STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **11**, March 2023, pp 196–215.
Published online in International Academic Press (www.IAPress.org)

Nonparametric Recursive Kernel Type Eestimators for the Moment Generating Function Under Censored Data

Salim Bouzebda[1], Issam Elhattab[1], Yousri Slaoui[2,*] and NourElhouda Taachouch[1]

[1] *LMAC, Université de Technologie de Compiégne, France*
[2] *Univ. Poitiers, Lab. Math. et Appl., Futuroscope Chasseneuil, France*

Abstract We are mainly concerned with kernel-type estimators for the moment-generating function in the present paper. More precisely, we establish the central limit theorem with the characterization of the bias and the variance for the nonparametric recursive kernel-type estimators for the moment-generating function under some mild conditions in the censored data setting. Finally, we investigate the methodology's performance for small samples through a short simulation study.

Keywords Moment generating function, Kernel type estimator, Stochastic approximation algorithm, Censored data

AMS 2010 subject classifications 62G08, 62L20, 60F10, 62G07

DOI: 10.19139/soic-2310-5070-1678

1. Introduction

Over years ago, [29] studied some properties of kernel density estimators introduced by [1] and [36]. Nonparametric regression function estimation has been the subject of intense investigation by statisticians and probabilists for many years, leading to the development of a large variety of methods. Kernel nonparametric function estimation methods have long attracted a great deal of attention. We advise the reader to see the following good references to the research literature in this area along with statistical applications consult [12], [37], [44], [14], [4, 3] and the references therein. The moment-generating function is an important tool for several statistical problems. Despite this importance, nonparametric estimation of the moment-generating function has received relatively scant attention. The moment-generating function is commonly thought of as a vehicle for obtaining the moments of a distribution. There are, however, other statistical settings in which it arises quite naturally. [30] used the moment-generating function to develop a method of estimating the parameters of a mixture of normal distributions. [16] used the empirical moment generating function to construct a test of separate families of distributions. Saddlepoint methods for approximating the pdf of a sample mean involve the moment-generating function of the underlying distribution (e.g., [31]). [10] proposed the moment-generating function to construct statistical tests for testing composite goodness-of-fit hypotheses on the exponential and bivariate Marshall-Olkin exponential distribution. [19] investigated the parametric estimation of the moment generating function. In the work of [24] tests of hypothesis are constructed for the family of skew-normal distributions. The proposed tests utilize the fact that the moment-generating function of the skew-normal variable satisfies a simple differential equation. [20] used a system of first-order partial differential equations that characterize the moment-generating function of the d-variate standard normal distribution to construct a class of affine invariant tests for normality

*Correspondence to: Yousri Slaoui (Email: Yousri.Slaoui@math.univ-poitiers.fr).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

in any dimension. This paper will consider the nonparametric recursive kernel-type estimators for the moment generating in the censored data setting by extending the previous work [6]. Recursive estimation, was proposed first in [34] and further investigation in many directions was given by [23], [13], [22], [26, 27], [40, 41] and [5].

This work concerns a nonparametric estimation of the recursive general kernel-type estimators for moment-generating functions for censored data defined by the stochastic approximation algorithm. To the best of our knowledge, the results presented here respond to a problem that has yet to be studied systematically up to the present, which was the basic motivation of the paper.

We start by giving some notation and definitions needed for the forthcoming sections. The problem of censoring is frequently encountered in certain statistical applications. The didactic example of censoring is arguably the study of the survival times of patients with a given chronic disease in a medical follow-up study lasting up to a fixed time t. If a patient is diagnosed with the disease at time s, then the survival time will be known if and only if the patient dies before time t. If this is not the case, then the only information available is that the survival time is equal to the censoring time $t-s$. In mathematical terms, the information available to the practitioner is the pair (T, C) defined in $\mathbb{R}\times\mathbb{R}$. Here T is the variable of interest, and C is a censoring variable. Throughout, we work with a sample $\{(T_i, C_i)_{1\leq i\leq n}\}$ of independent and identically distributed replicæ of (T, C), $n\geq 1$. Actually, in the right censorship model, the pairs (T_i, C_i), $1\leq i\leq n$, are not directly observed and the corresponding information is given by

$$Z_i := \min\{T_i, C_i\} \text{ and } \delta_i := \mathbb{1}\{T_i\leq C_i\},\ \ 1\leq i\leq n,$$

with $\mathbb{1}\{A\}$ standing for the indicator function of A. Accordingly, the observed sample is

$$\mathcal{D}_n = \{(Z_i, \delta_i), i = 1, \ldots, n\}.$$

For $-\infty < t < \infty$, set

$$F(t) = \mathbb{P}(T\leq t),\ G(t) = \mathbb{P}(C\leq t),\ \text{and } H(t) = \mathbb{P}(Z\leq t),$$

the right-continuous distribution functions of T, C and Z respectively. For example, survival data in clinical trials or failure time data in reliability studies are often subject to censoring. More specifically, 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, for recent references, see [7, 8, 42]. Let $T_1, T_2, \ldots T_n$ be a sequence of independent random variables with common distribution function $F(x)$, $x\in\mathbb{R}$ and probability density function $f(\cdot)$ with respect to the Lebesgue measure. Suppose that the moment generating function

$$C(t) = \int \exp(xt)\, f(x)dx,\ \ t\in\mathbb{R},$$

exists on a non-degenerate subset I of $\mathbb{R}$, necessarily containing the origin. Let us recall that, to construct a stochastic algorithm, which approximates the function f at a given point x, we need to define an algorithm of search of the zero of the function $h: y\rightarrow f(x)-y$. Following Robbins-Monro's procedure, this algorithm is defined by setting $f_0(x)\in\mathbb{R}$, and, for all $n\geq 1$,

$$f_n(x) = f_{n-1}(x) + \gamma_n W_n(x),$$

where $W_n(x)$ is an observation of the function h at the point $f_{n-1}(x)$, and the stepsize (γ_n) is a sequence of positive real numbers that go to zero. To define $W_n(x)$, we follow the approach of [32, 33] and of [39], and we introduce a kernel K (that is, a function satisfying $\int_{\mathbb{R}} K(x)dx = 1$), and a bandwidth (h_n) (that is, a sequence of positive real numbers that goes to zero), and sets $W_n(x) = h_n^{-1}\delta_n G(Z_n)^{-1}K\left(h_n^{-1}[x-Z_n]\right) - f_{n-1}(x)$. Then, the estimator f_n to recursively estimate the function f at the point x can be written as

$$f_n(x) = (1-\gamma_n)\, f_{n-1}(x) + \gamma_n h_n^{-1}\delta_n G(Z_n)^{-1}K\left(h_n^{-1}[x-Z_n]\right). \tag{1}$$

The function $G(\cdot)$ is generally unknown and has to be estimated. We will denote by $G_n(\cdot)$ the Kaplan-Meier estimator of the function $G(\cdot)$, see [21]. Namely, adopting the conventions $\prod_\emptyset = 1$ and $0^0 = 1$ and setting

$$N_n(u) = \sum_{i=1}^{n}\mathbb{1}\{Z_i\geq u\},$$

we have

$$G_n(u) = 1 - \prod_{i:Z_i \leq u} \left\{ \frac{N_n(Z_i) - 1}{N_n(Z_i)} \right\}^{(1-\delta_i)}, \quad \text{for } u \in \mathbb{R}.$$

The estimator $\widehat{f}_n$ to recursively estimate the function f at the point x

$$\widehat{f}_n(x) = (1-\gamma_n)\,\widehat{f}_{n-1}(x) + \gamma_n h_n^{-1} \delta_n G_n(Z_n)^{-1} K\left(h_n^{-1}[x - Z_n]\right), \tag{2}$$

where the stepsize (γ_n) is a sequence of positive real numbers that goes to zero, satisfying $\sum_{n\geq 1}\gamma_n = \infty$ and $\sum_{n\geq 1}\gamma_n^2 < \infty$ to ensure the almost sure convergence (see [13]), and the bandwidth (h_n) is a sequence of positive real numbers that go to zero. By using the equation (2), it follows that

$$\widehat{C}_n(t) = (1-\gamma_n)\,\widehat{C}_{n-1}(t) + \gamma_n h_n^{-1} \delta_n G_n(Z_n)^{-1} \int_{\mathbb{R}} \exp(xt)\, K\left(h_n^{-1}[x - Z_n]\right) dx. \tag{3}$$

Moreover, we set $\widehat{C}_0(t) = 0$ and

$$\Pi_n = \prod_{j=1}^{n} (1-\gamma_j),$$

then, we will investigate the following family of estimators

$$\widehat{C}_n(t) = \Pi_n \sum_{i=1}^{n} \gamma_i \Pi_i^{-1} h_i^{-1} \delta_i G_n(Z_i)^{-1} \int_{\mathbb{R}} \exp(xt)\, K\left(h_i^{-1}[x - Z_i]\right) dx. \tag{4}$$

Let

$$\widetilde{C}_n(t) = \int_{\mathbb{R}} \exp(xt)\, \widetilde{f}_n(x)\, dx, \tag{5}$$

where

$$\widetilde{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \frac{\delta_i}{1 - G_n(Z_i)} K\left(h_n^{-1}[x - Z_i]\right).$$

The recursive scheme offers many advantages to recursive estimators: they are easy implementation and do not require extensive data storage. More precisely, from a practical point of view, this arrangement provides important savings in computational time and storage memory, which is a consequence of the estimate updating independent of the data's history, providing a decisive computational advantage. The main drawback of the classical kernel estimator is using all data at each estimation step.

