



HAL
open science

Nonparametric recursive method for moment generating function kernel-type estimators

Salim Bouzebda, Yousri Slaoui

► **To cite this version:**

Salim Bouzebda, Yousri Slaoui. Nonparametric recursive method for moment generating function kernel-type estimators. *Statistics and Probability Letters*, 2022, 184, pp.109422. 10.1016/j.spl.2022.109422 . hal-04389629

HAL Id: hal-04389629

<https://univ-poitiers.hal.science/hal-04389629>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Nonparametric Recursive Method for Moment Generating Function Kernel-Type Estimators

Salim Bouzebda^{2*} and Yousri Slaoui^{3†}

¹LMAC, Université de Technologie de Compiègne, France

²Univ. Poitiers, Lab. Math. et Appl., Futuroscope Chasseneuil, France

Abstract

In the present paper, we are mainly concerned with the kernel type estimators for the moment generating function. More precisely, we establish the central limit theorem together 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. Finally, we investigate the performance of the methodology for small samples through a short simulation study.

Key words and phrases: Moment generating function; Kernel type estimator; Stochastic approximation algorithm.
MSC: 62G08, 62L20, 60F10, 62G07.

1 Introduction

Over years ago, Parzen (1962a) studied some properties of kernel density estimators introduced by Akaike (1954) and Rosenblatt (1956). Nonparametric regression function estimation has been the subject of intense investigation by both statisticians and probabilists for many years and this has led to the development of a large variety of methods. Kernel nonparametric function estimation methods have long attracted a great deal of attention, for good sources of references to research literature in this area along with statistical applications consult Devroye and Györfi (1985), Devroye (1987), Scott (1992), Wand and Jones (1995), Eggermont and LaRiccia (2001), Bouzebda and Nemouchi (2020); Bouzebda and El-hadjali (2020); Bouzebda *et al.* (2021) 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 only 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. Quandt and Ramsey (1978) used the moment generating function to develop a method of estimating the parameters of a mixture of normal distributions. Epps *et al.* (1982) used on the empirical moment generating function to construct a test of separate families of distributions. Saddlepoint methods for approximating the pdf of a sample mean involves the moment generating function of the underlying distribution (e.g.; Reid (1988)). Csörgő and Welsh (1989) 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. Gbur and Collins (1989) investigated the parametric moment generating function. In the work of Meintanis (2007) 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. Henze and Visagie (2020) 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 in any dimension. In this paper we will consider the nonparametric recursive kernel-type estimators for the multivariate moment generating function. Recursive estimation, was proposed first in Robbins and Monro (1951) and further investigation in many directions was given by Ljung (1978), Tsybakov (1990), Duflo (1997), Kushner and Yin (2003), Mokkadem *et al.* (2009a,b), Slaoui (2014b, 2015).

This work concerns a nonparametric estimation of the recursive general kernel-type estimators for moment generating function defined by the stochastic approximation algorithm. To the best of our knowledge, the results presented here respond to a problem that has not been studied systematically up to the present, which was the basic motivation of the paper.

We start by giving some notation and definitions that are needed for the forthcoming sections. Let d be a fixed natural number, and $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be a sequence of independent d -dimensional random vectors with common distribution function $F(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and probability density function $f(\cdot)$ with respect to the Lebesgue measure. Suppose that the moment generating function

$$C(\mathbf{t}) = \int \exp(\langle \mathbf{t}, \mathbf{x} \rangle) dF(\mathbf{x}), \quad \mathbf{t} = (t_1, \dots, t_d) \in \mathbb{R}^d,$$

*e-mail address: Salim.Bouzebda@utc.fr

†e-mail address: Yousri.Slaoui@math.univ-poitiers.fr

exists on a non-degenerate d -dimensional subset I of \mathbb{R}^d , necessarily containing the origin, where $\langle \mathbf{t}, \mathbf{x} \rangle = \sum_{k=1}^d t_k x_k$ denotes the usual inner product. Throughout the paper we write bold letters for vectors, e.g., $\mathbf{x} = (x_1, \dots, x_d)$ is a d -dimensional vector. For $\mathbf{u} := (u_1, \dots, u_d) \in \mathbb{R}^d$ and $\mathbf{v} := (v_1, \dots, v_d) \in \mathbb{R}^d$, we write $\mathbf{u} \leq \mathbf{v}$ when $u_j \leq v_j$ for $j = 1, \dots, d$ and $|\mathbf{t}| = \langle \mathbf{t}, \mathbf{t} \rangle^{\frac{1}{2}}$ and $\|\mathbf{t}\| = \max(|t_1|, \dots, |t_d|)$ will denote the length and maximum-norm on \mathbb{R}^d . For each $n \geq 1$ and $\mathbf{x} \in \mathbb{R}^d$, let $\mathbf{F}_n(\cdot)$ be the empirical distribution function of $\mathbf{X}_1, \dots, \mathbf{X}_n$, i.e.,

$$\mathbf{F}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \leq \mathbf{x}\}. \quad (1.1)$$

When considered as a pointwise estimator of $F(\cdot)$, $\mathbf{F}_n(\cdot)$ is an unbiased and strongly consistent estimator of $F(\cdot)$ (see [Serfling \(1980\)](#) for more details). Based on the empirical distribution function, we introduce the empirical moment generating function

$$\mathbf{C}_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n \exp(\langle \mathbf{t}, \mathbf{X}_j \rangle) = \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) d\mathbf{F}_n(\mathbf{x}), \quad \text{for } \mathbf{t} \in \mathbb{R}^d.$$

