}} \delta_n \overline{G}^{-1}(Y_n) K_0 \left(\frac{x - X_n}{h_n} \right) K_1 \left(\frac{t - Y_n}{h_n} \right) \\ &= \alpha_1 \widetilde{f}_{n_1}(x, t) + \sum_{k=n_1}^{n-1} \zeta_k \frac{\beta_k}{h_k^{d+1}} \delta_k \overline{G}^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \\ &+ \frac{\beta_n}{h_n^{d+1}} \delta_n \overline{G}^{-1}(Y_n) K_0 \left(\frac{x - X_n}{h_n} \right) K_1 \left(\frac{t - Y_n}{h_n} \right), \end{aligned}$$

where $\alpha_1 = \prod_{j=n_1+1}^n (1 - \beta_j)$ and $\zeta_k = \prod_{j=k+1}^n (1 - \beta_j)$. In order to highlight the performance of our proposed estimator in terms of computational cost, and since

\bar{G}_n converge to G (see Lemma 5.4), we can consider the following estimator as a consistent estimator of (2.3)

$$\begin{aligned}\widehat{f}_n(x, t) &= \prod_{j=n_1+1}^n (1 - \beta_j) \widehat{f}_{n_1}(x, t) \\ &+ \sum_{k=n_1}^{n-1} \left[\prod_{j=k+1}^n (1 - \beta_j) \right] \frac{\beta_k}{h_k^{d+1}} \delta_k \bar{G}_{n_2}^{-1}(Y_k) K_0\left(\frac{x - X_k}{h_k}\right) K_1\left(\frac{t - Y_k}{h_k}\right) \\ &+ \frac{\beta_n}{h_n^{d+1}} \delta_n \bar{G}_n^{-1}(Y_n) K_0\left(\frac{x - X_n}{h_n}\right) K_1\left(\frac{t - Y_n}{h_n}\right) \\ &= \alpha_1 \widehat{f}_{n_1}(x, t) + \sum_{k=n_1}^{n-1} \zeta_k \frac{\beta_k}{h_k^{d+1}} \delta_k \bar{G}_{n_2}^{-1}(Y_k) K_0\left(\frac{x - X_k}{h_k}\right) K_1\left(\frac{t - Y_k}{h_k}\right) \\ &+ \frac{\beta_n}{h_n^{d+1}} \delta_n \bar{G}_n^{-1}(Y_n) K_0\left(\frac{x - X_n}{h_n}\right) K_1\left(\frac{t - Y_n}{h_n}\right).\end{aligned}$$

Further, it follows from (2.4), that

$$\begin{aligned}\widehat{f}_n(x) &= \prod_{j=n_1+1}^n (1 - \gamma_j) \widehat{f}_{n_1}(x) + \sum_{k=n_1}^{n-1} \left[\prod_{j=k+1}^n (1 - \gamma_j) \right] \frac{\gamma_k}{h_k} K_0\left(\frac{x - X_k}{h_k}\right) \\ &+ \frac{\gamma_n}{h_n} \delta_n \bar{G}_n^{-1}(Y_n) K_0\left(\frac{x - X_n}{h_n}\right) \\ &= \tilde{\alpha}_1 \widehat{f}_{n_1}(x, t) + \sum_{k=n_1}^{n-1} \tilde{\zeta}_k \frac{\gamma_k}{h_k} K_0\left(\frac{x - X_k}{h_k}\right) + \frac{\gamma_n}{h_n} K_0\left(\frac{x - X_n}{h_n}\right),\end{aligned}$$

where $\tilde{\alpha}_1 = \prod_{j=n_1+1}^n (1 - \gamma_j)$ and $\tilde{\zeta}_k = \prod_{j=k+1}^n (1 - \gamma_j)$. It is clear, that we can use a data-driven procedure to construct an optimal bandwidth based on the first sample of size n_1 and separately an optimal bandwidth based on the second sample of size $n - n_1$. Then the proposed estimator can be viewed as a linear combination of two estimators, which improve the computational cost significantly.

4. Simulation experiments

In our simulation study, we consider three sample size, $n = 100$, $n = 200$ and $n = 300$ and the three models used in Khardani and Semmar (2014) : 1- sinus case : $Y = \sin(1.5X) + \varepsilon$, 2- parabolic case : $Y = 1 + X^2 + \varepsilon$, 3- exponential cas : $Y = \exp(X - 0.2) + \varepsilon$. In the three considered models, kernels K_0 and K_1 are considered to be the standard normal distributions $K_0(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$, $K_1(z_1, \dots, z_d) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} \sum_{i=1}^d z_i^2\right)$, X and ε are i.i.d. and follow respectively the normal distribution $\mathcal{N}(0, I_d)$ and $\mathcal{N}(0, 0.1I_d)$, with $d \in \{1, 2\}$. For each considered model and sample size n , we use various proportions of censoring levels, with the minimum equal to 0% and the maximum equal to 40% and we approximate the average MSE of the estimator and the standard deviations of the average MSE using $N = 500$

trials of sample size n ;

$$\overline{MSE} = \frac{1}{N} \sum_{k=1}^N \left\{ \widehat{f}^{[k]}(Y|X) - f(Y|X) \right\}^2$$

and

$$sdMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N \left\{ \left\{ \widehat{f}^{[k]}(Y|X) - f(Y|X) \right\}^2 - \overline{MSE} \right\}^2},$$

where $\widehat{f}^{[k]}(\cdot)$ is the estimator computed from the k th sample. In order to assess the impact of the censoring in the efficiency of the considered estimators we variate the rate of censoring for each models by generating the censoring variables C by an exponential distribution $\mathcal{E}(\lambda_1)$ shifted by λ_2 (for the exponential model), by a normal distribution $\mathcal{N}(0, \sigma_1)$ (for sinus case) and by $\mathcal{N}(0, \sigma_2)$ (for parabolic case). Thus, the behavior of the considered estimators is evaluated over a several parameters, such as the sample size n , the rate of censoring τ controlled by $(\lambda_1, \lambda_2, \sigma_1, \sigma_2)$ and the dimension d .

Remark 4.1. Since the optimal bandwidth (h_n) given in (3.8) belongs to $\mathcal{GS}(-1/(d+5))$, then according to the assumption **A5**(i), $a = 1/(d+5)$.

Moreover, in order to select the bandwidth selection, we consider a cross-validation procedure, defined as follows; firstly, we choose a sequence of bandwidth $h_i = C i^{-\nu}$ for $i = 1, \dots, n$, with $C \in \{0.5, 1, 1.5 \dots, 10\}$ and $\nu \in \{0.001, 0.002, \dots, 1\}$. Then, the cross-validation criterion

$$CV(C, \nu) = \frac{1}{n} \sum_{i=1}^n (\widehat{g}(Y|X) - \widehat{g}_{-i}(Y|X))^2,$$

where $\widehat{g}(Y|X)$ represents one of the considered estimators including the recursive and the non-recursive one, $\widehat{g}_{-i}(Y|X)$ represents one of the considered estimators using all the points except the i^{th} observation (X_i, Y_i) , for $i \in 1, \dots, n$.

The stepsizes $((\beta_n), (\gamma_n))$ equal respectively to $((n^{-1}), (n^{-1}))$, $((n^{-1}), ([1-a]n^{-1}))$, $([1-a]n^{-1}), (n^{-1})$ or $([1-a]n^{-1}), ([1-a]n^{-1})$. These four choices are referred to as **Recursive 1, 2, 3 and 4** respectively. Noting that the special cases $(\beta_n) = ([1-a]n^{-1})$, respectively $(\gamma_n) = ([1-a]n^{-1})$ represent the case which minimizes the variance of the numerator respectively the denominator of the proposed estimators (2.2) and the special cases $(\beta_n) = (n^{-1})$, respectively $(\gamma_n) = (n^{-1})$ represent the case which minimizes the MSE of the numerator respectively the denominator of the proposed estimators (2.2), see, for instance [Slaoui \(2015, 2016\)](#) in the framework of the regression function.