An outline of the remainder of the present paper is as follows. In Section 2, we will provide some notation and assumptions that we will use in our analysis. Section 3 is devoted to the main results of the present work. The finite sample performance of the proposed methodology is illustrated through Monte Carlo simulations in Section 4. Section 5 contains brief concluding remarks. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to Section 6.

2. Notation and assumptions

Throughout this paper, let us unburden our notation by writing

$$\mu_2(K) = \int_{\mathbb{R}} z^2 K(z) dz, \quad R(K) = \int_{\mathbb{R}} K^2(z)\, dz,$$

and

$$\xi = \lim_{n\to+\infty} (n\gamma_n)^{-1}. \tag{6}$$

First, let us set the following definition of the class of regularly varying sequences.

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

$$(v_n)_{n\geq 1} \in \mathcal{GS}(\gamma) \text{ if } \lim_{n\to+\infty} n\left[1 - \frac{v_{n-1}}{v_n}\right] = \gamma. \tag{7}$$

Condition (7) was introduced by [18] to define regularly Varying sequences (see also [2]). Noting that the acronym $\mathcal{GS}$ stands 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. For our main theoretical results, we need the following assumptions.

Assumptions:

($\mathbf{A1}$) $K : \mathbb{R} \to \mathbb{R}$ is a continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)\,dz = 1$, and, $\int_{\mathbb{R}} zK(z)\,dz = 0$ and $\int_{\mathbb{R}} z^2K(z)\,dz < \infty$.

($\mathbf{A2}$) (i) $(\gamma_n)_{n\geq 1} \in \mathcal{GS}(-\alpha)$, with $\alpha \in \left(\frac{1}{2}, 1\right]$,
(ii) $(h_n)_{n\geq 1} \in \mathcal{GS}(-a)$, with $a \in (0, \alpha]$,
(iii) $\lim_{n\to+\infty}(n\gamma_n) \in \left(\min\{a, \frac{\alpha-a}{2}\}, \infty\right]$;

($\mathbf{A3}$) the density function $f(\cdot)$ is bounded and differentiable.

Discussion on the assumptions:

- Assumptions ($\mathbf{A1}$) and ($\mathbf{A3}$) are standard in the framework of nonparametric kernel estimation (see, for instance, [39]).
- Assumption($\mathbf{A2}$) is widely used on the stochastic approximation algorithms (see, for instance, [26]).
- Assumption $(\mathbf{A2})(iii)$ on the limit of $(n\gamma_n)$ as n goes to infinity is usual in the framework of stochastic approximation algorithms. It implies in particular that the limit of $\left([n\gamma_n]^{-1}\right)$ is finite.
- To understand better the use the assumption ($\mathbf{A2}$), it is advised to consider the easiest sequence belonging to $\mathcal{GS}(\gamma)$, which is n^{γ}, one can check that for $(a_n) \in \mathcal{GS}(a)$ and $(b_n) \in \mathcal{GS}(b)$, we have $(a_nb_n) \in \mathcal{GS}(a+b)$ and $\left(a_nb_n^{-1}\right) \in \mathcal{GS}(a-b)$. For a sequences v_n belonging to $\mathcal{GS}(\gamma)$ with positive γ, we have $\lim_{n\to\infty} v_n = \infty$ and for sequences w_n belonging to $\mathcal{GS}(\beta)$ with negative β, we have $\lim_{n\to\infty} w_n = 0$. Then, it comes from $(\mathbf{A2})(i)$ that, $\gamma_n \to 0$, $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$, the assumption $(\mathbf{A2})(ii)$ ensures that $h_n \to 0$ and $\gamma_n/h_n \to 0$, the assumption $(\mathbf{A2})(iii)$, is very useful for the applicability of Lemma 1.
- The intuition behind the use of such bandwidth h_n belonging to $\mathcal{GS}(-a)$ is that the ratio h_{n-1}/h_n is equal to $1 + a/n + o(1/n)$, the application of Lemma 1 under the assumption ($\mathbf{A2}$), ensures that the bias and the variance depend only on h_n and not on $h_1, \ldots, h_n$.

3. Main results

Our first result is the following, which gives the bias and the variance of $\widehat{C}_n(\cdot)$ respectively.

Proposition 1 (Bias and Variance of $\widehat{C}_n(t)$)
Let Assumptions ($\mathbf{A1}$)-($\mathbf{A3}$) hold.

1. If $a \in \left(0, \frac{\alpha}{5}\right]$, then

$$\mathbb{E}[\widehat{C}_n(t)] - C(t) = \frac{h_n^2}{2(1-2a\xi)} t^2 \mu_2(K) \int_{\mathbb{R}} \exp(xt)\, f(x)\,dx + o\left(h_n^2\right). \tag{8}$$

2. If $a \in \left(\frac{\alpha}{5}, 1\right)$, then

$$\mathbb{E}[\widehat{C}_n(t)] - C(t) = o\left(\sqrt{\gamma_n h_n^{-1}}\right). \tag{9}$$

3. If $a \in \left(0, \frac{\alpha}{5}\right)$, then
$$\mathrm{Var}[\widehat{C}_n(t)] = o\left(h_n^4\right). \tag{10}$$
4. If $a \in \left[\frac{\alpha}{5}, 1\right)$, then
$$\mathrm{Var}[\widehat{C}_n(t)] = \frac{\gamma_n}{h_n} \frac{1}{(2-(\alpha-a)\xi)\,\overline{G}(t)} R(K) \int_{\mathbb{R}} \exp(2tx) f(x)\, dx + o\left(\frac{\gamma_n}{h_n}\right). \tag{11}$$

The bias and the Variance of the estimator $\widehat{C}_n(\cdot)$ defined by the stochastic approximation algorithm (5) then heavily depend on the choice of the stepsize (γ_n).

By following the proof of Proposition 1, we obtain this corollary.

Corollary 1 (Bias and Variance of $\widetilde{C}_n(t)$)
Let Assumptions $(\mathbf{A1})$, $(\mathbf{A1})\,(\mathbf{i})-(\mathbf{ii})$ and $(\mathbf{A3})$ hold.
$$\mathbb{E}[\widetilde{C}_n(t)] - C(t) = \frac{h_n^2}{2} t^2 \mu_2(K) \int_{\mathbb{R}} \exp(xt) f(x)\, dx + o\left(h_n^2\right), \tag{12}$$
and
$$\mathrm{Var}[\widetilde{C}_n(t)] = \frac{\gamma_n}{h_n \overline{G}(t)} R(K) \int_{\mathbb{R}} \exp(2xt) f(x)\, dx + o\left(\frac{\gamma_n}{h_n}\right). \tag{13}$$

Now, let us state the following theorem, which gives the asymptotic normality of the generalized recursive estimator $\widehat{C}_n(\cdot)$ defined in (3) and the generalized nonrecursive estimator $\widetilde{C}_n(\cdot)$ defined in (5) respectively.

3.1. Asymptotic normality

Let us now state the following theorem, which gives the weak convergence rate of the estimator $\widehat{C}_n(\cdot)$ defined in (5). Below, we write $Z \overset{\mathcal{D}}{=} \mathcal{N}(\mu, \sigma^2)$ whenever the random Variable Z follows a normal law with expectation μ and Variance σ^2, $\overset{\mathcal{D}}{\to}$ denotes the convergence in distribution and $\overset{\mathbb{P}}{\to}$ the convergence in probability.

Theorem 1 (Weak pointwise convergence rate of $\widehat{C}_n(\cdot)$)
Let the assumptions $(\mathbf{A1})$-$(\mathbf{A3})$ hold.

1. If there exists $c \geq 0$ such that $\gamma_n^{-1} h_n^5 \to c$, then
$$\sqrt{\gamma_n^{-1} h_n}(\widehat{C}_n(t) - C(t)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2(1-2a\xi)} t^2 \mu_2(K) \int_{\mathbb{R}} \exp(xt) f(x)\, dx, \frac{R(K)}{(2-(\alpha-a)\xi)\,\overline{G}(t)} \int_{\mathbb{R}} \exp(2xt) f(x)\, dx\right).$$
2. If $\gamma_n^{-1} h_n^5 \to \infty$, then
$$\frac{1}{h_n^2}(\widehat{C}_n(t) - C(t)) \overset{\mathbb{P}}{\to} \frac{t^2 \mu_2(K)}{(2(1-2a\xi))} \int_{\mathbb{R}} \exp(xt) f(x)\, dx.$$

The following corollary can be easily derived from the proof of Theorem 1.

Corollary 2 (Weak pointwise convergence rate of $\widetilde{C}_n(t)$)
Let the assumptions $(\mathbf{A1})$, $(\mathbf{A1})\,(\mathbf{i})-(\mathbf{ii})$ and $(\mathbf{A3})$ hold.

1. If there exists $c \geq 0$ such that $nh_n^5 \to c$, then
$$\sqrt{nh_n}(\widetilde{C}_n(t) - C(t)) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2} t^2 \mu_2(K) \int_{\mathbb{R}^d} \exp(xt) f(x)\, dx, \frac{R(K)}{\overline{G}(t)} \int_{\mathbb{R}} \exp(2xt) f(x)\, dx\right).$$

2. If $nh_n^5 \to \infty$, then

$$\frac{1}{h_n^2}(\widetilde{C}_n(t) - C(t)) \xrightarrow{\mathbb{P}} \frac{1}{2}t^2\mu_2(K)\int_{\mathbb{R}} \exp(xt) f(x)\, dx.$$

Remark 1 1. The rate of convergence of the recursive estimator $\widehat{C}_n(t)$ is $\sqrt{\gamma_n^{-1} h_n}$, while the rate of convergence of the recursive estimator $\widetilde{C}_n(t)$ is $\sqrt{nh_n}$.