Now, given the information that $F(\cdot)$ is absolutely continuous, it is more appropriate to consider a smooth estimator of $F(\cdot)$ rather than the empirical distribution function. For this purpose, let $K(\cdot)$ be a kernel function on \mathbb{R}^d ($\int_{\mathbb{R}^d} K(\mathbf{z}) d\mathbf{z} = 1$), $\mathcal{K}(\mathbf{z}) = \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_d} K(u_1, \dots, u_d) du_1 \dots du_d = \int_{-\infty}^{\mathbf{z}} K(\mathbf{u}) d\mathbf{u}$ and consider the standard kernel estimator of $F(\cdot)$ based on $\mathbf{X}_1, \dots, \mathbf{X}_n$

$$\tilde{\mathbf{F}}_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \mathcal{K}\left(\frac{\mathbf{x} - \mathbf{X}_j}{h_n}\right), \quad \text{for } \mathbf{x} \in \mathbb{R}^d, \quad (1.2)$$

where $h_n > 0$ is a bandwidth parameter. This estimator was considered for the first time in [Nadaraya \(1964\)](#), and it is constructed by integrating out the Parzen–Rosenblatt kernel density estimate ([Rosenblatt \(1956\)](#) and [Parzen \(1962b\)](#)),

$$\tilde{\mathbf{F}}_n(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} f_n(\mathbf{u}) d\mathbf{u},$$

where $f_n(\mathbf{x}) = \frac{1}{nh_n^d} \sum_{j=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_j}{h_n}\right)$. The almost sure uniform consistency of $\tilde{\mathbf{F}}_n(\cdot)$ was established in [Yamato \(1973\)](#) with the only smoothness condition that $F(\cdot)$ be continuous, while [Yukich \(1989\)](#) extended this result to higher dimensions. Moreover, it has been shown by several authors that the asymptotic performance of $\tilde{\mathbf{F}}_n(\cdot)$ is better than that of the empirical distribution function $\mathbf{F}_n(\cdot)$ for an appropriate choice of the kernel $K(\cdot)$ and on the bandwidth h_n , for more details see [Azzalini \(1981\)](#), [Reiss \(1981\)](#), [Falk \(1983\)](#), [Swanepoel \(1988\)](#), [Abdous \(1993\)](#) and [Swanepoel \(2021\)](#). A typical measure of accuracy of $\tilde{\mathbf{F}}_n(\cdot)$ is the Mean Weighted Integrated Squared Error (MWISE), defined as

$$MWISE(\tilde{\mathbf{F}}_n) := \mathbb{E} \left[\int_{\mathbb{R}^d} \left\{ \tilde{\mathbf{F}}_n(\mathbf{x}) - F(\mathbf{x}) \right\}^2 w(\mathbf{x}) dF(\mathbf{x}) \right], \quad (1.3)$$

where $w(\cdot)$ is some weight function. Several authors have derived asymptotic expressions for (1.3), see for instance [Swanepoel \(1988\)](#), [Jones \(1990\)](#) and [Altman and Léger \(1995\)](#). Notice that the classical kernel estimator of $C(\mathbf{t})$ is defined to be

$$\tilde{\mathbf{C}}_n(\mathbf{t}) = \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) d\tilde{\mathbf{F}}_n(\mathbf{x}) = \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f_n(\mathbf{x}) d\mathbf{x}, \quad \text{for } \mathbf{t} \in \mathbb{R}^d. \quad (1.4)$$

The main motivation of the current work is based of the following class of estimators of the distribution function F introduced in [Slaoui \(2014b\)](#)

$$\hat{\mathbf{F}}_n(\mathbf{x}) = (1 - \gamma_n) \hat{\mathbf{F}}_{n-1}(\mathbf{x}) + \gamma_n \mathcal{K}\left(\frac{\mathbf{x} - \mathbf{X}_n}{h_n}\right). \quad (1.5)$$

The estimator (1.5) was proposed in [Slaoui \(2014b\)](#) by using the Robbins-Monro algorithm, 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 $\sum_{n \geq 1} \gamma_n^2 < \infty$ in order to ensure the almost sure convergence (see [Duflo \(1997\)](#)), and the bandwidth (h_n) is a sequence of positive real numbers that goes to zero. [Slaoui \(2014b\)](#) showed that for an appropriate choice of the stepsize (γ_n), the bandwidth (h_n) and the kernel $K(\cdot)$, the estimator $\hat{\mathbf{F}}_n(\cdot)$ outperformed the estimator $\tilde{\mathbf{F}}_n(\cdot)$ in terms of MWISE, the author consider the weight function equal to the density function ($w(\cdot) = f(\cdot)$). It comes that

$$d\hat{\mathbf{F}}_n(\mathbf{x}) = (1 - \gamma_n) d\hat{\mathbf{F}}_{n-1}(\mathbf{x}) + \gamma_n h_n^{-d} K\left(\frac{\mathbf{x} - \mathbf{X}_n}{h_n}\right) d\mathbf{x}. \quad (1.6)$$

In this paper, we consider the following recursive estimator

$$\widehat{\mathbf{C}}_n(\mathbf{t}) = \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) d\widehat{\mathbf{F}}_n(\mathbf{x}), \quad \text{for } \mathbf{t} \in \mathbb{R}^d, \quad (1.7)$$

By combining equations (1.6) and (1.7), it follows that

$$\widehat{\mathbf{C}}_n(\mathbf{t}) = (1 - \gamma_n) \widehat{\mathbf{C}}_{n-1}(\mathbf{t}) + \frac{\gamma_n}{h_n^d} \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) K\left(\frac{\mathbf{x} - \mathbf{X}_n}{h_n}\right) d\mathbf{x}. \quad (1.8)$$

This recursive scheme offers many advantages to recursive estimators: they are of easy implementation and they do not require extensive storage of data. 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 fact that the estimate updating is independent of the history of the data providing a decisive computational advantage. The main drawback of the classical kernel estimator is the use of all data at each step of estimation. The estimators that we consider is given in more general form including as particular case the estimator $\widetilde{\mathbf{C}}_n(\mathbf{t})$, see Remark 2.