Remark 4.2. The choice $(\gamma_n) = ([1-a]n^{-1})$ is equivalent to the choice $(\gamma_n) = (h_n / \sum_{k=1}^n h_k)$, this two choices belong to the class of estimators which minimizes the asymptotic variance of the recursive kernel density estimators (for more details see [Mokkadem et al., 2009](#)).

$Y = \sin(1.5X) + \varepsilon$						
d=1	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.47169	0.34755	0.45528	0.34661	0.44123	0.34543
Recursive 1	0.49923	0.35809	0.49374	0.35087	0.48287	0.35026
Recursive 2	0.47490	0.34643	0.46969	0.34628	0.45679	0.34564
Recursive 3	0.50849	0.36673	0.49961	0.36614	0.48587	0.36146
Recursive 4	0.47133	0.34781	0.46427	0.34446	0.45537	0.34265
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.50374	0.34491	0.48050	0.34540	0.46866	0.33591
Recursive 1	0.52690	0.36925	0.50962	0.35586	0.46906	0.33225
Recursive 2	0.48405	0.34879	0.47177	0.34016	0.46035	0.33241
Recursive 3	0.56004	0.39488	0.53764	0.37361	0.52663	0.37624
Recursive 4	0.49605	0.35918	0.48073	0.34653	0.45601	0.33132
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.89366	0.79130	0.75139	0.63007	0.60624	0.53639
Recursive 1	0.69192	0.46411	0.60886	0.38604	0.50076	0.34334
Recursive 2	0.57606	0.38179	0.51977	0.34627	0.48482	0.33539
Recursive 3	0.80475	0.58158	0.68932	0.45206	0.60226	0.39245
Recursive 4	0.63578	0.43047	0.56176	0.36897	0.54207	0.34603
d=2	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.43293	0.33216	0.42792	0.32986	0.40528	0.31968
Recursive 1	0.45928	0.34876	0.45428	0.32207	0.44358	0.32157
Recursive 2	0.43594	0.31783	0.43217	0.31887	0.42936	0.31727
Recursive 3	0.46675	0.33567	0.45875	0.33465	0.44667	0.33284
Recursive 4	0.43367	0.32938	0.42628	0.32619	0.41823	0.31458
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.46249	0.32838	0.44318	0.32978	0.43234	0.32763
Recursive 1	0.48378	0.33898	0.46786	0.32674	0.43156	0.30498
Recursive 2	0.44438	0.32216	0.43345	0.31224	0.42456	0.34512
Recursive 3	0.51447	0.36287	0.49351	0.34494	0.48340	0.34536
Recursive 4	0.45593	0.32921	0.44165	0.31823	0.41856	0.31486
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.82731	0.72638	0.68974	0.57832	0.55649	0.49285
Recursive 1	0.63534	0.42625	0.55889	0.35437	0.45966	0.31562
Recursive 2	0.52872	0.35246	0.47745	0.31786	0.44556	0.30785
Recursive 3	0.73842	0.53584	0.63775	0.41959	0.55825	0.36416
Recursive 4	0.58615	0.39527	0.52365	0.33945	0.49782	0.31735

TABLE 4.1. Average MSE s and standard deviation of MSE (approximated using $N = 500$ trials) of the non-recursive estimator (2.1) and four recursive estimators; recursive 1 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), (n^{-1}))$, recursive 2 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), ([1-a]n^{-1}))$, recursive 3 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1-a]n^{-1}), (n^{-1}))$ and recursive 4 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1-a]n^{-1}), ([1-a]n^{-1}))$. Here we consider the sinus model $Y = \sin(1.5X) + \varepsilon$, $X \sim \mathcal{N}(0, I_d)$ and $\varepsilon \sim \mathcal{N}(0, 0.1I_d)$, with $d \in \{1, 2\}$ and we consider three sample sizes $n = 100$, $n = 200$ and $n = 300$.

$$Y = 1 + X^2 + \varepsilon$$

d=1	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.32769	0.31714	0.30391	0.31198	0.28309	0.31694
Recursive 1	0.34989	0.30718	0.36159	0.31387	0.28573	0.31105
Recursive 2	0.31787	0.31363	0.32957	0.30926	0.27265	0.30109
Recursive 3	0.37645	0.30476	0.38472	0.32429	0.27983	0.31104
Recursive 4	0.34012	0.30847	0.35229	0.31245	0.27637	0.31225
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.39926	0.34010	0.39073	0.33699	0.38957	0.32745
Recursive 1	0.44993	0.34646	0.44842	0.33971	0.44698	0.33567
Recursive 2	0.42835	0.33860	0.42232	0.33608	0.41742	0.33528
Recursive 3	0.46821	0.35069	0.46492	0.34198	0.46423	0.33839
Recursive 4	0.43682	0.34247	0.43437	0.33434	0.43444	0.33263
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.45167	0.37771	0.41885	0.37020	0.41249	0.37353
Recursive 1	0.48201	0.36498	0.46823	0.35601	0.45645	0.35189
Recursive 2	0.46431	0.34759	0.44149	0.34528	0.43427	0.33518
Recursive 3	0.51110	0.38720	0.49747	0.37411	0.48358	0.40111
Recursive 4	0.47373	0.36587	0.46035	0.35497	0.45824	0.35339
d=2	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.30189	0.29118	0.27796	0.28637	0.26332	0.29187
Recursive 1	0.32145	0.28698	0.33147	0.28963	0.26448	0.28341
Recursive 2	0.29182	0.28714	0.30456	0.28486	0.25043	0.27957
Recursive 3	0.34548	0.27784	0.35465	0.29865	0.25644	0.28549
Recursive 4	0.31264	0.28349	0.32776	0.28672	0.25386	0.28646
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.36629	0.31884	0.35843	0.30935	0.35685	0.30586
Recursive 1	0.41347	0.31828	0.41182	0.31285	0.41098	0.30289
Recursive 2	0.39458	0.31284	0.38756	0.30979	0.38162	0.30636
Recursive 3	0.42984	0.32089	0.42764	0.31138	0.42306	0.310483
Recursive 4	0.41872	0.31465	0.39848	0.30697	0.39858	0.30335
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.41464	0.34673	0.38485	0.33979	0.37854	0.34744
Recursive 1	0.44214	0.33261	0.42924	0.32623	0.41989	0.32104
Recursive 2	0.42644	0.31936	0.40526	0.31429	0.39496	0.30618
Recursive 3	0.46916	0.35525	0.45628	0.34366	0.44388	0.36845
Recursive 4	0.43486	0.33684	0.42269	0.32847	0.42632	0.32873

TABLE 4.2. Average MSE s and standard deviation of MSE (approximated using $N = 500$ trials) of the non-recursive estimator (2.1) and four recursive estimators; recursive 1 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), (n^{-1}))$, recursive 2 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), ([1 - a]n^{-1}))$, recursive 3 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1 - a]n^{-1}), (n^{-1}))$ and recursive 4 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1 - a]n^{-1}), ([1 - a]n^{-1}))$. Here we consider the parabolic model $Y = 1 + X^2 + \varepsilon$, $X \sim \mathcal{N}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 0.1)$ and we consider three sample sizes $n = 100$, $n = 200$ and $n = 300$.