2. In the case when $(\gamma_n) = \left(n^{-1}\right)$, the bias, variance and the rate of convergence of the two estimators $\widehat{C}_n(t)$ and $\widetilde{C}_n(t)$ are the same.
3. The advantage of recursive estimators on their nonrecursive version is that their update, from a sample of size n to one of size $n+1$, require fewer computations. This property can be generalized if we suppose that we receive two sets of data separately, the first one of cardinal n_1 smaller or equal to $n-1$ and the second set of cardinal $n - n_1$. We infer from (5) that,

$$\begin{aligned}
\widehat{C}_n(t) &= \prod_{j=n_1+1}^{n}(1-\gamma_j)\,\widehat{C}_{n_1}(t) \\
&\quad + \sum_{k=n_1}^{n-1}\left[\prod_{j=k+1}^{n}(1-\gamma_j)\right]\gamma_k h_k^{-1}\delta_k G_n(Z_k)^{-1}\int_{\mathbb{R}}\exp(xt)\, K\left(\frac{x-Z_k}{h_k}\right)dx. \\
&= \alpha_1\widehat{C}_{n_1}(t) + \sum_{k=n_1}^{n-1}\beta_k\gamma_k h_k^{-1}\delta_k G_n(Z_k)^{-1}\int_{\mathbb{R}}\exp(xt)\, K\left(\frac{x-Z_k}{h_k}\right)dx,
\end{aligned}$$

where $\alpha_1 = \prod_{j=n_1+1}^{n}(1-\gamma_j)$ and $\beta_k = \prod_{j=k+1}^{n}(1-\gamma_j)$. Then the proposed estimator can be viewed as a linear combination of two estimators, which improves the computational cost significantly.

4. Simulation results

In this section, series of experiments are conducted to examine the performance of the proposed estimators given in (5). The computing program codes are implemented in R. More precisely, we consider the case of drawing i.i.d. univariate random samples X_i, $i = 1, \ldots, n$. We consider the exponential $\mathcal{E}(1)$ for which $C(t) = 1/(1-t)$. In our simulation study, we make use of the following kernels:

- the gaussian kernel:

$$K(x/h) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2h^2},$$

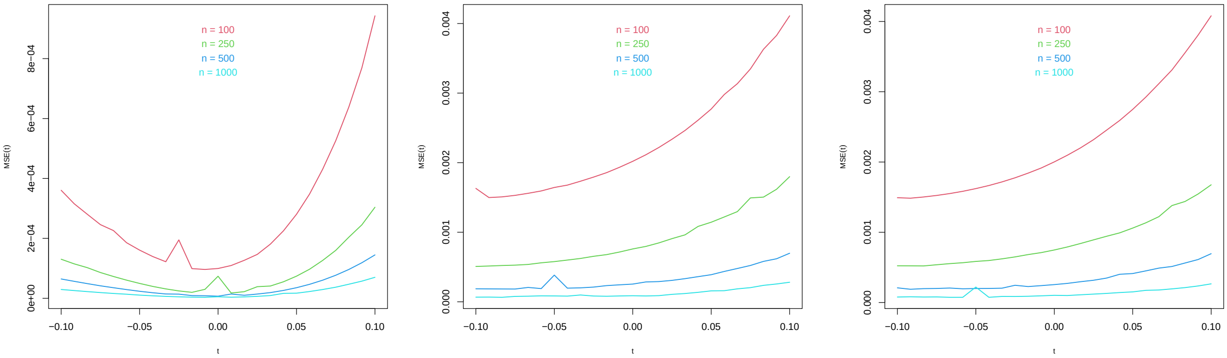
- the [15] kernel:

$$K(x/h) = \frac{3}{4}(1-(x/h)^2)\mathbb{1}\{|x/h| \leq 1\},$$

- the quadratic kernel :

$$K(u) = \frac{15}{16}\left(1-u^2\right)^2 \mathbb{1}\{|u| \leq 1\}.$$

Here, h is the smoothing bandwidth. We adopt the "normal scale rule" or the rule-of-thumb method, see for instance [38], to select the bandwidth, i.e., we chose h to be $\alpha_h\hat{\sigma}(X)n^{-1/5}$ where α_h is some positive constant and $\hat{\sigma}(X)$ is the standard deviation of X. These frameworks allow us to examine the finite sample properties of our estimators in (5). To this end, we compute our estimators for each of the three kernels presented above and some values of α_h and $n \in \{100, 250, 500, 1000\}$. The parameter α_h is calculated by minimizing the L_2 distance between $f_n(\cdot)$ and $f(\cdot)$, i.e.,



(a) (b) (c)

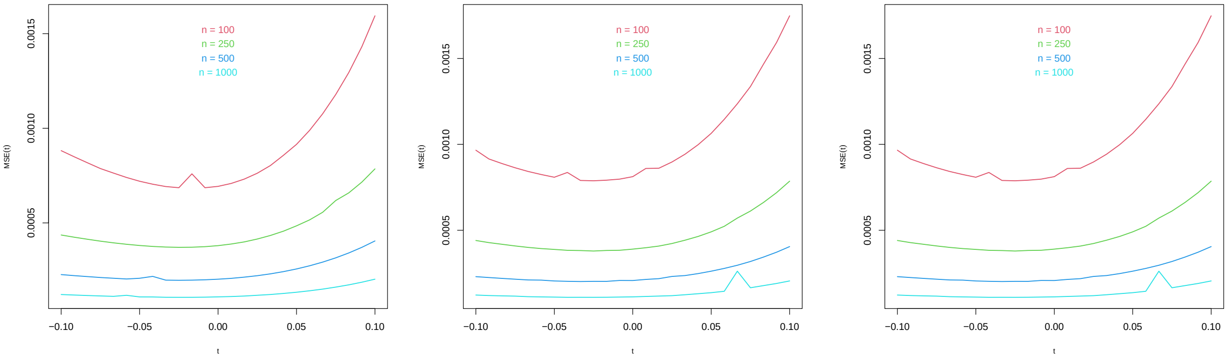
Figure 1. The local MSE of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.18$ in (a), the Epanečnikov kernel with $\alpha_h = 0.37$ in (b) and the quadratic kernel with $\alpha_h = 0.44$ in (c). Complete data.

$$\arg\min_{\alpha_h \in A} \sum_{i=1}^{\ell} (f_n(t_i) - f(t_i))^2,$$

where A is an appropriately chosen set. In our simulation $A = [0, 001, 10]$. We have chosen the uniform discretization $t_1, \ldots, t_\ell$ with $\ell = 50$ of $[-0, 10, 0, 10]$. The choice of α_h is not optimal since we are choosing this to minimize the distance between the densities rather than between the moment-generating functions. This choice is sufficient for our needs. The flexibility of this choice is due to the rule-of-thumb method. For the sake of effective calculations of these measures, the theoretical density can be replaced by the empirical counterparts based, for example, on 10000 simulations. For each setting, we consider three local measures are given for a given t and for any estimate (say $\widetilde{C}_n(t)$), let

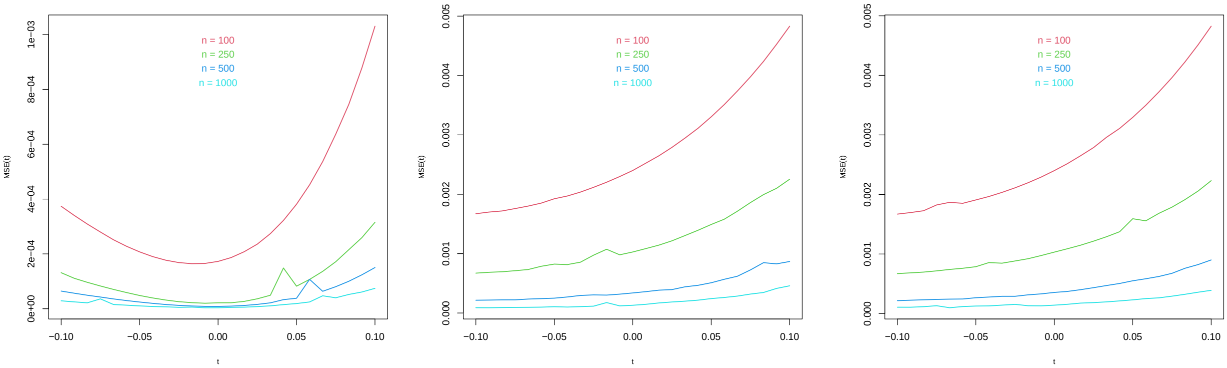
- the (local) bias: $\mathrm{Bias}(t) := \mathbb{E}\left[\widetilde{C}_n(t)\right] - C(t)$,
- the (local) variance: $\mathrm{Var}(t) := \mathbb{E}\left[\left(\widetilde{C}_n(t) - \mathbb{E}\left[\widetilde{C}_n(t)\right]\right)^2\right]$,
- the (local) mean square-error: $\mathrm{MSE}(t) := \mathbb{E}\left[\left(\widetilde{C}_n(t) - C(t)\right)^2\right]$.