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 by means of 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

Through-out this paper, let us unburden our notation by writing

$$\mu_j(K) = \int_{\mathbb{R}^d} z_j^2 K(\mathbf{z}) d\mathbf{z}, \quad R(K) = \int_{\mathbb{R}^d} K^2(\mathbf{z}) d\mathbf{z}, \quad \text{where } \mathbf{z} = (z_1, \dots, z_d),$$

and

$$\xi = \lim_{n \rightarrow +\infty} (n\gamma_n)^{-1}. \quad (2.1)$$

First of all, let us set the following definition of a 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

$$(v_n)_{n \geq 1} \in \mathcal{GS}(\gamma) \text{ if } \lim_{n \rightarrow +\infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \quad (2.2)$$

Condition (2.2) was introduced by Galambos and Seneta (1973) to define regularly varying sequences (see also Bojanic and Seneta (1995)). 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 theoretical main results, we need the following assumptions.

Assumptions:

(A1) $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is a continuous bounded function satisfying:

$$\int_{\mathbb{R}^d} K(\mathbf{z}) d\mathbf{z} = 1, \quad \forall j \in \{1, \dots, d\}, \quad \int_{\mathbb{R}^d} z_j K(\mathbf{z}) d\mathbf{z} = 0 \quad \text{and} \quad \mu_j(K) = \int_{\mathbb{R}^d} z_j^2 K(\mathbf{z}) d\mathbf{z} < \infty;$$

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

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

Discussion on the assumptions:

- Assumptions (A1) and (A3) are standard in the framework of nonparametric kernel estimation (see for instance Slaoui (2014a)).
- Assumption (A2) is widely used on the stochastic approximation algorithms (see for instance Mokkadem et al. (2009a)).

- Assumption **(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 $([n\gamma_n]^{-1})$ is finite.
- To understand better the use the assumption **(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 $(a_n b_n) \in \mathcal{GS}(a+b)$ and $(a_n b_n^{-1}) \in \mathcal{GS}(a-b)$. For a sequences v_n belonging to $\mathcal{GS}(\gamma)$ with positive γ , we have $\lim_{n \rightarrow \infty} v_n = \infty$ and for sequences w_n belonging to $\mathcal{GS}(\beta)$ with negative β , we have $\lim_{n \rightarrow \infty} w_n = 0$. Then, it comes from **(A2)**(i) that, $\gamma_n \rightarrow 0$, $\sum_n \gamma_n = \infty$ and $\sum_n \gamma_n^2 < \infty$, the assumption **(A2)**(ii) ensures that $h_n \rightarrow 0$ and $\gamma_n/h_n^d \rightarrow 0$, the assumption **(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 **(A2)**, ensures that the bias and the variance depend only on h_n and not on h_1, \dots, h_n .

3 Main results

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

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

1. If $a \in \left(0, \frac{\alpha}{d+4}\right]$, then

$$\mathbb{E}[\widehat{\mathbf{C}}_n(\mathbf{t})] - \mathbf{C}(\mathbf{t}) = \frac{h_n^2}{2(1-2a\xi)} \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o(h_n^2). \quad (3.1)$$

2. If $a \in \left(\frac{\alpha}{d+4}, 1\right)$, then

$$\mathbb{E}[\widehat{\mathbf{C}}_n(\mathbf{t})] - \mathbf{C}(\mathbf{t}) = o\left(\sqrt{\gamma_n h_n^{-d}}\right). \quad (3.2)$$

3. If $a \in \left(0, \frac{\alpha}{d+4}\right)$, then

$$\text{Var}[\widehat{\mathbf{C}}_n(\mathbf{t})] = o(h_n^4). \quad (3.3)$$

4. If $a \in \left[\frac{\alpha}{d+4}, 1\right)$, then

$$\text{Var}[\widehat{\mathbf{C}}_n(\mathbf{t})] = \frac{\gamma_n}{h_n^d} \frac{1}{2 - (\alpha - ad)\xi} R(K) \int_{\mathbb{R}} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o\left(\frac{\gamma_n}{h_n^d}\right). \quad (3.4)$$

The bias and the variance of the estimator $\widehat{\mathbf{C}}_n(\cdot)$ defined by the stochastic approximation algorithm (1.8) then heavily depend on the choice of the stepsize (γ_n) .

Remark 1. The combination of (3.1) and (3.4), ensure that, when $a = \frac{\alpha}{d+4}$, we can obtain the asymptotic expression of the bias and the variance of the proposed estimator together, and then, we can obtain the weak pointwise convergence rate (see Theorem 1 and Corollary 2). Then, the case $a = \frac{\alpha}{d+4}$ is a very interesting case. We can also observe that, when $(\gamma_n) \in \mathcal{GS}(-1)$ (e.g. $(\gamma_n) = (n^{-1})$), $\alpha = 1$.

Remark 2. The estimator $\widetilde{\mathbf{C}}_n(\mathbf{t})$, can be written recursively as follows:

$$\widetilde{\mathbf{C}}_n(\mathbf{t}) = \left(1 - \frac{1}{n}\right) \widetilde{\mathbf{C}}_{n-1}(\mathbf{t}) + \frac{\gamma_n}{h_n^d} \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) K\left(\frac{\mathbf{x} - \mathbf{X}_n}{h_n}\right) d\mathbf{x}.$$

Then, the estimator $\widetilde{\mathbf{C}}_n(\mathbf{t})$ is a special case of the considered estimator $\widehat{\mathbf{C}}_n(\mathbf{t})$, with the choice $(\gamma_n) = (n^{-1})$.