$Y = \exp(X - 0.2) + \varepsilon$						
d=1	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.42833	0.36027	0.41430	0.35918	0.40384	0.35864
Recursive 1	0.43161	0.35150	0.42285	0.34915	0.41921	0.35558
Recursive 2	0.43280	0.35182	0.42327	0.34951	0.41728	0.33615
Recursive 3	0.42708	0.32801	0.40997	0.31353	0.39704	0.31252
Recursive 4	0.40618	0.34053	0.39809	0.34555	0.38464	0.34481
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.45615	0.35219	0.40481	0.34873	0.38593	0.34156
Recursive 1	0.43922	0.34366	0.42326	0.34238	0.36921	0.33958
Recursive 2	0.43969	0.34527	0.41813	0.33615	0.34570	0.31342
Recursive 3	0.42708	0.32801	0.41704	0.31812	0.37973	0.31553
Recursive 4	0.41945	0.33378	0.39439	0.31991	0.36854	0.31447
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.82192	0.72970	0.62419	0.45912	0.42344	0.43347
Recursive 1	0.66738	0.70964	0.59954	0.50787	0.49785	0.47703
Recursive 2	0.61095	0.53304	0.52550	0.41610	0.41229	0.40523
Recursive 3	1.01691	0.97070	0.83108	0.88199	0.64612	0.62362
Recursive 4	0.74369	0.69710	0.70782	0.56174	0.69365	0.52963
d=2	n = 100		n = 200		n = 300	
0%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.39773	0.33026	0.38084	0.32921	0.37969	0.32964
Recursive 1	0.39678	0.32228	0.38469	0.32048	0.38484	0.32635
Recursive 2	0.39743	0.32246	0.38846	0.32025	0.38338	0.30852
Recursive 3	0.39205	0.30102	0.37638	0.28776	0.36449	0.28615
Recursive 4	0.37248	0.31229	0.36582	0.31703	0.35326	0.31616
10%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.41833	0.32358	0.37172	0.32092	0.35469	0.31472
Recursive 1	0.40332	0.31563	0.38834	0.31426	0.33189	0.31107
Recursive 2	0.40346	0.31638	0.38344	0.30814	0.31784	0.28766
Recursive 3	0.39295	0.30102	0.38235	0.29219	0.34856	0.28348
Recursive 4	0.38557	0.30867	0.36224	0.29255	0.33839	0.28814
40%	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$	\overline{MSE}	$sdMSE$
non-recursive	0.75465	0.66984	0.57396	0.42144	0.38878	0.39895
Recursive 1	0.61268	0.65198	0.55058	0.46694	0.45695	0.43765
Recursive 2	0.56093	0.48924	0.48232	0.38197	0.37844	0.37128
Recursive 3	0.91222	0.89343	0.76283	0.80934	0.59912	0.57424
Recursive 4	0.68235	0.63989	0.64974	0.51628	0.63674	0.48614

TABLE 4.3. Average MSE s and standard deviation of MSE (approximated using $N = 500$ trials) of the non-recursive estimator (2.1) and four recursive estimators; recursive 1 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), (n^{-1}))$, recursive 2 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = ((n^{-1}), ([1 - a]n^{-1}))$, recursive 3 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1 - a]n^{-1}), (n^{-1}))$ and recursive 4 correspond to the estimator (2.2) with the choice $((\beta_n), (\gamma_n)) = (([1 - a]n^{-1}), ([1 - a]n^{-1}))$. Here we consider the exponential model $Y = \exp(X - 0.2) + \varepsilon$, $X \sim \mathcal{N}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 0.1)$ and we consider three sample sizes $n = 100$, $n = 200$ and $n = 300$.

We can conclude from Tables 4.1, 4.2 and 4.3 that

- In the three considered models, the $sdMSE$ of the recursive estimator give better results compared to the non-recursive estimator.
- In the third model, the recursive estimator outperformed the non-recursive estimator in all the considered censoring levels, while in the first and the second model, the recursive estimator outperformed the non-recursive estimator in all the considered censoring levels, with exception of the non-censoring case.
- In many situations, the proposed recursive estimators with the stepsizes $((\beta_n), (\gamma_n)) = ((n^{-1}), ([1 - a] n^{-1}))$ outperformed the non-recursive estimator (2.1) in terms of estimation error.
- The estimators get closer to the true density function as sample size increases and the proportion of the censoring level decreases.
- The Average $MSEs$ decrease as the sample size increases.
- The Average $MSEs$ increase as the percentage of censoring τ decreases.
- The Average $MSEs$ decrease as the dimension d increases.

5. Proof

Throughout this section we use the following notation:

$$\Pi_n = \prod_{i=1}^n (1 - \gamma_j), \quad Q_n = \prod_{j=1}^n (1 - \beta_j).$$

Let us first state the following technical lemma.

Lemma 5.1. *Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, $(\beta_n) \in \mathcal{GS}(-\beta)$, $m_1 > 0$ such that $m_1 - v^*\epsilon_1 > 0$ where ϵ_1 is defined in (3.2), $m_2 > 0$ and ϵ_2 is defined in (3.3). We have*

$$\begin{aligned} & \lim_{n \rightarrow +\infty} v_n \Pi_n^{m_1} Q_n^{m_2} \sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} \\ &= \begin{cases} \frac{1}{(cm_1+m_2-v^*\epsilon_2)} & \text{if } \gamma_n \equiv c\beta_n \\ \frac{1}{m_2} & \text{if } \gamma_n = o(\beta_n) \\ \frac{1}{m_1-v^*\epsilon_1} & \text{if } \beta_n = o(\gamma_n) \end{cases} \end{aligned} \tag{5.1}$$

Moreover, for all positive sequence (α_n) such that $\lim_{n \rightarrow +\infty} \alpha_n = 0$, and all $\delta \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} v_n \Pi_n^{m_1} Q_n^{m_2} \left[\sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0. \tag{5.2}$$

Lemma 5.2. *(Bias and variance of $f_n(x, t)$). Under Assumptions A1, A2, A4 and A5(i) – (iii), and assuming that $f^{(2)}(x, t)$ is continuous at (x, t) . If $a \in [0, \beta/(d + 5)]$, then*

$$\begin{aligned} & \mathbb{E} \left[\widehat{f}_n(x, t) \right] - f(x, t) \\ &= \frac{1}{2(1 - 2a\epsilon_2)} h_n^2 \left(\sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x, t) \right) + m_2 f^{(2)}(x, t) \right) + o(h_n^2). \end{aligned} \tag{5.3}$$

If $a \in]\beta/(d + 5), 1[$, then

$$\mathbb{E} \left[\widehat{f}_n(x, t) \right] - f(x, t) = o \left(\sqrt{\beta_n h_n^{-d-1}} \right). \tag{5.4}$$

If $a \in [\beta/(d + 5), 1[$, then

$$\begin{aligned} \text{Var} \left[\widehat{f}_n(x, t) \right] &= \frac{1}{(2 - (\beta - (d + 1)a) \epsilon_2)} \frac{\beta_n}{h_n^{d+1}} \frac{f(x, t)}{G(t)} \int_{\mathbb{R}} K_0^2(z) dz \int_{\mathbb{R}^d} K_1^2(y) dy. \end{aligned} \tag{5.5}$$

If $a \in [0, \beta/(d + 5)[$, then

$$\text{Var} \left[\widehat{f}_n(x, t) \right] = o(h_n^4). \tag{5.6}$$

Lemma 5.3. (Bias and variance of f_n). Under Assumptions **A3**, **A4**(i), **A5**(iv)–(v), and assuming that $f^{(2)}(x)$ is continuous at x .

If $a \in [0, \alpha/(d + 4)]$, then

$$\mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) = \frac{1}{2(\alpha - 2a\epsilon_1)} h_n^2 \sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x) \right) + o(h_n^2). \tag{5.7}$$

If $a \in]\alpha/(d + 4), 1[$, then

$$\mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) = o \left(\sqrt{\gamma_n h_n^{-d}} \right). \tag{5.8}$$

If $a \in [\alpha/(d + 4), 1[$, then

$$\text{Var} \left[\widehat{f}_n(x) \right] = \frac{1}{(2 - (\alpha - ad) \epsilon_1)} \frac{\gamma_n}{h_n^d} f(x) \int_{\mathbb{R}^d} K_0^2(z) dz. \tag{5.9}$$

If $a \in [0, \alpha/(d + 4)[$, then

$$\text{Var} \left[\widehat{f}_n(x) \right] = o(h_n^4).$$

Lemma 5.1 is widely applied throughout the proofs. Lemmas 5.1 and 5.2 are proved in Sections 5.1 and 5.2, respectively. The proof of Lemma 5.3 is given in Mokkadem et al. (2009).