The same remark that $C(t)$ can be replaced by the empirical counterparts based, for example, on 10000 simulations. We will consider different intensities of censoring in the sample. The desired censoring rates (proportions) (cr) are 5%, 10% or 30%.
Notice that, as in any other inferential context, the greater the sample size is, the better the performance is. Simple inspection of the results reported in the Figures 1, 3, 5 and 7 show local MSE for $\widetilde{C}_n(t)$), while Figures 2, 4, 6 and 8 show local MSE for $\widehat{C}_n(t)$ allows us to deduce that large values of the sample size n gives smaller MSE. Figures 9, 11, 13 and 15, show local variance results for $\widetilde{C}_n(t)$), Figures 10, 12, 14 and 16, show local variance results for $\widehat{C}_n(t)$ display the results for the bias and the variance for the nonrecursive and the recursive estimators. As in the results for the MSE, we have good performance of the estimators for the normal distribution. From figures, the best results are obtained when the data is complete, and the results in the censoring case are satisfactory when the censoring rate is moderate 5%, 10% and 30% and the performance deteriorates when the censoring rate increase.



(a) (b) (c)

Figure 2. The local MSE of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.14$ in (a), the Epanečnikov kernel with $\alpha_h = 0.29$ in (b) and the quadratic kernel with $\alpha_h = 0.35$ in (c). Complete data.

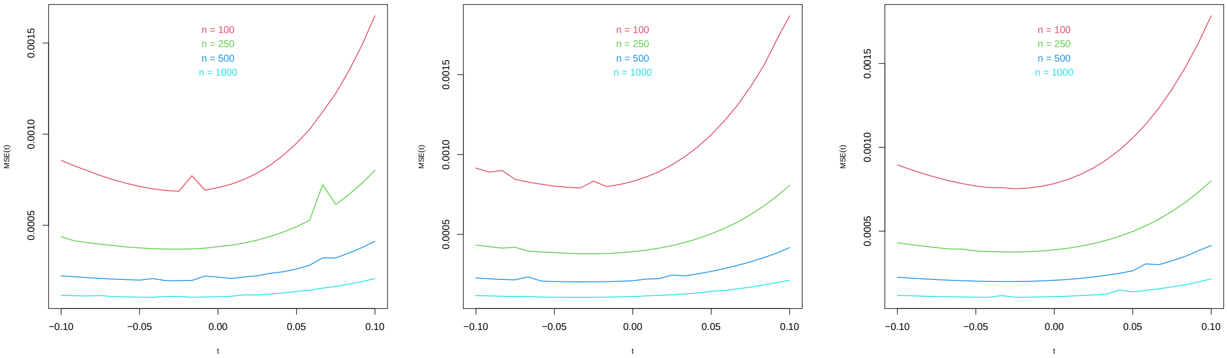


(a) (b) (c)

Figure 3. The local MSE of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.19$ in (a), the Epanečnikov kernel with $\alpha_h = 0.40$ in (b) and the quadratic kernel with $\alpha_h = 0.48$ in (c). Censored data, $cr = 0.05$.

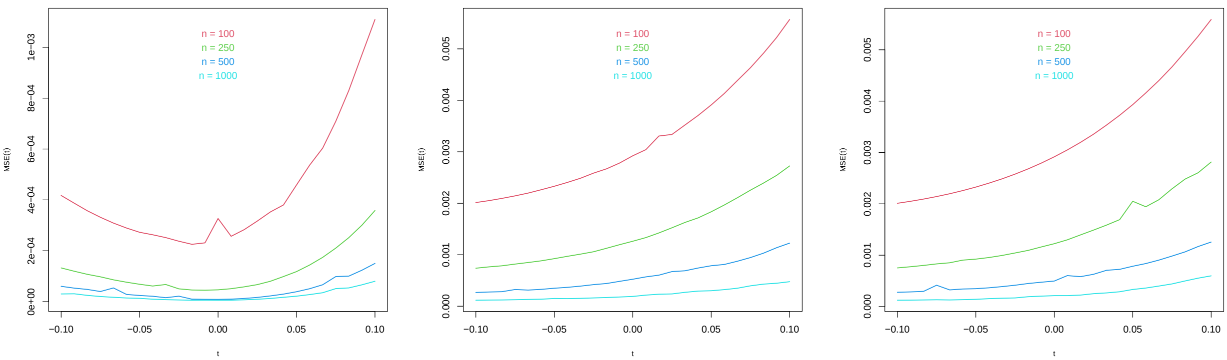
5. Conclusion

In this paper, we have considered estimating the nonparametric moment-generating function in the censored data setting. We have investigated the asymptotic properties of the nonparametric recursive kernel-type estimators for the moment-generating function. More precisely, we obtained the central limit theorem together with the characterization of the bias and the Variance of these estimators under general conditions. A future research direction would be to study the problem of estimation in nonparametric moment-generating function models as such investigated in this work in the setting of serially dependent observations (mixing or weak dependent), which requires nontrivial mathematics that goes well beyond the scope of the present paper. We plan to extend the current work to some applied fields, such as psychological data.



(a) (b) (c)

Figure 4. The local MSE of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.14$ in (a), the Epanečnikov kernel with $\alpha_h = 0.30$ in (b) and the quadratic kernel with $\alpha_h = 0.36$ in (c). Censored data, $cr = 0.05$



(a) (b) (c)

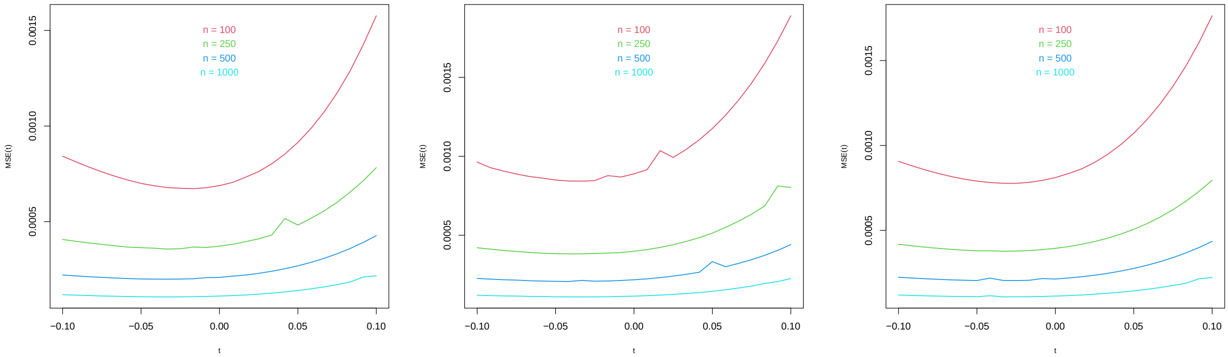
Figure 5. The local MSE of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.21$ in (a), the Epanečnikov kernel with $\alpha_h = 0.43$ in (b) and the quadratic kernel with $\alpha_h = 0.52$ in (c). Censored data, $cr = 0.10$

6. Proofs

This section is devoted to the proof of our results. The previously presented notation continues to be used in the following.
For any distribution function (df) $L(\cdot)$ recall that

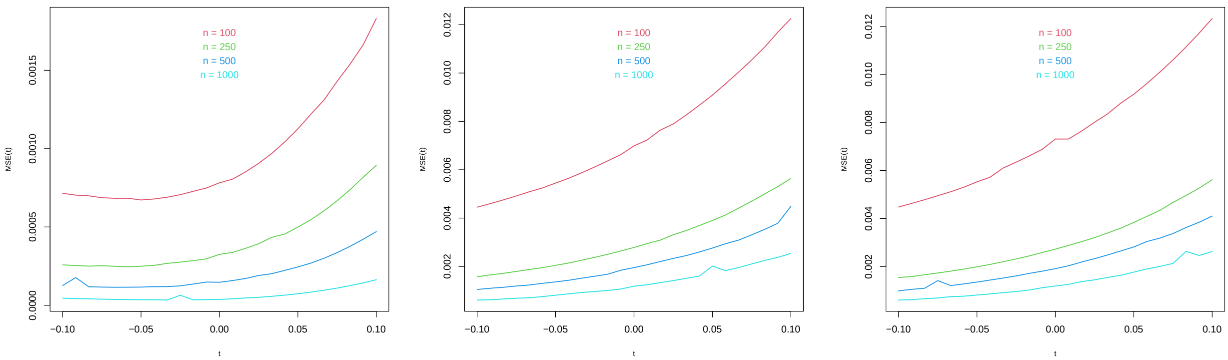
$$\tau_L = \sup\{t : L(t) < 1\}$$

be its support's right endpoint. Further, we will denote by τ_F (resp. τ_G) the upper endpoints of $F(\cdot)$ (resp. of $G(\cdot)$). In the following we assume that $\tau_F < \infty$, $G(\tau_F) > 0$, $\tau_H < \min(\tau_F, \tau_G)$ and C is independent to $(\mathbf{X}, T)$.
Now, we define the sequence (m_n) by setting



(a) (b) (c)

Figure 6. The local MSE of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.15$ in (a), the Epanečnikov kernel with $\alpha_h = 0.30$ in (b) and the quadratic kernel with $\alpha_h = 0.37$ in (c). Censored data, $cr = 0.10$



(a) (b) (c)

Figure 7. The local MSE of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.31$ in (a), the Epanečnikov kernel with $\alpha_h = 0.64$ in (b) and the quadratic kernel with $\alpha_h = 0.77$ in (c). Censored data, $cr = 0.30$