The following Corollary is a consequence of the Proposition 1.

Corollary 1 (Bias and variance of $\widetilde{\mathbf{C}}_n(\mathbf{t})$). Let Assumptions **(A1)**, **(A2)** (i) – (ii) and **(A3)** hold.

$$\mathbb{E}[\widetilde{\mathbf{C}}_n(\mathbf{t})] - \mathbf{C}(\mathbf{t}) = \frac{(d+4)}{2(d+2)} h_n^2 \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o(h_n^2), \quad (3.5)$$

and

$$\text{Var}[\widetilde{\mathbf{C}}_n(\mathbf{t})] = \frac{(d+4)}{(2d+4)} \frac{\gamma_n}{h_n^d} R(K) \int_{\mathbb{R}} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o\left(\frac{\gamma_n}{h_n^d}\right). \quad (3.6)$$

Now, let us state the following theorems which gives respectively the asymptotic normality of the generalized recursive estimator $\widehat{\mathbf{C}}_n(\cdot)$ defined in (1.8) and the generalized nonrecursive estimator $\widetilde{\mathbf{C}}_n(\cdot)$ defined in (1.4).

3.1 Asymptotic normality of $\widehat{\mathbf{C}}_n(\cdot)$ and $\widetilde{\mathbf{C}}_n(\mathbf{t})$

Let us now state the following theorem, which gives the weak convergence rate of the estimator $\widehat{\mathbf{C}}_n(\cdot)$ defined in (1.8). 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 1 (Weak pointwise convergence rate of $\widehat{\mathbf{C}}_n(\cdot)$). Let the assumptions **(A1)**-**(A3)** hold.

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

$$\begin{aligned} & \sqrt{\gamma_n^{-1}h_n^d}(\widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \\ & \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2(1-2a\xi)} \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x}, \right. \\ & \left. \frac{1}{2 - (\alpha - ad)\xi} R(K) \int_{\mathbb{R}} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} \right). \end{aligned}$$

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

$$\frac{1}{h_n^2}(\widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \xrightarrow{\mathbb{P}} \frac{1}{2(1-2a\xi)} \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x}.$$

The following corollary is an immediate consequence of Theorem 1.

Corollary 2 (Weak pointwise convergence rate of $\widetilde{\mathbf{C}}_n(\mathbf{t})$). Let the assumptions **(A1)**, **(A2)** (i) – (ii) and **(A3)** hold.

1. If there exists $c \geq 0$ such that $nh_n^{d+4} \rightarrow c$, then

$$\begin{aligned} & \sqrt{nh_n^d}(\widetilde{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \\ & \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}\left(\frac{\sqrt{c}}{2} \frac{(d+4)}{(d+2)} \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x}, \frac{(d+4)}{(2d+4)} R(K) \int_{\mathbb{R}} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} \right). \end{aligned}$$

2. If $nh_n^{d+4} \rightarrow \infty$, then

$$\frac{1}{h_n^2}(\widetilde{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \xrightarrow{\mathbb{P}} \frac{1}{2} \frac{(d+4)}{(d+2)} \sum_{j=1}^d t_j^2 \mu_j^2(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x}.$$

Theorem 2 (Uniform convergence of $\widehat{\mathbf{C}}_n(\mathbf{t})$). Let the assumptions **(A1)**-**(A3)** hold, $f(\cdot)$ is uniformly continuous and there exists $\eta > 0$ such that $\mathbf{z} \rightarrow \|\mathbf{z}\|^\eta |f(\mathbf{z})|$ is a bounded function. We let \mathcal{C} be a compact set of \mathbb{R}^d . Then, we have

$$\sup_{\mathbf{t} \in \mathcal{C}} \left| \widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right| = o(1) \quad \text{a.s. as } n \rightarrow \infty.$$

The following corollary is an immediate consequence of 2.

Corollary 3 (Uniform convergence of $\widetilde{\mathbf{C}}_n(\mathbf{t})$). Let the assumptions **(A1)**, **(A2)** (i) – (ii) and **(A3)** hold, $f(\cdot)$ is uniformly continuous and there exists $\eta > 0$ such that $\mathbf{z} \rightarrow \|\mathbf{z}\|^\eta |f(\mathbf{z})|$ is a bounded function. We let \mathcal{C} be a compact set of \mathbb{R}^d . Then, we have

$$\sup_{\mathbf{t} \in \mathcal{C}} \left| \widetilde{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right| = o(1) \quad \text{a.s. as } n \rightarrow \infty.$$

Remark 3. 1. The rate of convergence of the recursive estimator $\widehat{\mathbf{C}}_n(\mathbf{t})$ is $\sqrt{\gamma_n^{-1}h_n^d}$, while the rate of convergence of the recursive estimator $\widetilde{\mathbf{C}}_n(\mathbf{t})$ is $\sqrt{nh_n^d}$.