5.1. Proof of Lemma 5.1. To prove Lemma 5.1, we first establish (5.2). Set

$$L_n = v_n \Pi_n^{m_1} Q_n^{m_2} \left[\sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0.$$

We have

$$L_n = \frac{v_n}{v_{n-1}} (1 - \gamma_n)^{m_1} (1 - \beta_n)^{m_2} L_{n-1} + \gamma_n \alpha_n \tag{5.10}$$

with, since $(v_n) \in \mathcal{GS}(v^*)$ and in view of (3.2),

$$\begin{aligned} \frac{v_n}{v_{n-1}} (1 - \gamma_n)^{m_1} &= \left(1 + \frac{v^*}{n} + o \left(\frac{1}{n} \right) \right) (1 - m_1 \gamma_n + o(\gamma_n)) \\ &= (1 - v^* \epsilon_1 \gamma_n + o(\gamma_n)) (1 - m_1 \gamma_n + o(\gamma_n)) \\ &= 1 - (m_1 - v^* \epsilon_1) \gamma_n + o(\gamma_n). \end{aligned}$$

and in view of (3.3),

$$(1 - \beta_n)^{m_2} = 1 - m_2\beta_n + o(\beta_n),$$

then, it follows that

$$\begin{aligned} & \frac{v_n}{v_{n-1}} (1 - \gamma_n)^{m_1} (1 - \beta_n)^{m_2} \\ &= (1 - (m_1 - v^*\epsilon_1)\gamma_n + o(\gamma_n)) (1 - m_2\beta_n + o(\beta_n)), \\ &= 1 - [(m_1 - v^*\epsilon_1)\gamma_n + m_2\beta_n] + o(\gamma_n + \beta_n) \end{aligned}$$

then,

$$\begin{aligned} & \frac{v_n}{v_{n-1}} (1 - \gamma_n)^{m_1} (1 - \beta_n)^{m_2} \\ &= \begin{cases} 1 - (cm_1 + m_2 - v^*\epsilon_2)\beta_n + o(\beta_n) & \text{if } \gamma_n \equiv c\beta_n \\ 1 - m_2\beta_n + o(\beta_n) & \text{if } \gamma_n = o(\beta_n) \\ 1 - (m_1 - v^*\epsilon_1)\gamma_n + o(\gamma_n) & \text{if } \beta_n = o(\gamma_n) \end{cases} \quad (5.11) \end{aligned}$$

and from (5.10), Set $A_1 \in]0, \min\{cm_1 + m_2 - v^*\epsilon_2; m_2\}[$ and $A_2 \in]0, m_1 - v^*\epsilon_1[$; for n large enough, we obtain

$$L_n \leq \begin{cases} (1 - A_1\beta_n)L_{n-1} + \gamma_n\alpha_n & \text{when } \lim_{n \rightarrow +\infty} \gamma_n\beta_n^{-1} \in [0, +\infty[\\ (1 - A_2\gamma_n)L_{n-1} + \gamma_n\alpha_n & \text{when } \lim_{n \rightarrow +\infty} \gamma_n\beta_n^{-1} = \infty \end{cases}$$

and (5.2) follows straightforwardly from the application of Lemma 4.1.1 in Dufflo (1997). Now, let:

$$\xi = \begin{cases} \frac{1}{(cm_1 + m_2 - v^*\epsilon_2)} & \text{if } \gamma_n \equiv c\beta_n \\ \frac{1}{m_2} & \text{if } \gamma_n = o(\beta_n) \\ \frac{1}{m_1 - v^*\epsilon_1} & \text{if } \beta_n = o(\gamma_n) \end{cases}$$

Moreover, we let C denote a positive generic constant that may vary from line to line; we have

$$v_n \Pi_n^{m_1} Q_n^{m_2} \sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} - \xi = v_n \Pi_n^{m_1} Q_n^{m_2} \left[\sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} - \xi P_n \right]$$

with, in view of (5.11),

$$\begin{aligned} P_n &= v_n^{-1} \Pi_n^{-m_1} Q_n^{-m_2} \\ &= \sum_{k=2}^n (v_k^{-1} \Pi_k^{-m_1} Q_k^{-m_2} - v_{k-1}^{-1} \Pi_{k-1}^{-m_1} Q_{k-1}^{-m_2}) + C \\ &= \sum_{k=2}^n v_k^{-1} \Pi_k^{-m_1} Q_k^{-m_2} \left[1 - \frac{v_k}{v_{k-1}} (1 - \gamma_k)^{m_1} (1 - \beta_k)^{m_2} \right] + C \\ &= \sum_{k=2}^n v_k^{-1} \Pi_k^{-m_1} Q_k^{-m_2} [\xi^{-1}(\gamma_k + \beta_k) + o(\gamma_k + \beta_k)] + C. \end{aligned}$$

It follows that

$$v_n \Pi_n^{m_1} Q_n^{m_2} \sum_{k=1}^n \Pi_k^{-m_1} Q_k^{-m_2} \frac{\gamma_k}{v_k} - \xi$$

$$= v_n \Pi_n^{m_1} Q_n^{m_2} \left[\sum_{k=2}^n v_k^{-1} \Pi_k^{-m_1} Q_k^{-m_2} o(\gamma_k + \beta_k) + C \right].$$

and (5.1) follows from the application of (5.2), which conclude the proof of Lemma 5.1.

5.2. *Proof of Lemma 5.2.* First, for x such that $\hat{f}_n(x) \neq 0$, we have

$$\begin{aligned} & \hat{f}_n(x, t) - f(x, t) \\ &= Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \delta_k \bar{G}_n^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) - f(x, t). \end{aligned}$$

The proof is standard, in the sense that is not affected by the dependence structure. Using the fact that for any measurable function ψ

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

First, it follows from (3.9), that

$$\tilde{f}_n(x, t) = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \delta_k \bar{G}^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right), \quad (5.12)$$

and we let

$$Z_k(x, t) = h_k^{-d-1} \mathbb{1}_{\{T_k \leq C_k\}} \bar{G}^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right). \quad (5.13)$$

Then, it follows that

$$\mathbb{E} \left[\tilde{f}_n(x, t) \right] - f(x, t) = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k \mathbb{E} \{ Z_k(x, t) \} - f(x, t).$$

The following Lemma 5.4 is widely used and plays a very important role in the proofs also in equation (5.14) (below) which shows the same asymptotic behavior of the two estimators $\hat{f}_n(x, t)$ and $\tilde{f}_n(x, t)$.

Lemma 5.4. (*Brunel and Comte, 2005*)

$$\mathbb{E} \left\{ \sup_{t \in A \subseteq [0, \tau]} |\bar{G}_n(t) - \bar{G}(t)| \right\} = O \left(\frac{1}{n} \right) \quad \text{as } n \rightarrow \infty.$$

First, using lemma 5.4, we have

$$\begin{aligned} & \mathbb{E} \left\{ \sup_{t \in A \subseteq [0, \tau]} |\hat{f}_n(x, t) - \tilde{f}_n(x, t)| \right\} \\ &= \mathbb{E} \left\{ \sup_{t \in A \subseteq [0, \tau]} \left| Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \delta_k \left(\bar{G}_n^{-1}(Y_k) - \bar{G}^{-1}(Y_k) \right) \right. \right. \\ & \quad \left. \left. K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \right| \right\} \\ &\leq \left\| K_1 \right\|_{\infty} \times \frac{\sup_{t \in A \subseteq [0, \tau]} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau + Dh_k) \bar{G}(\tau + Dh_k)} Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \mathbb{E} \left\{ K_0 \left(\frac{x - X_k}{h_k} \right) \right\} \end{aligned}$$

where D is related to the support of K_1 . Since $\bar{G}_n(\tau + Dh_k)$ and $\bar{G}(\tau + Dh_k) \rightarrow G(\tau) > 0$, we have under Lemmas 5.1 and 5.4

$$\mathbb{E} \left\{ \sup_{t \in A \subseteq [0, \tau]} \left| \hat{f}_n(x, t) - \tilde{f}_n(x, t) \right| \right\} = O \left(\frac{1}{nh_n^d} \right) = o(1). \tag{5.14}$$