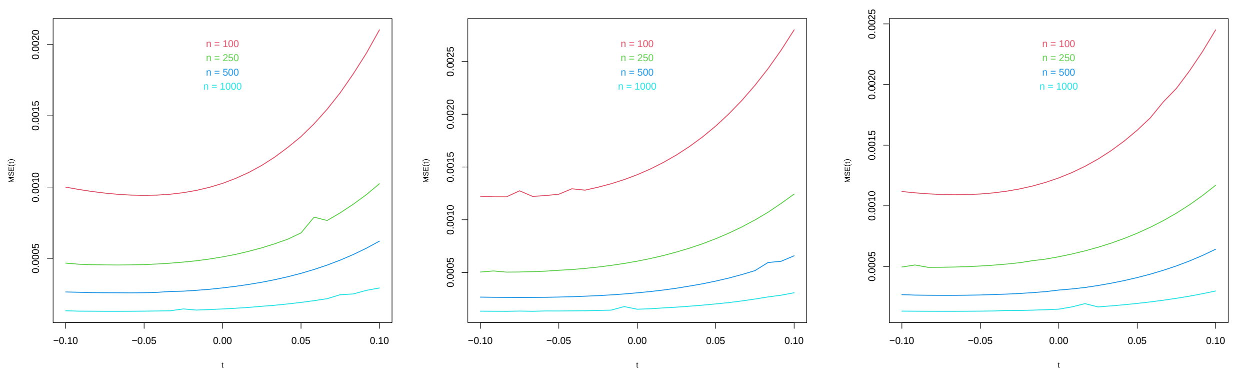
$$(m_n) \quad = \quad \begin{cases} \dfrac{\log\log n}{\sqrt{\gamma_n^{-1} h_n^2}} & \text{if} \quad \dfrac{\log\log n}{\sqrt{\gamma_n^{-1} h_n^6}} = \infty, \\ h_n^2 & \quad \text{otherwise}. \end{cases} \tag{14}$$

Further, we consider the following notation throughout this section

$$\mathcal{T}_n(t) = h_n^{-1} \delta_n G(Z_n)^{-1} \int_{\mathbb{R}} \exp(xt)\, K\left(\frac{x - X_n}{h_n}\right) dx,$$

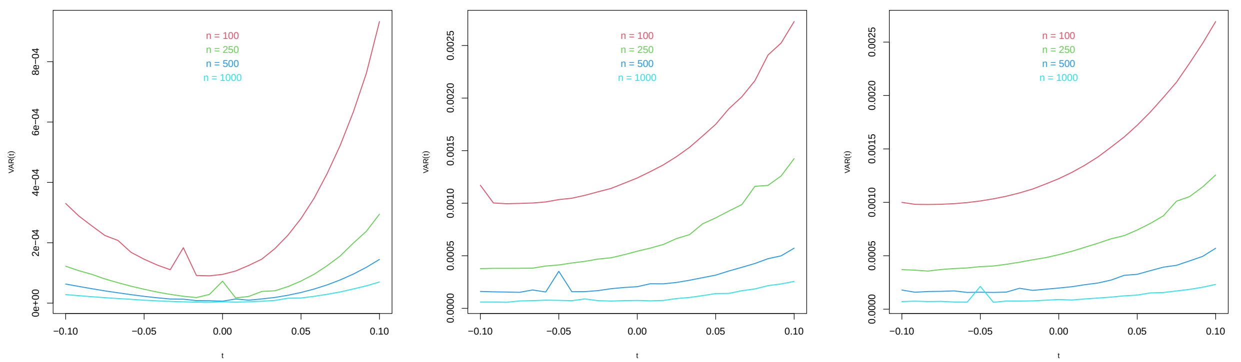
and we use the fact that,

$$\mathbb{1}_{\{T_1 \le C_1\}} \varphi(Z_1) = \mathbb{1}_{\{T_1 \le C_1\}} \varphi(T_1)$$



(a) (b) (c)

Figure 8. The local MSE of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.17$ in (a), the Epanečnikov kernel with $\alpha_h = 0.35$ in (b) and the quadratic kernel with $\alpha_h = 0.43$ in (c). Censored data, $cr = 0.30$



(a) (b) (c)

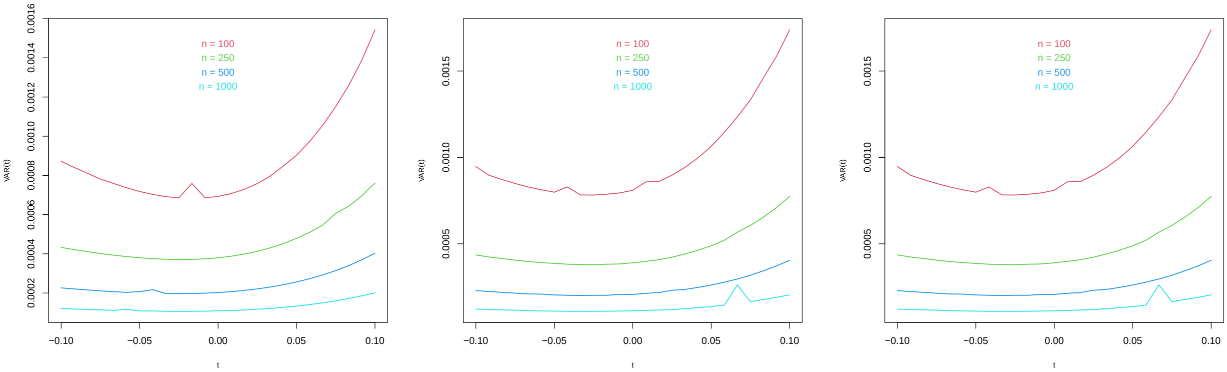
Figure 9. The local VAR of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.18$ in (a), the Epanečnikov kernel with $\alpha_h = 0.37$ in (b) and the quadratic kernel with $\alpha_h = 0.44$ in (c). Complete data.

for all measurable function $\varphi(\cdot)$. Then, we readily obtain that

$$\mathcal{T}_n(t) = h_n^{-1}\mathbb{1}_{\{T_n<C_n\}}G(T_n)^{-1}\int_{\mathbb{R}}\exp(xt)\,K\left(\frac{x-X_n}{h_n}\right)dx. \tag{15}$$

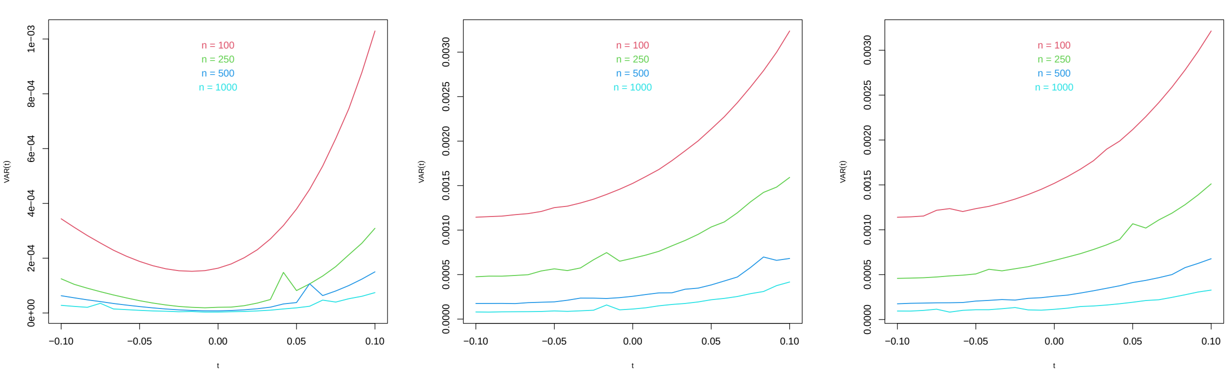
Let,

$$C_n(t) = \Pi_n\sum_{i=1}^{n}\gamma_i\Pi_i^{-1}h_i^{-1}\delta_i G(Z_i)^{-1}\int_{\mathbb{R}}\exp(xt)\,K\left(h_i^{-1}[x-X_i]\right)dx. \tag{16}$$



(a) (b) (c)

Figure 10. The local VAR of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.14$ in (a), the Epanečnikov kernel with $\alpha_h = 0.29$ in (b) and the quadratic kernel with $\alpha_h = 0.35$ in (c). Complete data.

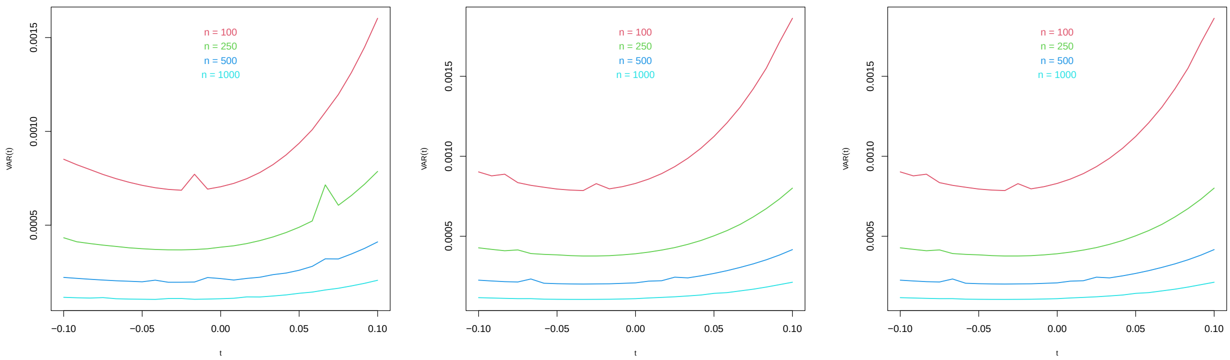


(a) (b) (c)

Figure 11. The local VAR of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.19$ in (a), the Epanečnikov kernel with $\alpha_h = 0.40$ in (b) and the quadratic kernel with $\alpha_h = 0.48$ in (c). Censored data, $cr = 0.05$