2. In the case when $(\gamma_n) = (n^{-1})$, the bias, variance and the rate of convergence of the two estimators $\widehat{\mathbf{C}}_n(\mathbf{t})$ and $\widetilde{\mathbf{C}}_n(\mathbf{t})$ are the same.
3. Clearly the empirical distribution function is a recursive estimator, it follows from (1.1), that

$$\mathbf{F}_n(\mathbf{x}) = \left(1 - \frac{1}{n}\right) \mathbf{F}_{n-1}(\mathbf{x}) + \frac{1}{n} \mathbb{1}\{\mathbf{X}_n \leq \mathbf{x}\}, \quad \text{which gives } \mathbf{C}_n(\mathbf{t}) = \left(1 - \frac{1}{n}\right) \mathbf{C}_{n-1}(\mathbf{t}) + \frac{1}{n} \exp(\langle \mathbf{t}, \mathbf{X}_n \rangle).$$

4 Simulation results

In this section, series of experiments are conducted in order to examine the performance of the proposed estimators given in (1.8). The computing program codes are implemented in R. The setup of our simulation study closely follows that of Slaoui (2014b). More precisely, we consider the case of drawing i.i.d. univariate random samples $X_i, i = 1, \dots, n$. We consider the exponential $\mathcal{E}(1)$, the standard normal $\mathcal{N}(0, 1)$ and the uniform $\mathcal{U}[0, 1]$ distributions. 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 Epanečnikov (1969) kernel:

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

- the quadratic kernel :

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

Here, h is the smoothing bandwidth. We adopt the “normal scale rule” or the rule-of-thumb method, see for instance Silverman (1986), 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 (1.8). 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.,

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

where A is an appropriately chosen set. In our simulation $A = [0, 001, 10]$. We have chosen the uniform discretization t_1, \dots, t_ℓ with $\ell = 50$ of $[-0, 10, 0, 10]$. The choice of α_h is not the optimal one, since we are choosing this in order 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 $\widetilde{\mathbf{C}}_n(t)$), let

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

The same remark that $C(t)$ can be replaced by the empirical counterparts based, for example, on 10000 simulations. 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, 2 and 3 local MSE for $\widetilde{\mathbf{C}}_n(t)$), (4, 5 and 6 local MSE for $\widehat{\mathbf{C}}_n(t)$) allows us to deduce that for large values of the sample size n gives smaller MSE. In the case of the normal distribution, the obtained are very satisfactory for the both estimators. This can be justified by the fact that the moment generating function for the normal distribution $C(t) = \exp(t^2/2)$ is very smooth. However the results for the uniform distribution are not similar to those for the normal distribution. This being said, the main problem is the difficulty to estimate the moment generating function for the uniform distribution $C(t) = (\exp(t) - 1)/t$ in the neighborhood of 0. However, one

can see that MSE is less than 10^{-4} in all cases for the uniform distribution. One can see that the moment generating function for the exponential distribution $C(t) = 1/(1-t)$ is not smooth like for the normal distribution. Figures (8, 9 and 10 for $\tilde{\mathbf{C}}_n(t)$) (11, 12 and 13 for $\hat{\mathbf{C}}_n(t)$) corroborate the preceding remark. Figures (7 and 14) display the results for $\mathbf{C}_n(t)$, where we can see the good performance of the estimator for the different distributions. Figures ((15, 16 and 17, local variance results) and (22, 23 and 24, local bias results) for $\tilde{\mathbf{C}}_n(t)$), ((18, 19 and 20, local variance results) and (25, 26 and 27, local bias results) for $\hat{\mathbf{C}}_n(t)$) display the results for the bias and the variance for the non-recursive and the recursive estimators. As in the results for the MSE, we have good performance of the estimators for the normal distribution. We do not have the same accuracy for the uniform distribution. Figures (21 (local variance) and 28 (local bias)) display the results for $\mathbf{C}_n(t)$, where we have the same conclusion as for the local MSE, the estimator behaves well in the different settings.

5 Conclusion

In this paper we have considered the estimation of the nonparametric moment generating function. 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 non trivial mathematics, that goes well beyond the scope of the present paper. It would be interesting to extend our work to the case of censored data that may be applied in several areas. We plan also to consider some bandwidth selection procedures like in [Slaoui \(2014a,b\)](#) with some real data applications related to this context.

6 Proofs

This section is devoted to the proof of our results. The previously presented notation continues to be used in the following. Throughout this section we use the following notation:

$$\begin{aligned}\Pi_n &= \prod_{j=1}^n (1 - \gamma_j), \\ Z_n(\mathbf{t}) &= h_n^{-d} \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) K\left(\frac{\mathbf{x} - \mathbf{X}_n}{h_n}\right) d\mathbf{x}.\end{aligned}\tag{6.1}$$

Before giving the outlines of the proofs, we state the following technical lemma, which is proved in [Mokkadem et al. \(2009a\)](#), and which is widely applied throughout the demonstrations.

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 (2.1). We have

$$\lim_{n \rightarrow +\infty} v_n \Pi_n^m \sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} = \frac{1}{m - v^*\xi}.\tag{6.2}$$

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

$$\lim_{n \rightarrow +\infty} v_n \Pi_n^m \left[\sum_{k=1}^n \Pi_k^{-m} \frac{\gamma_k}{v_k} \alpha_k + \delta \right] = 0.\tag{6.3}$$

Let us underline that the application of Lemma 1 requires Assumption (A2)(iii) on the limit of $(n\gamma_n)$ as n goes to infinity.

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

First, in view of (6.1), we infer that

$$\begin{aligned}\hat{\mathbf{C}}_n(\mathbf{t}) - \mathbf{C}(\mathbf{t}) &= (1 - \gamma_n) \left(\hat{\mathbf{C}}_{n-1}(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right) + \gamma_n (Z_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \\ &= \sum_{k=1}^{n-1} \left[\prod_{j=k+1}^n (1 - \gamma_j) \right] \gamma_k (Z_k(\mathbf{t}) - \mathbf{C}(\mathbf{t})) + \gamma_n (Z_n(\mathbf{t}) - \mathbf{C}(\mathbf{t})) + \left[\prod_{j=1}^n (1 - \gamma_j) \right] \left(\hat{\mathbf{C}}_0(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right) \\ &= \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k (Z_k(\mathbf{t}) - \mathbf{C}(\mathbf{t})) + \Pi_n \left(\hat{\mathbf{C}}_0(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right).\end{aligned}\tag{6.4}$$