Now, using the properties of conditional expectation, the independence between (X, T) and C , and Taylor's expansion with integral, we have

$$\begin{aligned} & \mathbb{E} \{ Z_k(x, t) \} - f(x, t) \\ &= h_k^{-d-1} \mathbb{E} \left\{ K_0 \left(\frac{x - X_1}{h_n} \right) \mathbb{E} \left[\mathbb{1}_{\{T_1 \leq C_1\}} \bar{G}^{-1}(Y_k) K_1 \left(\frac{t - Y_k}{h_k} \right) \middle| X_k, T_k \right] \right\} - f(x, t) \\ &= h_k^{-d-1} \mathbb{E} \left\{ \bar{G}^{-1}(T_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - T_k}{h_k} \right) \mathbb{E} \left[\mathbb{1}_{\{T_k \leq C_k\}} \middle| X_k, T_k \right] \right\} - f(x, t) \\ &= h_k^{-d-1} \mathbb{E} \left\{ K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - T_k}{h_k} \right) \right\} - f(x, t) \end{aligned} \tag{5.15}$$

$$\begin{aligned} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0(u) K_1(v) [f(x - h_k u, t - h_k v) - f(x, t)] dudv \\ &= \frac{h_k^2}{2} \left(\sum_{j=1}^d \mu_j^2 f_{j,j}^{(2)}(x, t) + m_2 f^{(2)}(x, t) \right) + \eta_k(x, t), \end{aligned} \tag{5.16}$$

with

$$\begin{aligned} \eta_k(x, t) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0(r) K_1(s) \left\{ f(x - h_k r, t - h_k s) - f(x, t) + rh_k \sum_{j=1}^d \frac{\partial f}{\partial x_j}(x, t) \right. \\ &+ sh_k \frac{\partial f}{\partial t}(x, t) - \frac{h_k^2}{2} \sum_{j=1}^d r_j^2 \frac{\partial^2 f}{\partial^2 x_j}(x, t) - sh_k^2 \sum_{j=1}^d r_j \frac{\partial^2 f}{\partial x_j \partial t}(x, t) \\ &\left. - s^2 \frac{h_k^2}{2} \frac{\partial^2 f}{\partial^2 t}(x, t) \right\} dr ds. \end{aligned}$$

Since $f_{i,j}^{(2)}$ is bounded and continuous at (x, t) for all $i, j \in \{1, \dots, d\}$, we have $\lim_{k \rightarrow \infty} \eta_k(x, t) = 0$. In the case when $a \leq \beta / (d + 5)$, we have $\lim_{n \rightarrow \infty} (n\beta_n) > 2a$, the application of Lemma 5.1, gives

$$\begin{aligned} & \mathbb{E} \left[\tilde{f}_n(x, t) \right] - f(x, t) \\ &= \frac{1}{2} Q_n \sum_{k=1}^n Q_k^{-1} \beta_k \left[\frac{h_k^2}{2} \left(\sum_{j=1}^d \mu_j^2 f_{j,j}^{(2)}(x, t) + m_2 f^{(2)}(x, t) + o(1) \right) \right] \\ &= \frac{1}{2(1 - 2a\epsilon_2)} h_n^2 \left[\sum_{j=1}^d \mu_j^2 f_{j,j}^{(2)}(x, t) + m_2 f^{(2)}(x, t) + o(1) \right], \end{aligned}$$

and (5.3) follows from (5.14). In the case when $a > \beta / (d + 5)$, we have $h_n^2 = o(\sqrt{\beta_n h_n^{-d-1}})$; since $\lim_{n \rightarrow \infty} (n\beta_n) > (\beta - (d + 1)a) / 2$, the application of

Lemma 5.1 ensures that

$$\begin{aligned}\mathbb{E} \left[\tilde{f}_n(x, t) \right] - f(x, t) &= \frac{1}{2} Q_n \sum_{k=1}^n Q_k^{-1} \beta_k o \left(\sqrt{\beta_k h_k^{-d-1}} \right) \\ &= o \left(\sqrt{\beta_n h_n^{-d-1}} \right),\end{aligned}$$

then, (5.4) follows from (5.14). Now, we have

$$\begin{aligned}\text{Var} \left[\tilde{f}_n(x, t) \right] &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 \text{Var} \left[Z_k(x, t) \right] \\ &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 h_k^{-(d+1)} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}} \frac{K_0^2(z) K_1^2(y)}{\overline{G}(t - yh_k)} f(x - zh_k, t - yh_k) \right. \\ &\quad \left. - h_k^{d+1} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0(z) K_1(y) f(x - zh_k, t - yh_k) \right)^2 \right] dz dy \\ &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 h_k^{-(d+1)} \left[\frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy + v_k(x, t) \right. \\ &\quad \left. - h_k^{d+1} \tilde{v}_k(x, t) \right]\end{aligned}$$

with

$$\begin{aligned}v_k(x, t) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) \left[\frac{f(x - zh_k, t - yh_k)}{\overline{G}(t - yh_k)} - \frac{f(x, t)}{\overline{G}(t)} \right] dz dy \\ \tilde{v}_k(x, t) &= \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0(z) K_1(y) f(x - zh_k, t - yh_k) \right)^2.\end{aligned}$$

Since \overline{G} and f are bounded and continuous, we have $\lim_{k \rightarrow \infty} v_k(x) = 0$ and $\lim_{k \rightarrow \infty} h_k^{d+1} \tilde{v}_k(x) = 0$. In the case when $a \geq \beta / (d + 5)$, we have $\lim_{n \rightarrow \infty} (n\beta_n) > (\beta - (d + 1)a) / 2$, and the application of Lemma 5.1 gives

$$\begin{aligned}\text{Var} \left[\tilde{f}_n(x, t) \right] &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k^2 h_k^{-(d+1)} \left[\frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy + o(1) \right] \\ &= \frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy \frac{1}{(2 - (\beta - (d + 1)a)\epsilon_2) h_n^{d+1}} + o \left(\frac{\beta_n}{h_n^{d+1}} \right),\end{aligned}$$

then, (5.5) follows from (5.14). In the case when $a < \beta / (d + 5)$, we have $\beta_n h_n^{-(d+1)} = o(h_n^4)$; since $\lim_{n \rightarrow \infty} (n\beta_n) > 2a$, the application of Lemma 5.1 ensures that

$$\begin{aligned}\text{Var} \left[\tilde{f}_n(x, t) \right] &= Q_n^2 \sum_{k=1}^n Q_k^{-2} \beta_k o(h_k^4) \\ &= o(h_n^4),\end{aligned}$$

then, (5.6) follows from (5.14).

5.3. Proof of Proposition 3.2.

Let us first note that, for x such that $\widehat{f}_n(x) \neq 0$, we have

$$\widehat{f}_n(t|x) - f(t|x) = B_n(x) \frac{f(x)}{\widehat{f}_n(x)}, \quad (5.17)$$

with

$$\begin{aligned} B_n(x) &= \frac{1}{f(x)} \left[\widehat{f}_n(x, t) - \widetilde{f}_n(x, t) \right] + \frac{1}{f(x)} \left[\widetilde{f}_n(x, t) \right. \\ &\quad \left. - f(x, t) \right] - \frac{f(t|x)}{f(x)} \left[\widehat{f}_n(x) - f(x) \right]. \end{aligned}$$

It follows from (5.17), that the asymptotic behavior of $\widehat{f}_n(t|x) - f(t|x)$ can be deduced from the one of $B_n(x)$. First, it follows from (5.3) and (5.7), in the case when $a \leq \beta / (d + 5)$ (thanks to the assumption **A5**(iv)) that

$$\begin{aligned} \mathbb{E}[\widehat{f}_n(t|x)] - f(t|x) &= \frac{1}{f(x)} \left[\frac{1}{2(1 - 2a\epsilon_2)} \left(\sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x, t) \right) + m_2 f^{(2)}(x, t) \right) h_n^2 \right] \\ &\quad - \frac{f(t|x)}{f(x)} \frac{h_n^2}{2(1 - 2a\epsilon_1)} \sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x) \right) + o(h_n^2), \end{aligned}$$

which proves (3.4). Now, in the case when $a > \beta / (d + 5)$, the combination of (5.4) and (5.8) gives (3.5).