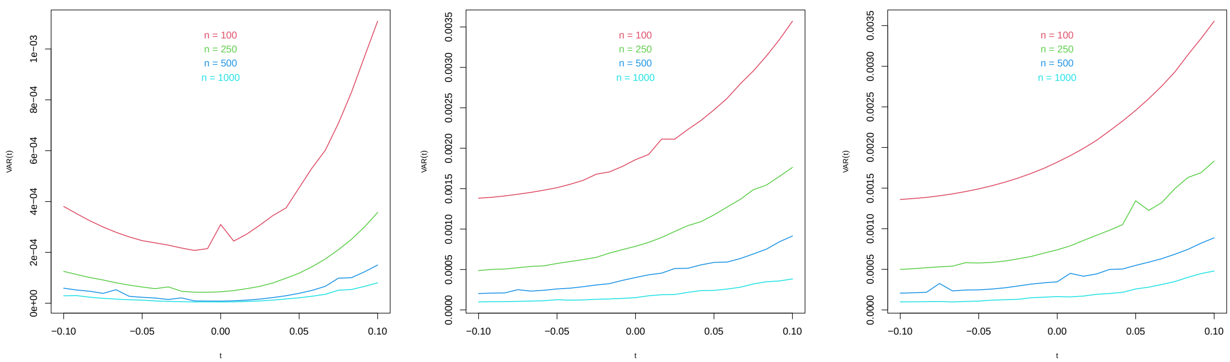
The combination between (4) and (16) ensure that

$$\begin{aligned}
&\left|\widehat{C}_n(t) - C_n(t)\right| \\
&= \Pi_n \left|\sum_{i=1}^{n} \Pi_i^{-1}\gamma_i h_i^{-1}\delta_i \left[\frac{1}{G_n(Z_i)} - \frac{1}{G(Z_i)}\right] \int_{\mathbb{R}} \exp(xt) K\left(h_i^{-1}[x - X_i]\right) dx\right| \\
&= \Pi_n \left|\sum_{i=1}^{n} \Pi_i^{-1}\gamma_i h_i^{-1}\mathbb{1}_{\{T_i < C_i\}} \left[\frac{1}{G_n(T_i)} - \frac{1}{G(T_i)}\right] \int_{\mathbb{R}} \exp(xt) K\left(h_i^{-1}[x - X_i]\right) dx\right| \\
&\leq \Pi_n \left|\sum_{i=1}^{n} \Pi_i^{-1}\gamma_i h_i^{-1} \left[\frac{G_n(T_i) - G(T_i)}{G_n(T_i)G(T_i)}\right] \int_{\mathbb{R}} \exp(xt) K\left(h_i^{-1}[x - X_i]\right) dx\right| \\
&\leq \frac{\sup_{t \leq \tau_H}(|G_n(t) - G(t)|)}{G_n(\tau_H)G(\tau_H)} \Pi_n \left|\sum_{i=1}^{n} \Pi_i^{-1}\gamma_i h_i^{-1} \int_{\mathbb{R}} \exp(xt) K\left(h_i^{-1}[x - X_i]\right) dx\right|.
\end{aligned}$$



(a) (b) (c)

Figure 12. The local VAR of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.14$ in (a), the Epanečnikov kernel with $\alpha_h = 0.30$ in (b) and the quadratic kernel with $\alpha_h = 0.36$ in (c). Censored data, $cr = 0.05$.



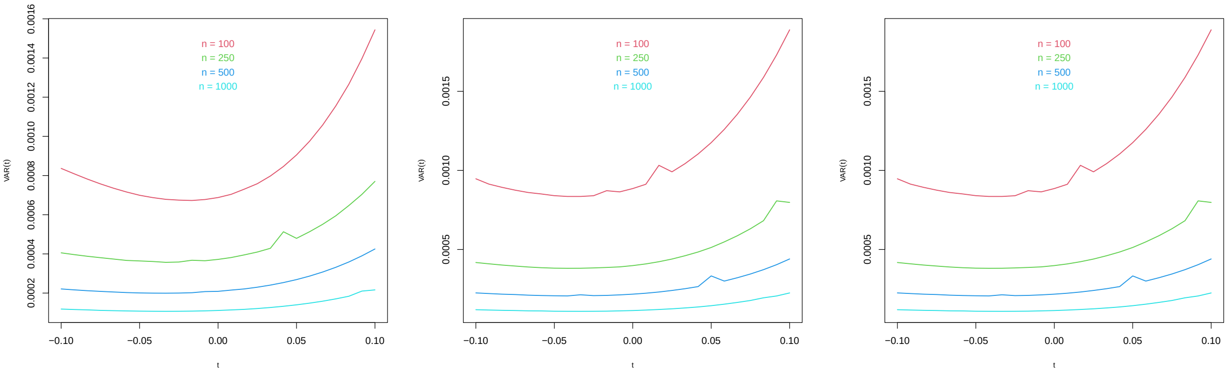
(a) (b) (c)

Figure 13. The local VAR of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.21$ in (a), the Epanečnikov kernel with $\alpha_h = 0.43$ in (b) and the quadratic kernel with $\alpha_h = 0.52$ in (c). Censored data, $cr = 0.10$

Then by using the strong law of large numbers (SLLN) and the law of iterated logarithm (LIL) on the censoring law (see formula (4.28) in [11], see also [17]), we have

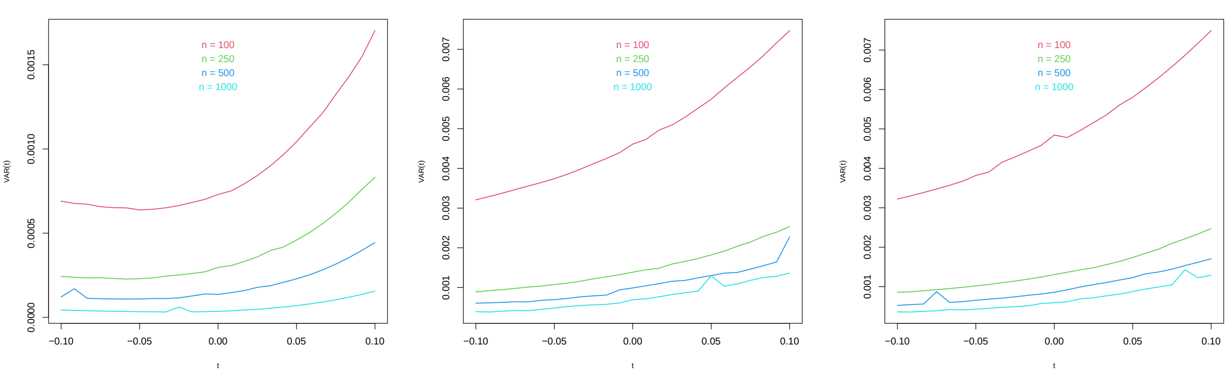
$$\sup_{x\in S}\left|\widehat{C}_n(t) - C_n(t)\right| = O\left(\sqrt{\frac{\log\log n}{nh_n^2}}\right) = o\left(m_n\right). \tag{17}$$

The following simple lemma will play an instrumental role in the sequel. This section is devoted to the proof of our results. The previously presented notation continues to be used in the following. Before giving the outlines of the proofs, we state the following technical lemma, proved in [26], and widely applied throughout the demonstrations.



(a) (b) (c)

Figure 14. The local VAR of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.15$ in (a), the Epanečnikov kernel with $\alpha_h = 0.30$ in (b) and the quadratic kernel with $\alpha_h = 0.37$ in (c). Censored data, $cr = 0.10$.



(a) (b) (c)

Figure 15. The local VAR of $\widetilde{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.31$ in (a), the Epanečnikov kernel with $\alpha_h = 0.64$ in (b) and the quadratic kernel with $\alpha_h = 0.77$ in (c). Censored data, $cr = 0.30$

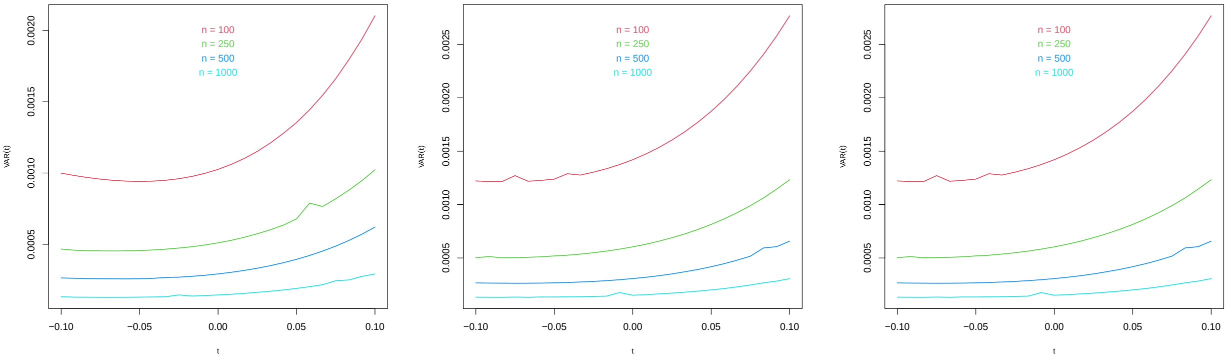
Lemma 1
Let $(v_n) \in \mathcal{GS}(v^*)$, $(\gamma_n) \in \mathcal{GS}(-\alpha)$, and $m > 0$ such that $m - v^*\xi > 0$ where ξ is defined in (6). We have

$$\lim_{n\to+\infty} v_n \Pi_n^m \sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}. \tag{18}$$

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

$$\lim_{n\to+\infty} v_n \Pi_n^m \left[\sum_{k=1}^{n} \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0. \tag{19}$$

Let us underline that the application of Lemma 1 requires Assumption **(A1)**(iii) on the limit of $(n\gamma_n)$ as n goes to infinity.