This readily implies that

$$\mathbb{E} \left(\widehat{\mathbf{C}}_n(\mathbf{t}) \right) - \mathbf{C}(\mathbf{t}) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k \left(\mathbb{E} [Z_k(\mathbf{t})] - \mathbf{C}(\mathbf{t}) \right) + \Pi_n \left(\widehat{\mathbf{C}}_0(\mathbf{t}) - \mathbf{C}(\mathbf{t}) \right). \quad (6.5)$$

Taylor's expansion with integral remainder ensures that

$$\begin{aligned} \mathbb{E} [Z_k(\mathbf{t})] - \mathbf{C}(\mathbf{t}) &= \int_{\mathbb{R}^{2d}} \{ \exp(\langle \mathbf{t}, \mathbf{x} + h_k \mathbf{z} \rangle) - \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \} K(\mathbf{z}) f(\mathbf{x}) d\mathbf{z} d\mathbf{x} \\ &= \frac{h_k^2}{2} \sum_{j=1}^d t_j^2 \mu_j(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + h_k^2 \delta_k(\mathbf{t}), \end{aligned} \quad (6.6)$$

where

$$\delta_k(\mathbf{t}) = h_k^{-2} \int_{\mathbb{R}^{2d}} f(\mathbf{x}) K(\mathbf{z}) \left[\{ \exp(\langle \mathbf{t}, \mathbf{x} + h_k \mathbf{z} \rangle) - \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \} - \frac{h_k^2}{2} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \sum_{j=1}^d (t_j z_j)^2 \right] d\mathbf{x} d\mathbf{z}.$$

We have $\lim_{k \rightarrow \infty} \delta_k(\mathbf{t}) = 0$. In the case $0 < a \leq \alpha/(d+4)$, we have $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$; the application of Lemma 1 then gives

$$\begin{aligned} \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] - \mathbf{C}(\mathbf{t}) &= \frac{1}{2} \sum_{j=1}^d t_j^2 \mu_j(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} \left\{ \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^2 [1 + o(1)] \right\} \\ &\quad + \Pi_n (C_0(\mathbf{t}) - \mathbf{C}(\mathbf{t})) \\ &= \frac{1}{2(1-2a\xi)} \sum_{j=1}^d t_j^2 \mu_j(K) \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} [h_n^2 + o(1)], \end{aligned}$$

and (3.1) follows. In the case $\alpha/(d+4) < a < 1$, we have $h_n^2 = o(\sqrt{\gamma_n h_n^{-d}})$. Since we have $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - ad)/2$, the application of Lemma 1 gives

$$\mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] - \mathbf{C}(\mathbf{t}) = \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k o\left(\sqrt{\gamma_k h_k^{-d}}\right) + O(\Pi_n) = o\left(\sqrt{\gamma_n h_n^{-d}}\right),$$

which gives (3.2). Now, since $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ is a sequence of independent d -dimensional random vectors with common distribution function $F(\mathbf{x})$, we have $\text{Cov}(Z_k(\mathbf{t}), Z_{k'}(\mathbf{t})) = 0$ for $k \neq k'$, then, it comes that

$$\begin{aligned} \text{Var} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] &= \Pi_n^2 \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var} [Z_k(\mathbf{t})] \\ &= \Pi_n^2 \sum_{k=1}^n \frac{\Pi_k^{-2} \gamma_k^2}{h_k^d} \left[\int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} + \mathbf{z} h_k \rangle) K(\mathbf{z}) d\mathbf{z} \right\} \left\{ \int_{\mathbb{R}^d} \exp(\langle \mathbf{t}, \mathbf{x} + \mathbf{z}' h_k \rangle) K(\mathbf{z}') d\mathbf{z}' \right\} f(\mathbf{x}) d\mathbf{x} \right. \\ &\quad \left. - h_k^d \left(\int_{\mathbb{R}^{2d}} K(\mathbf{z}) \exp(\langle \mathbf{t}, \mathbf{x} - \mathbf{z} h_k \rangle) f(\mathbf{x}) d\mathbf{x} d\mathbf{z} \right)^2 \right] \\ &= \Pi_n^2 \sum_{k=1}^n \frac{\Pi_k^{-2} \gamma_k^2}{h_k^d} \left[R(K) \int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + \nu_k(\mathbf{t}) - h_k^d \tilde{\nu}_k(\mathbf{t}) \right], \end{aligned} \quad (6.7)$$

where

$$\begin{aligned} \nu_k(\mathbf{t}) &= \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \{ \exp(\langle \mathbf{t}, \mathbf{x} + \mathbf{z} h_k \rangle) - \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \} K(\mathbf{z}) d\mathbf{z} \right\} \\ &\quad \times \left\{ \int_{\mathbb{R}^d} \{ \exp(\langle \mathbf{t}, \mathbf{x} + \mathbf{z}' h_k \rangle) - \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \} K(\mathbf{z}') d\mathbf{z}' \right\} f(\mathbf{x}) d\mathbf{x} \\ \tilde{\nu}_k(\mathbf{t}) &= \left(\int_{\mathbb{R}^{2d}} \exp(\langle \mathbf{t}, \mathbf{x} + \mathbf{z} h_k \rangle) K(\mathbf{z}) f(\mathbf{x}) d\mathbf{x} d\mathbf{z} \right)^2. \end{aligned}$$