Now, in the case when $a \geq \beta / (d + 5)$, the combinations of (5.17), (5.5), (5.9), (5.14) ensures that,

$$\begin{aligned} &\text{Var} \left(\widehat{f}_n(t|x) \right) \\ &= \frac{1}{f^2(x)} \left[\frac{f(x, t)}{\overline{G}(t)} \right] \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy \frac{1}{2 - (\beta - (d + 1)a)\epsilon_2} \frac{\beta_n}{h_n^{d+1}} \\ &\quad + \frac{f^2(t|x)}{f^2(x)} \left[f(x) \int_{\mathbb{R}^d} K_0^2(z) dz \frac{1}{2 - (1 - ad)\epsilon_2} \right] \frac{\gamma_n}{h_n^d} \\ &\quad - \frac{f(t|x)}{f^2(x)} \text{cov} \left(\widetilde{f}_n(x, t), \widehat{f}_n(x) \right) + o \left(\frac{\beta_n}{h_n^{d+1}} \right), \end{aligned} \quad (5.18)$$

with

$$\text{cov} \left(\widetilde{f}_n(x, t), \widehat{f}_n(x) \right) = \mathbb{E}[\widetilde{f}_n(x, t) \widehat{f}_n(x)] - \mathbb{E}[\widetilde{f}_n(x, t)] \mathbb{E}[\widehat{f}_n(x)].$$

Moreover, it follows from (2.5) and (2.6), that

$$\begin{aligned} &\widetilde{f}_n(x, t) \widehat{f}_n(x) \\ &= Q_n \sum_{k=1}^n Q_k^{-1} \beta_k h_k^{-d-1} \delta_k \overline{G}^{-1}(Y_k) K_0 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \\ &\quad \times \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-d} K_0 \left(\frac{x - X_k}{h_k} \right) \\ &= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k \gamma_k h_k^{-2d-1} \delta_k \overline{G}^{-1}(Y_k) K_0^2 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \end{aligned}$$

$$\begin{aligned}
& + Q_n \Pi_n \sum_{\substack{i,j=1 \\ i \neq j}}^n Q_i^{-1} \Pi_j^{-1} \beta_i \gamma_j h_i^{-d-1} h_j^{-d} \delta_i \overline{G}^{-1}(Y_k) K_0 \left(\frac{x - X_i}{h_i} \right) \\
& K_0 \left(\frac{x - X_j}{h_j} \right) K_1 \left(\frac{t - Y_i}{h_i} \right) \\
& =: \Theta_n^1 + \Theta_n^2.
\end{aligned}$$

Then, we have

$$\mathbb{E}[\tilde{f}_n(x, t) \hat{f}_n(x)] = \mathbb{E}[\Theta_n^1] + \mathbb{E}[\Theta_n^2].$$

First, the application of Lemma 5.1 ensures that,

$$\begin{aligned}
\mathbb{E}[\Theta_n^1] &= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k \gamma_k h_k^{-2d-1} \mathbb{E} \left[\delta_k \overline{G}^{-1}(Y_k) K_0^2 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \right] \\
&= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k \gamma_k h_k^{-2d-1} \mathbb{E} \left[K_0^2 \left(\frac{x - X_k}{h_k} \right) K_1 \left(\frac{t - Y_k}{h_k} \right) \right. \\
&\quad \left. \mathbb{E} \left[\delta_k \overline{G}^{-1}(Y_k) | (X, T) \right] \right] \\
&= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k \gamma_k h_k^{-2d-1} \\
&\quad \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2 \left(\frac{x - z}{h_k} \right) K_1 \left(\frac{t - y}{h_k} \right) f(z, y) dz dy \\
&= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k \gamma_k h_k^{-d} \\
&\quad \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(u) K_1(v) f(x - uh_k, t - vh_k) dudv \tag{5.19} \\
&= Q_n \Pi_n \sum_{k=1}^n Q_k^{-1} \Pi_k^{-1} \beta_k O(\gamma_k h_k^{-d}) \\
&= o \left(\frac{\beta_n}{h_n^d} \right). \tag{5.20}
\end{aligned}$$

Moreover, a conditional expectation ensures that

$$\begin{aligned}
& \mathbb{E}[\Theta_n^2] \\
&= \mathbb{E} \left[Q_n \Pi_n \sum_{i,j=1,}^n Q_i^{-1} \Pi_j^{-1} \beta_i \gamma_j h_i^{-d-1} h_j^{-d} \delta_i \overline{G}^{-1}(Y_i) K_0 \left(\frac{x - X_i}{h_i} \right) \right. \\
&\quad \left. K_0 \left(\frac{x - X_j}{h_j} \right) K_1 \left(\frac{t - Y_i}{h_i} \right) \right] \\
&= Q_n \Pi_n \sum_{\substack{i,j=1 \\ i \neq j}}^n Q_i^{-1} \Pi_j^{-1} \beta_i \gamma_j h_i^{-d-1} h_j^{-d} \\
&\quad \times \mathbb{E} \left[K_0 \left(\frac{x - X_i}{h_i} \right) K_0 \left(\frac{x - X_j}{h_j} \right) K_1 \left(\frac{t - Y_i}{h_i} \right) \mathbb{E} \left\{ \delta_i \overline{G}^{-1}(Y_i) | (X_i, X_j, Y_i) \right\} \right]
\end{aligned}$$

$$\begin{aligned}
&= Q_n \Pi_n \sum_{\substack{i,j=1 \\ i \neq j}}^n Q_i^{-1} \Pi_j^{-1} \beta_i \gamma_j h_i^{-d-1} h_j^{-d} \underbrace{\mathbb{E} \left[K_0 \left(\frac{x - X_i}{h_i} \right) K_1 \left(\frac{t - Y_i}{h_i} \right) \right]}_{\varphi_n^1} \\
&\times \underbrace{\mathbb{E} \left[K_0 \left(\frac{x - X_j}{h_j} \right) \right]}_{\varphi_n^2}. \tag{5.21}
\end{aligned}$$

The combination of (5.15) and (5.16), implies that

$$\varphi_n^1 = h_i^{d+1} \left[f(x, t) + \frac{h_i^2}{2} \left(\sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x, t) \right) + m_2 f^{(2)}(x, t) \right) + o(h_i^2) \right]. \tag{5.22}$$

Moreover, a Taylor expansion ensures that

$$\varphi_n^2 = h_j^d \left[f(x) + \frac{h_j^2}{2} \sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x) \right) + o(h_j^2) \right]. \tag{5.23}$$

Then, the application of Lemma 5.1 together with (5.21), (5.22), (5.23), gives

$$\mathbb{E}[\Theta_n^2] - \mathbb{E}[\widehat{f}_n(x, t)] \mathbb{E}[\widehat{f}_n(x)] = o(h_n^4). \tag{5.24}$$

Consequently, in the case when $a \geq \beta / (d + 5)$, the combination of (5.18), (5.19), (5.14) and (5.24) ensures that,

$$\begin{aligned}
\text{Var} \left(\widehat{f}_n(t|x) \right) &= \frac{1}{f^2(x)} \left[\frac{f(x, t)}{G(t)} \right] \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy \\
&\quad \frac{1}{(2 - (\beta - (d + 1)a) \epsilon_2)} \frac{\beta_n}{h_n^{d+1}} + o \left(\frac{\beta_n}{h_n^{d+1}} \right),
\end{aligned}$$

which proves (3.7). Now, in the case when $a < \beta / (d + 5)$, we have $\beta_n h_n^{-d-1} = o(h_n^4)$, and the application of Lemma 5.1 ensures that

$$\text{Var} \left(\widehat{f}_n(t|x) \right) = o(h_n^4),$$

which gives (3.6).

5.4. *Proof of Theorem 3.3.* Let us first state the following corollaries, which give the weak convergence rate of the estimator $\widehat{f}_n(x)$ defined in (2.6) and the weak convergence rate of the estimator $\widehat{f}_n(x, t)$ defined in (2.5) respectively.