(a) (b) (c)

Figure 16. The local VAR of $\widehat{C}_n(t)$ estimator for the exponential distribution. The used kernel and α_h are: the gaussian kernel with $\alpha_h = 0.17$ in (a), the Epanečnikov kernel with $\alpha_h = 0.35$ in (b) and the quadratic kernel with $\alpha_h = 0.43$ in (c). Censored data, $cr = 0.30$.

We denote by $\mathfrak{C}$ a constant Varying from line to line. Our proofs are organized as follows. Propositions 1 in Section 6.1, Theorem 1 in Section 6.2.

6.1. *Proof of Proposition 1*

We first note that we have

$$\widehat{C}_n(t) - C\left(t\right) \quad = \quad \widehat{C}_n(t) - C_n(t) + C_n(t) - C\left(t\right). \tag{20}$$

Then, it follows from (17), that the asymptotic behavior of $\widehat{C}_n(t) - C\left(t\right)$ can be deduced from the one of $C_n\left(t\right) - C\left(t\right)$. Moreover, in view of (16) and (15), we can write that

$$\mathbb{E}\left[C_n(t)\right] - C\left(t\right) = \Pi_n \sum_{i=1}^{n} \Pi_i^{-1}\gamma_i \left\{\mathbb{E}\left[\mathcal{T}_i\left(t\right)\right] - C\left(t\right)\right\}.$$

Since we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{T}_i\left(t\right)\right] &= \mathbb{E}\left[h_i^{-1}\mathbb{E}\left[\mathbb{1}_{\{T_i<C_i\}}|T_i, X_i\right] G(T_i)^{-1}\int_{\mathbb{R}} \exp\left(xt\right) K\left(h_i^{-1}\left[x - X_i\right]\right)\right] dx \\
&= \mathbb{E}\left[h_i^{-1}\int_{\mathbb{R}} \exp\left(xt\right) K\left(h_i^{-1}\left[x - X_i\right]\right)\right] dx.
\end{aligned}$$

Moreover, in view of (16), we infer that

$$\begin{aligned}
C_n\left(t\right) - C\left(t\right) &= \left(1-\gamma_n\right)\left(C_{n-1}\left(t\right) - C\left(t\right)\right) + \gamma_n\left(\mathcal{T}_n\left(t\right) - C\left(t\right)\right) \\
&= \sum_{k=1}^{n-1}\left[\prod_{j=k+1}^{n}\left(1-\gamma_j\right)\right]\gamma_k\left(\mathcal{T}_k\left(t\right) - \widehat{C}_n\right) + \gamma_n\left(\mathcal{T}_n\left(t\right) - C\left(t\right)\right) \\
&\quad + \left[\prod_{j=1}^{n}\left(1-\gamma_j\right)\right]\left(C_0\left(t\right) - C\left(t\right)\right) \\
&= \Pi_n \sum_{k=1}^{n} \Pi_k^{-1}\gamma_k\left(\mathcal{T}_k\left(t\right) - C\left(t\right)\right) + \Pi_n\left(C_0\left(t\right) - C\left(t\right)\right).
\end{aligned} \tag{21}$$

This readily implies that

$$\mathbb{E}\left(C_n\left(t\right)\right)-C\left(t\right) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_k\left(\mathbb{E}\left(\mathcal{T}_k\left(t\right)\right)-C\left(t\right)\right)+\Pi_n\left(C_0\left(t\right)-C\left(t\right)\right). \tag{22}$$

Taylor's expansion with integral remainder ensures that

$$\begin{aligned}\mathbb{E}\left[\mathcal{T}_k\left(t\right)\right]-C\left(t\right) &= \int_{\mathbb{R}^2}\left\{\exp\left(t\left(x+zh_k\right)\right)-\exp\left(tx\right)\right\}K\left(z\right)f(x)dzdx.\\ &= \frac{h_k^2}{2}t^2\mu_2\left(K\right)\int_{\mathbb{R}}\exp\left(xt\right)f\left(x\right)dx+h_k^2\delta_k\left(t\right),\end{aligned} \tag{23}$$

where

$$\delta_k\left(t\right)=h_k^{-2}\int_{\mathbb{R}^2}f\left(x\right)K\left(z\right)\left[\left\{\exp\left(t\left(x+zh_k\right)\right)-\exp\left(tx\right)\right\}-t^2z^2\frac{h_k^2}{2}\exp\left(xt\right)\right]dxdz.$$

We have $\lim_{k\to\infty}\delta_k\left(t\right)=0$. In the case $a\leq\alpha/5$, we have $\lim_{n\to\infty}\left(n\gamma_n\right)>2a$; the application of Lemma 1 then gives

$$\begin{aligned}\mathbb{E}\left[C_n\left(t\right)\right]-C\left(t\right) &= \frac{1}{2}t^2\mu_2\left(K\right)\int_{\mathbb{R}}\exp\left(xt\right)f\left(x\right)dx\left\{\Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_kh_k^2[1+o(1)]\right\}\\ &\quad+\Pi_n\left(C_0\left(t\right)-C\left(t\right)\right)\\ &= \frac{1}{2(1-2a\xi)}t^2\mu_2\left(K\right)\int_{\mathbb{R}}\exp\left(xt\right)f\left(x\right)dx\left[h_n^2+o(1)\right],\end{aligned}$$

and (8) follows form the combination of (17) and (20). In the case $a>\alpha/5$, we have $h_n^2=o\left(\sqrt{\gamma_nh_n^{-1}}\right)$. Since we have $\lim_{n\to\infty}\left(n\gamma_n\right)>\left(\alpha-a\right)/2$, the application of Lemma 1 gives

$$\mathbb{E}\left[C_n\left(t\right)\right]-C\left(t\right) = \Pi_n\sum_{k=1}^{n}\Pi_k^{-1}\gamma_ko\left(\sqrt{\gamma_kh_k^{-1}}\right)+O\left(\Pi_n\right)=o\left(\sqrt{\gamma_nh_n^{-1}}\right),$$

the combination of (17) and (20) gives (9). Now, since $X_1, X_2, \ldots X_n$ is a sequence of independent uni-dimensional random vectors with common distribution function $F(x)$, we have $\mathcal{C}ov\left(Z_k\left(t\right), Z'_k\left(t\right)\right)=0$ for $k\neq k'$, then, it comes that

$$\begin{aligned}\mathrm{Var}\left[C_n\left(t\right)\right] &= \frac{\Pi_n^2}{\overline{G}(t)}\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2\mathrm{Var}\left[\mathcal{T}_k\left(t\right)\right]\\ &= \frac{\Pi_n^2}{\overline{G}(t)}\sum_{k=1}^{n}\frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[\int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\exp\left(t\left(x+zh_k\right)\right)K\left(z\right)dz\right\}\right.\\ &\quad\times\left\{\int_{\mathbb{R}}\exp\left(t\left(x+z'h_k\right)\right)K\left(z'\right)dz'\right\}f\left(x\right)dx\\ &\quad\left.-h_k\left(\int_{\mathbb{R}^2}K\left(z\right)\exp\left(t\left(x-zh_k\right)\right)f\left(x-zh_k\right)dxdz\right)^2\right]\\ &= \frac{\Pi_n^2}{\overline{G}(t)}\sum_{k=1}^{n}\frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[R\left(K\right)\int_{\mathbb{R}}\exp\left(2xt\right)f\left(x\right)dx+\nu_k\left(t\right)-h_k\widetilde{\nu}_k\left(t\right)\right],\end{aligned}$$

where

$$\begin{aligned}
\nu_k(t) &= \int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\{\exp(t(x+zh_k))-\exp(xt)\}K(z)\,dz\right\} \\
&\quad\times\left\{\int_{\mathbb{R}}\{\exp(t(x+z'h_k))-\exp(xt)\}K(z')\,dz'\right\}f(x)\,dx \\
\widetilde{\nu}_k(t) &= \left(\int_{\mathbb{R}^2}\exp(t(x+zh_k))K(z)f(x)\,dxdz\right)^2.
\end{aligned}$$

In view of **(A3)**, we have $\lim_{k\to\infty}\nu_k(t)=0$ and $\lim_{k\to\infty}h_k\widetilde{\nu}_k(t)=0$, we let $\varepsilon_k(t)=\nu_k(t)-h_k\widetilde{\nu}_k(t)$, we have $\lim_{k\to\infty}\varepsilon_k(t)=0$. In the case $a\geq\alpha/5$, we have $\lim_{n\to\infty}(n\gamma_n)>(\alpha-a)/2$, we make use of Lemma 1 to infer that

$$\begin{aligned}
\operatorname{Var}[C_n(t)] &= \frac{\Pi_n^2}{\overline{G}(t)}\sum_{k=1}^{n}\frac{\Pi_k^{-2}\gamma_k^2}{h_k}\left[R(K)\int_{\mathbb{R}}\exp(2xt)f(x)\,dx+\varepsilon_k(t)\right] \\
&= \frac{1}{(2-(\alpha-a)\xi)\,\overline{G}(t)}\frac{\gamma_n}{h_n^d}\left[R(K)\int_{\mathbb{R}}\exp(2xt)f(x)\,dx+o(1)\right],
\end{aligned}$$

the combination of (17) and (20) gives (10). When $a<\alpha/5$, we have $\gamma_n h_n^{-1}=o\left(h_n^4\right)$. Then, since $\lim_{n\to\infty}(n\gamma_n)>2a$, we apply Lemma 1 to infer that

$$\operatorname{Var}[C_n(t)] = \Pi_n^2\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k o\left(h_k^4\right)=o\left(h_n^4\right),$$

the combination of (17) and (20) proves (11). □

6.2. *Proof of Theorem 1*

First, it comes from (21) and (22), that

$$C_n(t)-\mathbb{E}[C_n(t)]=\Pi_n\sum_{k=1}^{n}Y_k(t),$$

where

$$Y_k(t)=\Pi_k^{-1}\gamma_k\left(\mathcal{T}_k(t)-\mathbb{E}(\mathcal{T}_k(t))\right).$$