In view of **(A3)**, we have $\lim_{k \rightarrow \infty} \nu_k(\mathbf{t}) = 0$ and $\lim_{k \rightarrow \infty} h_k^d \tilde{\nu}_k(\mathbf{t}) = 0$, we let $\varepsilon_k(\mathbf{t}) = \nu_k(\mathbf{t}) - h_k^d \tilde{\nu}_k(\mathbf{t})$, we have $\lim_{k \rightarrow \infty} \varepsilon_k(\mathbf{t}) = 0$. In the case $\alpha/(d+4) \leq a < 1$, we have $\lim_{n \rightarrow \infty} (n\gamma_n) > (\alpha - ad)/2$, we make use of Lemma 1 to infer that

$$\begin{aligned} \text{Var} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] &= \Pi_n^2 \sum_{k=1}^n \frac{\Pi_k^{-2} \gamma_k^2}{h_k^d} \left[R(K) \int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + \varepsilon_k(\mathbf{t}) \right] \\ &= \frac{1}{2 - (\alpha - ad)\xi} \frac{\gamma_n}{h_n^d} \left[R(K) \int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o(1) \right], \end{aligned}$$

which gives (3.4). When $0 < a < \alpha/(d+4)$, we have $\gamma_n h_n^{-d} = o(h_n^4)$. Then, since $\lim_{n \rightarrow \infty} (n\gamma_n) > 2a$, we apply Lemma 1 to infer that

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

which proves (3.3). □

6.2 Proof of Theorem 1

First, it comes from (6.4) and (6.5), that

$$\widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] = \Pi_n \sum_{k=1}^n Y_k(\mathbf{t}),$$

where

$$Y_k(\mathbf{t}) = \Pi_k^{-1} \gamma_k (Z_k(\mathbf{t}) - \mathbb{E}(Z_k(\mathbf{t}))).$$

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

$$\begin{aligned} v_n^2 &= \sum_{k=1}^n \text{Var}(Y_k(\mathbf{t})) = \sum_{k=1}^n \Pi_k^{-2} \gamma_k^2 \text{Var}(Z_k(\mathbf{t})) \\ &= R(K) \sum_{k=1}^n \frac{\Pi_k^{-2} \gamma_k^2}{h_k^d} \left[\int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o(1) \right] \\ &= R(K) \frac{1}{\Pi_n^2} \frac{\gamma_n}{h_n^d} \left[\frac{1}{2 - (\alpha - ad)\xi} \int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} + o(1) \right]. \end{aligned}$$

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

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

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

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

and we thus obtain

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

Then the application of Lyapunov's Theorem ensures that

$$\sqrt{\gamma_n^{-1} h_n^d} \left(\widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] \right) \xrightarrow{D} \mathcal{N} \left(0, \frac{1}{2 - (\alpha - ad)\xi} R(K) \int_{\mathbb{R}^d} \exp(2\langle \mathbf{t}, \mathbf{x} \rangle) f(\mathbf{x}) d\mathbf{x} \right). \quad (6.8)$$

Now, in the case when $a > \alpha/(d+4)$, Part 1 of Theorem 1 follows from the combination of (3.2) and (6.8). Moreover, in the case when $a = \alpha/(d+4)$, Parts 1 and 2 of Theorem 1 follow from the combination of (3.1) and (6.8). In the case $a < \alpha/(d+4)$, (3.3) implies that

$$h_n^{-2} \left(\widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbb{E} \left(\widehat{\mathbf{C}}_n(\mathbf{t}) \right) \right) \xrightarrow{\mathbb{P}} 0,$$

and the application of (3.1) gives Part 2 of Theorem 1. □

6.3 Proof of Theorem 2

First, using the compactness property of the set \mathcal{C} , we infer that, for some $(\mathbf{t}_k)_{1 \leq k \leq \gamma_n}$, $\mathcal{C} \subset \bigcup_{k=1}^{\gamma_n} B(\mathbf{t}_k, a_n)$, with $a_n = h_n^{d+1}$, where $B(\mathbf{t}_k, a_n)$, is an open ball of centre \mathbf{t}_k and radius a_n . Now, for any $\mathbf{t} \in \mathcal{C}$, we set $\tilde{k}(\mathbf{t}) = \arg \min_k \|\mathbf{t}_k - \mathbf{t}\|$. Then, for any $\mathbf{t} \in \mathcal{C}$, we have

$$\begin{aligned} \sup_{\mathbf{t} \in \mathcal{C}} \left| \widehat{\mathbf{C}}_n(\mathbf{t}) - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] \right| &\leq \sup_{\mathbf{t} \in \mathcal{C}} \left| \widehat{\mathbf{C}}_n(\mathbf{t}) - \widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) \right| + \sup_{\mathbf{t} \in \mathcal{C}} \left| \widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) \right] \right| \\ &\quad + \sup_{\mathbf{t} \in \mathcal{C}} \left| \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) \right] - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}) \right] \right| =: \mathcal{T}_{1,n} + \mathcal{T}_{2,n} + \mathcal{T}_{3,n}. \end{aligned} \quad (6.9)$$

First, it follows from (1.8) and from some analysis considerations that for any $\mathbf{t} \in \mathcal{C}$

$$\left| \widehat{\mathbf{C}}_n(\mathbf{t}) - \widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) \right| \leq \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k^{-d} \int_{\mathbb{R}^d} \left| \exp(\langle \mathbf{t}, \mathbf{x} \rangle) - \exp(\langle \mathbf{t}_{\tilde{k}}, \mathbf{x} \rangle) \right| \left| K \left(\frac{\mathbf{X}_k - \mathbf{x}}{h_k} \right) \right| d\mathbf{x} \leq \mathfrak{C} \|dK\|_{\infty} \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k h_k,$$

we then get $\mathcal{T}_{1,n} = o(1)$ and $\mathcal{T}_{3,n} = o(1)$. Now, we set $\rho > 0$ and M such that

$$\|f\|_{\infty} \int_{\|\mathbf{z}\| > M} |K(\mathbf{z})| d\mathbf{z} \leq \mathfrak{C}.$$