Corollary 5.5 (Weak pointwise convergence rate). *Let Assumptions A3, A4(i), A5(iv) – (v) hold, and assume that $f^{(2)}(x)$ is continuous at x .*

(1) *If there exists $c \geq 0$ such that $\gamma_n^{-1} h_n^{d+4} \rightarrow c$, then*

$$\begin{aligned}
&\sqrt{\beta_n^{-1} h_n^{d+1}} \left(\widehat{f}_n(x) - f(x) \right) \\
&\xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\sqrt{c}}{2} \frac{1}{2(\alpha - 2a\epsilon_1)} \sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x) \right), \frac{1}{(2 - (\alpha - ad) \epsilon_1)} f(x) \int_{\mathbb{R}^d} K_0^2(z) dz \right).
\end{aligned}$$

(2) If $\gamma_n^{-1}h_n^{d+4} \rightarrow \infty$, then

$$\frac{1}{h_n^2} \left(\widehat{f}_n(x) - f(x) \right) \xrightarrow{\mathbb{P}} \frac{1}{2(\alpha - 2a\epsilon_1)} \sum_{j=1}^d \left(\mu_j^2 f_{j,j}^{(2)}(x) \right).$$

Corollary 5.6 (Weak pointwise convergence rate). *Let Assumptions **A1**, **A2**, **A4** and **A5(i) – (iii)** hold, and assume that $f^{(2)}(x, t)$ is continuous at (x, t) .*

(1) If there exists $c \geq 0$ such that $\beta_n^{-1}h_n^{d+5} \rightarrow c$, then

$$\sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(x, t) - f(x, t) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\sqrt{c}}{2} E_1^{1/2}, E_2 \right).$$

(2) If $\beta_n^{-1}h_n^{d+5} \rightarrow \infty$, then

$$\frac{1}{h_n^2} \left(\widehat{f}_n(x, t) - f(x, t) \right) \xrightarrow{\mathbb{P}} \frac{1}{2} C_1^{1/2}.$$

Corollary 5.6 is proved in Section 5.4.1, while the proof of Corollary 5.5 is given in Mokkadem et al. (2009).

Now, it follows from (5.17), for x such that $f_n(x) \neq 0$, that

$$\begin{aligned} & \sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(t|x) - f(t|x) \right) \\ &= \frac{1}{f(x)} \sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(x, t) - f(x, t) \right) - \frac{f(t|x)}{f(x)} \sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(x) - f(x) \right) \\ &= \frac{1}{f(x)} \sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(x, t) - f(x, t) \right) \\ &\quad - \frac{f(t|x)}{f(x)} \sqrt{\beta_n^{-1}h_n \gamma_n} \sqrt{\gamma_n h_n^d} \left(f_n(x) - \mathbb{E} \left[\widehat{f}_n(x) \right] \right) \\ &\quad - \frac{f(t|x)}{f(x)} \sqrt{\beta_n^{-1}h_n^{d+1}} \left(\mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) \right). \end{aligned} \quad (5.25)$$

Part 1 of Theorem 3.3 follows from the combination of (5.7), (5.25) and the first part of Corollary 5.6.

Moreover, it follows from (5.17) that, for x such that $f_n(x) \neq 0$, we have

$$\begin{aligned} h_n^{-2} \left(\widehat{f}_n(t|x) - f(t|x) \right) &= \frac{1}{f(x)} h_n^{-2} \left(\widehat{f}_n(x, t) - f(x, t) \right) \\ &\quad - \frac{f(t|x)}{f(x)} h_n^{-2} \left(\widehat{f}_n(x) - f(x) \right). \end{aligned} \quad (5.26)$$

Part 2 of Theorem 3.3 follows from the combination of (5.26), the second part of Corollary 5.5 and the second part of Corollary 5.6.

5.4.1. *Proof of Corollary 5.6.* Let us first assume that, if $a \geq \beta/(d+5)$, then

$$\sqrt{\beta_n^{-1}h_n^{d+1}} \left(\widehat{f}_n(x, t) - \mathbb{E} \left[\widehat{f}_n(x, t) \right] \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, E_2). \quad (5.27)$$

In the case when $a > \beta/(d+5)$, Part 1 of Corollary 5.6 follows from the combination of (5.4) and (5.27). In the case when $a = \beta/(d+5)$, Parts 1 and 2 of Corollary 5.6 follow from the combination of (5.3) and (5.27). In the case when $a < \beta/(d+5)$, (5.6) implies that

$$h_n^{-2} \left(\widehat{f}_n(x, t) - \mathbb{E} \left[\widehat{f}_n(x, t) \right] \right) \xrightarrow{\mathcal{D}} 0,$$

and the application of (5.3) gives Part 2 of Corollary 5.6. We now prove (5.27). In view of (2.5) and (5.13), we have

$$\widehat{f}_n(x, t) - \mathbb{E}[\widehat{f}_n(x, t)] = Q_n \sum_{k=1}^n Q_k^{-1} \beta_k (Z_k(x, t) - \mathbb{E}[Z_k(x, t)]).$$

Set

$$Y_k(x, t) = Q_k^{-1} \beta_k (Z_k(x, t) - \mathbb{E}[Z_k(x, t)]).$$

The application of Lemma 5.1 ensures that

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n \text{Var}(Y_k(x, t)) \\ &= \sum_{k=1}^n Q_k^{-2} \beta_k^2 \text{Var}(Z_k(x, t)) \\ &= \sum_{k=1}^n Q_k^{-2} \beta_k^2 h_k^{-(d+1)} \left[\frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy + o(1) \right] \\ &= \frac{1}{Q_n^2} \frac{\beta_n}{h_n^{d+1}} \left[\frac{1}{2 - (\beta - a(d+1)) \epsilon_2} \frac{f(x, t)}{\overline{G}(t)} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_0^2(z) K_1^2(y) dz dy + o(1) \right]. \end{aligned}$$

On the other hand, we have, for all $p > 0$,

$$\mathbb{E}[|Z_k(x, t)|^{2+p}] = O\left(\frac{1}{h_k^{(d+1)(1+p)}}\right).$$

By using the fact that $\lim_{n \rightarrow \infty} (n\beta_n) > (\beta - a(d+1))/2$, implies that there exists $p > 0$ such that

$$\lim_{n \rightarrow \infty} (n\beta_n) > \frac{1+p}{2+p} (\beta - a(d+1)).$$

The application of Lemma 5.1, ensures that

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}[|Y_k(x, t)|^{2+p}] &= O\left(\sum_{k=1}^n Q_k^{-2-p} \beta_k^{2+p} \mathbb{E}[|Z_k(x, t)|^{2+p}]\right) \\ &= O\left(\sum_{k=1}^n \frac{\Pi_k^{-2-p} \beta_k^{2+p}}{h_k^{(d+1)(1+p)}}\right) \\ &= O\left(\frac{\beta_n^{1+p}}{Q_n^{2+p} h_n^{(d+1)(1+p)}}\right), \end{aligned}$$

from which we conclude that

$$\frac{1}{v_n^{2+p}} \sum_{k=1}^n \mathbb{E}[|Y_k(x, t)|^{2+p}] = O\left([\beta_n h_n^{-d-1}]^{p/2}\right) = o(1).$$

The convergence in (5.27) then follows from the application of Lyapounov's Theorem.

6. Conclusion

This paper proposes an adaptive recursive kernel conditional density estimation under censored data (2.2). The proposed estimators follow asymptotically the normal distribution. The proposed estimators are compared to the non-recursive conditional density estimator for censored data (2.1). We showed that using a specific cross-validation bandwidth selection and some particularly stepsizes, the proposed recursive estimators often provides better results compared to the non-recursive conditional density estimator in terms of estimation error.

However the main advantage of the recursive method is that it runs considerably faster than the classical one, see for instance, Slaoui (2014a) in the framework of density estimation, and Slaoui (2014b) in the framework of distribution estimation.

We plan to make an extensions of our proposed estimators by considering Bernstein polynomials rather than kernels and to propose an adaptation of the estimators developed in Jmaei et al. (2017) and Slaoui and Jmaei (2019) in the context of conditional density estimation under censoring data.

In conclusion, the proposed adaptive recursive estimators often provides better results compared to the non-recursive under censoring data in terms of estimation error and much better in terms of computational costs.

Acknowledgments

The authors would like to thank the Editor, the Associate Editor and the referees for their very helpful comments, which led to considerable improvement of the original version of the paper and a more sharply focused presentation.