Since $\lim_{n\to\infty}(n\gamma_n)>(\alpha-a)/2$, the application of Lemma 1 ensures that

$$\begin{aligned}
v_n^2 &= \sum_{k=1}^{n}\operatorname{Var}(Y_k(t))=\sum_{k=1}^{n}\Pi_k^{-2}\gamma_k^2\operatorname{Var}(\mathcal{T}_k(t)) \\
&= \frac{R(K)}{\overline{G}(t)}\sum_{k=1}^{n}\frac{\Pi_k^{-2}\gamma_k^2}{h_n}\left[\int_{\mathbb{R}}\exp(2xt)f(x)\,dx+o(1)\right] \\
&= \frac{R(K)}{\Pi_n^2\overline{G}(t)}\frac{\gamma_n}{h_n}\left[\frac{1}{2-(\alpha-a)\xi}\int_{\mathbb{R}}\exp(2xt)f(x)\,dx+o(1)\right].
\end{aligned}$$

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

$$\mathbb{E}\left[|\mathcal{T}_k(t)|^{2+p}\right] = O\left(\frac{1}{h_k^{(1+p)}}\right),$$

and, since $\lim_{n\to\infty}(n\gamma_n) > (\alpha - a)/2$, there exists $p > 0$ such that $\lim_{n\to\infty}(n\gamma_n) > \frac{1+p}{2+p}(\alpha - a)$. Applying Lemma 1, we get

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\left[|Y_k(x)|^{2+p}\right] &= O\left(\sum_{k=1}^{n}\Pi_k^{-2-p}\gamma_k^{2+p}\mathbb{E}\left[|\mathcal{T}_k(t)|^{2+p}\right]\right)\\ &= O\left(\sum_{k=1}^{n}\frac{\Pi_k^{-2-p}\gamma_k^{2+p}}{h_k^{(1+p)}}\right) = O\left(\frac{\gamma_n^{1+p}}{\Pi_n^{2+p}h_n^{d(1+p)}}\right),\end{aligned}$$

and we thus obtain

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

Then the application of Lyapunov's Theorem ensures that

$$\sqrt{\gamma_n^{-1}h_n}\left(C_n(t) - \mathbb{E}[C_n(t)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{2-(\alpha-a)\xi}\frac{R(K)}{\overline{G}(t)}\int_{\mathbb{R}}\exp(2xt)f(x)\,dx\right). \tag{24}$$

Now, in the case when $a > \alpha/5$, Part 1 of Theorem 1 follows from the combination of (9), (17), (20) and (24). Moreover, in the case when $a = \alpha/5$, Parts 1 and 2 of Theorem 1 follow from the combination of (8), (17), (20) and (24). In the case $a < \alpha/5$, (11) implies that

$$h_n^{-2}\left(C_n(t) - \mathbb{E}(C_n(t))\right) \xrightarrow{\mathbb{P}} 0,$$

and the combination of (8), (17) and (20) gives Part 2 of Theorem 1.

□

Funding

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

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgement

The authors are indebted to the Editor-in-Chief, the Associate Editor and the referee for their valuable comments and suggestions, which led to a considerable improvement of the manuscript.

REFERENCES

1. Akaike, H. (1954). An approximation to the characteristic function. *Ann. Inst. Statist. Math., Tokyo*, **6**, 127–132.
2. Bojanic, R. and Seneta, E. (1973). A unified theory of regularly Varying sequences. *Math. Z.*, **134**, 91–106.
3. Bouzebda, S. and El-hadjali, T. (2020). Uniform convergence rate of the kernel regression estimator adaptive to intrinsic dimension in presence of censored data. *J. Nonparametr. Stat.*, **32**(4), 864–914.
4. Bouzebda, S. and Nemouchi, B. (2020). Uniform consistency and uniform in bandwidth consistency for nonparametric regression estimates and conditional U-statistics involving functional data. *J. Nonparametr. Stat.*, **32**(2), 452–509.

5. Bouzebda, S. and Slaoui, Y. (2020). Nonparametric recursive method for kernel-type function estimators for censored data. *J. Stoch. Anal.*, **1**(3), Art. 4, 19.
6. Bouzebda, S. and Slaoui, Y. (2022). Nonparametric recursive method for moment generating function kernel-type estimators. *Statist. Probab. Lett.*, page 109422.
7. Bouzebda, S. and Soukarieh, I. (2023a). Non-parametric conditional U-processes for locally stationary functional random fields under stochastic sampling design. *Mathematics*, **11**(1), 1–70.
8. Bouzebda, S., Nezzal, A., and Zari, T. (2023b). Uniform consistency for functional conditional U-statistics using delta-sequences. *Mathematics*, **11**(1), 1–39.
9. Clarkson, J. A. and Adams, C. R. (1933). On definitions of bounded Variation for functions of two Variables. *Trans. Amer. Math. Soc.*, **35**(4), 824–854.
10. Csörgő, S. and Welsh, A. H. (1989). Testing for exponential and Marshall-Olkin distributions. *J. Statist. Plann. Inference*, **23**(3), 287–300.
11. Deheuvels, P. and Einmahl, J. H. J. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications. *Ann. Probab.*, **28**(3), 1301–1335.
12. Devroye, L. (1987). *A course in characteristic estimation*, volume 14 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA.
13. Duflo, M. (1997). Random iterative models. Collection Applications of Mathematics, *Springer, Berlin*.
14. Eggermont, P. P. B. and LaRiccia, V. N. (2001). *Maximum penalized likelihood estimation. Vol. I*. Springer Series in Statistics. Springer-Verlag, New York. characteristic estimation.
15. Epanečnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Teor. Verojatnost. i Primenen.*, **14**, 156–162.
16. Epps, T. W., Singleton, K. J., and Pulley, L. B. (1982). A test of separate families of distributions based on the empirical moment generating function. *Biometrika*, **69**(2), 391–399.
17. Földes, A. and Rejtő, L. (1981). A LIL type result for the product limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**(1), 75–86.
18. Galambos, J. and Seneta, E. (1973). Regularly Varying sequences. *Proc. Amer. Math. Soc.*, **41**, 110–116.
19. Gbur, E. E. and Collins, R. A. (1989). Estimation of the moment generating function. *Comm. Statist. Simulation Comput.*, **18**(3), 1113–1134.
20. Henze, N. and Visagie, J. (2020). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Ann. Inst. Statist. Math.*, **72**(5), 1109–1136.
21. Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.
22. Kushner and Yin (2003). Stochastic approximation and recursive algorithms and applications, *Stoch. Model. Appl. Probab.*, **35** *Springer*.
23. Ljung, L. (1978). Strong convergence of a stochastic approximation algorithm, *Ann. Statist.* **6**, 680–696.
24. Meintanis, S. G. (2007). A Kolmogorov-Smirnov type test for skew normal distributions based on the empirical moment generating function. *J. Statist. Plann. Inference*, **137**(8), 2681–2688.
25. Mokkadem, A. and Pelletier, M. (2007). Compact law of the iterated logarithm for matrix-normalized sums of random vectors. *Theory Probab. Appl.*, **52**(4), 2459–2478.
26. Mokkadem, A. Pelletier, M. and Slaoui, Y. (2009a). The stochastic approximation method for the estimation of a multiVariate probability density. *J. Statist. Plann. Inference*, **139**, 2459–2478.
27. Mokkadem, A. Pelletier, M. and Slaoui, Y. (2009b). Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA. Lat. Am. J. Probab. Math. Stat.*, **6**, 63–114.
28. Nadaraya, E. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9**(1), 141–142.
29. Parzen, E. (1962). On estimation of a probability characteristic function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
30. Quandt, R. E. and Ramsey, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.*, **73**(364), 730–752. With comments and a rejoinder by the authors.
31. Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.*, **3**(2), 213–238. With comments and a rejoinder by the author.
32. Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes. I. *Studia Sci. Math. Hungar.*, **8**, 391–398.
33. Révész, P. (1977). How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist. Ser. Statist.*, **8**(1), 119–126.
34. Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Anal. Math. Statist.*, **22**, 400–407.
35. Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), 41–55.
36. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a characteristic function. *Ann. Math. Statist.*, **27**, 832–837.
37. Scott, D. W. (1992). *Multivariate characteristic estimation*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. Theory, practice, and visualization, A Wiley-Interscience Publication.
38. Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
39. Slaoui, Y. (2014a). Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *Journal of Probability and Statistics*, **2014**, ID 739640, doi:10.1155/2014/739640.
40. Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23**, 306–325.
41. Slaoui, Y. (2015). Plug-In order selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.*, **69**, 483–509.
42. Soukarieh, I. and Bouzebda, S. (2023). Renewal type bootstrap for increasing degree U-process of a Markov chain. *J. Multivariate Anal.*, **195**, Paper No. 105143.

43. Yamato, H. (1972/73). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, **15**(3-4), 69–78.
44. Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.