Moreover, Lemma 1 ensures that $\Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k = 1 + o(1)$, then, it comes that

$$\begin{aligned} \left| \widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) - \mathbb{E} \left[\widehat{\mathbf{C}}_n(\mathbf{t}_{\tilde{k}}) \right] \right| &\leq \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k |\mathbb{E} [Z_k(\mathbf{t})]| \\ &\leq \mathfrak{C} + \int_{\|\mathbf{z}\| \leq M} \exp(\langle \mathbf{t}, \mathbf{x} \rangle) |f(\mathbf{x})| |K(\mathbf{z})| d\mathbf{z} \\ &\quad + \exp(\langle \mathbf{t}, \mathbf{x} \rangle) \Pi_n \sum_{k=1}^n \Pi_k^{-1} \gamma_k \int_{\|\mathbf{z}\| > M} |K(\mathbf{z})| |f(\mathbf{x} - \mathbf{z}h_k) - f(\mathbf{x})| d\mathbf{z}. \end{aligned}$$

Then, the uniform continuity of $f(\cdot)$ combined with the dominate convergence and the existence of $\eta > 0$ such that $\mathbf{z} \rightarrow \|\mathbf{z}\|^\eta |f(\mathbf{x})|$ is a bounded function ensure that $\mathcal{T}_{2,n} = o(1)$. Then the combination of (3.1) and (6.9) concludes the proof of Theorem 2. \square

Supplementary Material

upplement to "Nonparametric Recursive Method for Moment Generating Function Kernel-Type Estimators" (,.pdf). We present here the figures for the simulations.

Funding

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

Acknowledgement

The authors are indebted to the Editor-in-Chief, the Associate Editor and the referees for their very valuable comments and suggestions which led to a considerable improvement of the manuscript. The authors thank Mr. Issam Elhattab for his important help with the simulation section.

References

- Abdous, B. (1993). Note on the minimum mean integrated squared error of kernel estimates of a distribution function and its derivatives. *Comm. Statist. Theory Methods*, **22**, 603–609.
- Akaike, H. (1954). An approximation to the characteristic function. *Ann. Inst. Statist. Math.*, Tokyo, **6**, 127–132.
- Altman, N. and Léger, Ch. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inference*, **46**, 195–214.
- Azzalini, A. (1981). A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method *Biometrika*, **68**, 326–328.

- Bojanic, R. and Seneta, E. (1973). A unified theory of regularly varying sequences. *Math. Z.*, **134**, 91–106.
- 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.
- Bouzebda, S., Elhattab, I., and Nemouchi, B. (2021). On the uniform-in-bandwidth consistency of the general conditional U -statistics based on the copula representation. *Journal of Nonparametric Statistics*, **33**(2), 321–358.
- 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.
- 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.
- Collomb, G. (1981). Estimation non-paramétrique de la régression: revue bibliographique. *Internat. Statist. Rev.*, **49**(1), 75–93.
- Csörgő, S. and Welsh, A. H. (1989). Testing for exponential and Marshall-Olkin distributions. *J. Statist. Plann. Inference*, **23**(3), 287–300.
- Devroye, L. (1987). *A course in characteristic estimation*, volume 14 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA.
- Devroye, L. and Györfi, L. (1985). *Nonparametric characteristic estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L_1 view.
- Duflo, M. (1997). Random iterative models. Collection Applications of Mathematics, *Springer, Berlin*.
- 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.
- Epanečnikov, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Teor. Veroyatnost. i Primenen.*, **14**, 156–162.
- 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.
- Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Stat. Neerl.*, **69**, 483–509.
- Galambos, J. and Seneta, E. (1973). Regularly varying sequences. *Proc. Amer. Math. Soc.*, **41**, 110–116.
- Gbur, E. E. and Collins, R. A. (1989). Estimation of the moment generating function. *Comm. Statist. Simulation Comput.*, **18**(3), 1113–1134.
- 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.
- Jones, M. C. (1990). The Performance of Kernel Density Functions in Kernel Distribution Function Estimation, *Statist. Probab. Lett.*, **9**, 129–132.
- Kushner and Yin (2003). Stochastic approximation and recursive algorithms and applications, *Stoch. Model. Appl. Probab.*, **35** Springer.
- Ljung, L. (1978). Strong convergence of a stochastic approximation algorithm, *Ann. Statist.* **6**, 680–696.
- 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.
- 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.
- 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.
- 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.
- Nadaraya, E. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9**(1), 141–142.

- Parzen, E. (1962a). On estimation of a probability characteristic function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
- Parzen, E. (1962b). On estimation of a probability density and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- 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.
- Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.*, **3**(2), 213–238. With comments and a rejoinder by the author.
- Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions. *Scand. J. Statist.*, **8**, 116–119.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Anal. Math. Statist.*, **22**, 400–407.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a characteristic function. *Ann. Math. Statist.*, **27**, 832–837.
- 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.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- 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.
- Slaoui, Y. (2014b). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23**, 306–325.
- Slaoui, Y. (2015). Plug-In order selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.*, **69**, 483–509.
- Swanepoel, J. W. H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function. *Comm. Statist. Theory Methods*, **17**, 3785–3799.
- Swanepoel, J. W. H. (2021). The law of the iterated logarithm and maximal smoothing principle for the kernel distribution function estimator. *J. Nonparametr. Stat.*, **33**(1), 156–169.
- Tsybakov, A. B. (1990). Recurrent estimation of the mode of a multidimensional distribution.
- Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, **15**(3-4), 69–78.
- Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.
- Yukich, J. E. (1989). A note on the limit theorems for perturbed empirical process. *Stochastic Process. Appl.*, **33**, 163–173.