References

- Bashtannyk, D. M. and Hyndman, R. J. Bandwidth selection for kernel conditional density estimation. *Comput. Statist. Data Anal.*, **36** (3), 279–298 (2001). [MR1836204](#).
- Beran, R. Nonparametric regression with randomly censored survival data. Technical report, Technical university of California, Berkeley (1981).
- Bertin, K., Lacour, C., and Rivoirard, V. Adaptive pointwise estimation of conditional density function. *Ann. Inst. Henri Poincaré Probab. Stat.*, **52** (2), 939–980 (2016). [MR3498017](#).
- Blum, J. R. Multidimensional stochastic approximation methods. *Ann. Math. Statistics*, **25**, 737–744 (1954). [MR65092](#).
- Bojanic, R. and Seneta, E. A unified theory of regularly varying sequences. *Math. Z.*, **134**, 91–106 (1973). [MR333082](#).
- Bouaziz, O. and Lopez, O. Conditional density estimation in a censored single-index regression model. *Bernoulli*, **16** (2), 514–542 (2010). [MR2668913](#).
- Brunel, E. and Comte, F. Penalized contrast estimation of density and hazard rate with censored data. *Sankhyā*, **67** (3), 441–475 (2005). [MR2235573](#).
- Brunel, E., Comte, F., and Lacour, C. Adaptive estimation of the conditional density in the presence of censoring. *Sankhyā*, **69** (4), 734–763 (2007). [MR2521231](#).
- Chen, H. Lower rate of convergence for locating a maximum of a function. *Ann. Statist.*, **16** (3), 1330–1334 (1988). [MR959206](#).

- Davies, H. I. Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.*, **15** (3-4), 49–54 (1972/73). [MR334390](#).
- Deheuvels, P. Sur l'estimation séquentielle de la densité. *C. R. Acad. Sci. Paris Sér. A-B*, **276**, A1119–A1121 (1973). [MR334393](#).
- Devroye, L. P. On the pointwise and the integral convergence of recursive kernel estimates of probability densities. *Utilitas Math.*, **15**, 113–128 (1979). [MR531624](#).
- Dippon, J. Accelerated randomized stochastic optimization. *Ann. Statist.*, **31** (4), 1260–1281 (2003). [MR2001650](#).
- Dippon, J. and Renz, J. Weighted means in stochastic approximation of minima. *SIAM J. Control Optim.*, **35** (5), 1811–1827 (1997). [MR1466929](#).
- Dąbrowska, D. M. Nonparametric regression with censored survival time data. *Scand. J. Statist.*, **14** (3), 181–197 (1987). [MR932943](#).
- Dąbrowska, D. M. Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.*, **17** (3), 1157–1167 (1989). [MR1015143](#).
- Dufflo, M. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin (1996). ISBN 3-540-60699-8. [MR1612815](#).
- Dufflo, M. *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin (1997). ISBN 3-540-57100-0. [MR1485774](#).
- Efromovich, S. Conditional density estimation in a regression setting. *Ann. Statist.*, **35** (6), 2504–2535 (2007). [MR2382656](#).
- Efromovich, S. Dimension reduction and adaptation in conditional density estimation. *J. Amer. Statist. Assoc.*, **105** (490), 761–774 (2010). [MR2724859](#).
- Fabian, V. Stochastic approximation of minima with improved asymptotic speed. *Ann. Math. Statist.*, **38**, 191–200 (1967). [MR207136](#).
- Fan, J., Yao, Q., and Tong, H. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, **83** (1), 189–206 (1996). [MR1399164](#).
- Fan, J. and Yim, T. H. A crossvalidation method for estimating conditional densities. *Biometrika*, **91** (4), 819–834 (2004). [MR2126035](#).
- Galambos, J. and Seneta, E. Regularly varying sequences. *Proc. Amer. Math. Soc.*, **41**, 110–116 (1973). [MR323963](#).
- Hall, P. and Heyde, C. C. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (1980). ISBN 0-12-319350-8. [MR624435](#).
- Hall, P. and Patil, P. On the efficiency of on-line density estimators. *IEEE Transactions on Information Theory*, **40** (5), 1504–1512 (1994). DOI: [10.1109/18.333864](#).
- Hall, P., Racine, J., and Li, Q. Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.*, **99** (468), 1015–1026 (2004). [MR2109491](#).
- Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.*, **5** (4), 315–336 (1996). [MR1422114](#).
- Hyndman, R. J. and Yao, Q. Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonparametr. Stat.*, **14** (3), 259–278 (2002). [MR1905751](#).
- Jmaei, A., Slaoui, Y., and Dellagi, W. Recursive distribution estimator defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr.*

- Stat.*, **29** (4), 792–805 (2017). [MR3740720](#).
- Kaplan, E. L. and Meier, P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481 (1958). [MR93867](#).
- Khardani, S. and Semmar, S. Nonparametric conditional density estimation for censored data based on a recursive kernel. *Electron. J. Stat.*, **8** (2), 2541–2556 (2014). [MR3285875](#).
- Khardani, S. and Slaoui, Y. Nonparametric relative regression under random censorship model. *Statist. Probab. Lett.*, **151**, 116–122 (2019). [MR3944666](#).
- Kohler, M., Máthé, K., and Pintér, M. Prediction from randomly right censored data. *J. Multivariate Anal.*, **80** (1), 73–100 (2002). [MR1889835](#).
- Kushner, H. J. and Clark, D. S. *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin (1978). ISBN 0-387-90341-0. [MR499560](#).
- Liang, H.-Y. and Liu, A.-A. Kernel estimation of conditional density with truncated, censored and dependent data. *J. Multivariate Anal.*, **120**, 40–58 (2013). [MR3072717](#).
- Lo, S. H., Mack, Y. P., and Wang, J. L. Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Related Fields*, **80** (3), 461–473 (1989). [MR976536](#).
- Marron, J. S. and Padgett, W. J. Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Ann. Statist.*, **15** (4), 1520–1535 (1987). [MR913571](#).
- Mokkadem, A. and Pelletier, M. A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist.*, **35** (4), 1749–1772 (2007). [MR2351104](#).
- Mokkadem, A., Pelletier, M., and Slaoui, Y. The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference*, **139** (7), 2459–2478 (2009). [MR2508006](#).
- Polonik, W. and Yao, Q. Conditional minimum volume predictive regions for stochastic processes. *J. Amer. Statist. Assoc.*, **95** (450), 509–519 (2000). [MR1803169](#).
- Révész, P. Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes. I. *Studia Sci. Math. Hungar.*, **8**, 391–398 (1973). [MR373198](#).
- Révész, P. How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Math. Operationsforsch. Statist. Ser. Statist.*, **8** (1), 119–126 (1977). [MR501557](#).
- Robbins, H. and Monro, S. A stochastic approximation method. *Ann. Math. Statistics*, **22**, 400–407 (1951). [MR42668](#).
- Roussas, G. G. and Tran, L. T. Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Ann. Statist.*, **20** (1), 98–120 (1992). [MR1150336](#).
- Ruppert, D. Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Probab.*, **10** (1), 178–187 (1982). [MR637384](#).
- Slaoui, Y. Large and moderate deviation principles for recursive kernel density estimators defined by stochastic approximation method. *Serdica Math. J.*, **39** (1), 53–82 (2013). [MR3087768](#).

- Slaoui, Y. Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *J. Probab. Stat.*, pp. Art. ID 739640, 11 (2014a). [MR3219439](#).
- Slaoui, Y. The stochastic approximation method for estimation of a distribution function. *Math. Methods Statist.*, **23** (4), 306–325 (2014b). [MR3295061](#).
- Slaoui, Y. Plug-in bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.*, **69** (4), 483–509 (2015). [MR3414708](#).
- Slaoui, Y. Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface*, **9** (3), 375–388 (2016). [MR3457504](#).
- Slaoui, Y. and Jmaei, A. Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials. *Stat. Interface*, **12** (3), 439–455 (2019). [MR3964528](#).
- Tjøstheim, D. Non-linear time series: a selective review. *Scand. J. Statist.*, **21** (2), 97–130 (1994). [MR1294588](#).
- Wegman, E. J. and Davies, H. I. Remarks on some recursive estimators of a probability density. *Ann. Statist.*, **7** (2), 316–327 (1979). [MR520242](#).
- Wolverton, C. T. and Wagner, T. J. Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory*, **IT-15**, 258–265 (1969). [MR275576](#).
- Yamato, H. Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist.*, **14**, 1–12; correction, *ibid.* 15 (1972), 133 (1970/71). [MR381187](#).