Statistics and Its Interface Volume 13 (2020) 399–410

Estimation of a distribution function using Lagrange polynomials with Tchebychev-Gauss points

Salima Helali* and Yousri Slaoui†

The estimation of the distribution function of a real random variable is an intrinsic topic in non parametric estimation. To this end, a distribution estimator based on Lagrange polynomials and Tchebychev-Gauss points, is introduced. Some asymptotic properties of the proposed estimator are investigated, such as its asymptotic bias, variance, mean squared error and Chung-Smirnov propriety. The asymptotic normality and the uniform convergence of the estimator are also established. Lastly, the performance of the proposed estimator is explored through a certain simulation examples.

AMS 2000 subject classifications: Primary 62E20, 97N50; secondary 41A50.
Keywords and phrases: Distribution estimator, Lagrange polynomials, Tchebychev-Gauss points, Asymptotic properties.

1. INTRODUCTION

Non parametric distribution estimation is undoubtedly a useful tool of data analysis, which is reflected by the multiple literary works addressing the topic. Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables having a common unknown distribution function F with associated density f supported on a compact interval. Within the framework of the nonparametric estimation, since we know that F is continuous, we consider the estimation of F by using smooth functions rather than the empirical distribution function, which is not continuous. Several methods have been set forward for smooth estimation of density and distribution functions. The most popular one, called kernel method, is introduced by Rosenblatt [22]. The advances were carried out by Parzen [21] to estimate density function. The kernel distribution estimator was identified by Nadaraya [19] as

$$\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}\left(\frac{x-X_i}{h_n}\right), \tag{1}$$

where $\mathcal{K}(z) = \int_{-\infty}^{z} K(u)du$, K is a kernel function and (h_n) is a bandwidth. The properties of $\tilde{F}_n$ have been known for a long time, for example its uniform convergence towards F with continuous f (Nadaraya [19], Winter [35], Yamato [37]), then unconditionally on f (Singh et al. [26]) and its asymptotic normality (Watson and Leadletter [34]). Winter [36] also demonstrated that $\tilde{F}_n$ checks the Chung-Smirnov property with probability 1.

However, Kernel methods have estimation problems at the edges, when we have a random variable X with distribution function supported on a compact interval. In order to overcome this problem, various methods such as the Bernstein polynomial density and distribution estimators were introduced first by Vitale [32] and then extended by Tenbusch [31], Babu et al. [2] and Kakizawa [14]. In particular, following Babu et al. [2], the Bernstein estimator of order $\nu > 0$ of the distribution F is defined as

$$\overline{F}_{n,\nu}(x) = \sum_{k=0}^{\nu}\widehat{F}_n\left(k/\nu\right)b_k(\nu, x), \tag{2}$$

with $\widehat{F}_n$ is the empirical distribution function and $b_k(\nu, x) = C_\nu^k x^k(1-x)^{\nu-k}$ is the Bernstein polynomial. This estimator is asymptotically unbiased. Babu et al. [2] found also that $\overline{F}_{n,\nu}$ to be uniformly strongly consistent. Babu and Chaubey [3] adapted the Bernstein estimator to the problem of estimating a multivariate distribution function (including the case of dependent observations under α mixing). Leblanc [15] reported that it has the Chung-Smirnov property, as $n \to \infty$.

In this paper, we present what appears to be a new method based on Lagrange polynomials and Tchebytchev-Gauss points. When we have a random variable X with distribution F supported on a compact interval $[a, b]$ such as $a < b$, we can transform X into Y, a random variable with support $[-1, 1]$ through the transformation $Y = \dfrac{X-(a+b)/2}{(b-a)/2}$. Transformations such as $Y = 2X/(1+X) - 1$ and $Y = 2\pi^{-1}\arctan(X)$ can be used to cover the cases of random variables X with support $\mathbb{R}_+$ and $\mathbb{R}$ respectively. Once the random variable X is transformed into Y, we can apply Lagrange polynomials with

*Corresponding author.
†First supporter of the project.

Tchebytchev-Gauss points to approximate the distribution function of Y on the interval $[-1,1]$. In the theoretical part of this paper, we consider the case where f is supported on $[-1,1]$, and we propose an estimator of order $m>0$ of the distribution F using Lagrange polynomial expressed as,

$$\widetilde{F}_{n,m}(x)=\sum_{i=1}^{m}\widehat{F}_n(x_i)\mathcal{L}_i(x), \tag{3}$$

where, for all $i = 1\dots m$, $x_i=\cos\left((2i-1)\pi/2m\right)$ are Tchebytchev-Gauss points, $\mathcal{L}_i(x)=\prod_{\substack{j=1\\ j\neq i}}^{m}\frac{x-x_j}{x_i-x_j}$ is the Lagrange polynomial, and $\widehat{F}_n$ denotes the empirical distribution function obtained from a random sample of size n. The points $(x_i)_{1\leq i\leq m}$ are the zeros of the Tchebytchev polynomial $T_m(x)=\cos(m\arccos(x))$. They are also the optimal choice of grids that give the best convergence $\sum_{i=1}^{m}v(x_i)\mathcal{L}_i(.)\to v(.)$ uniformly, when $m\to\infty$, for any continuous function v of class C^k (for $k\geq 1$) on the interval $[-1,1]$. This result was studied by Jackson in the early 20th century (see [10], [11]). His results can also be found in [5], chapter 4, section 6, page 147, which is the main idea of the proposed estimator. As an excellent reference for properties of Lagrange polynomial with Tchebytchev-Gauss points in the deterministic case, we refer the reader to Austin [1]. To the best of our knowledge, the estimator presented here has not been studied so far, which stands for the basic motivation of the paper. The main objective of this paper is to study the properties of the distribution estimator (3). We consider first the mean squared error for a fixed x, for $-1<x<1$, and split it into bias squared and variance terms. Then, we establish the uniform convergence of this estimator, the Chung-Smirnov property and the (pointwise) asymptotic normality of the proposed estimator. Basically, the remainder of the paper is organized as follows. In the next section, we display the assumptions and notations. In Section 3, we exhibit our main results. Section 4 highlights a simulation study that compares the performance of the proposed estimator $\tilde{F}_{n,m}$ with the Bernstein estimator (2) and with the kernel (standard Gaussian kernel) estimator (1). Section A provides the proofs of our theoretical results.

2. ASSUMPTIONS AND NOTATIONS

We consider the following definition.

Definition 2.1.
Let g be a function defined on $[-1,1]$. g is said to be Lipschitz of order $\alpha\in(0,1]$ if there exists a positive constant c such that

$$\mid g(x)-g(y)\mid\leq c\mid x-y\mid^{\alpha},$$

for all $x,y\in[-1,1]$. For convenience, we write $g\in Lip(\alpha,c)$.

To study the asymptotic behaviours of the estimator (3) inside the interval $[-1,1]$, the following assumption is considered:

$(\mathcal{A}_1)$ F is of class C^2 on $[-1,1]$.
$(\mathcal{A}_2)$ f and f' are bounded.

Throughout this paper, we let $i=1\dots m$, $x\in[-1,1]$ for $m\geq 1$, and we consider the following notations:

$\theta_i=(2i-1)\pi/2m$, $\sigma^2(x)=F(x)(1-F(x))$,
$x_i=\cos(\theta_i)$: Tchebytchev-Gauss points,
$A_m(x)=\sum_{i=1}^{m}F(x_i)\mathcal{L}_i(x)$,
$b(x)=f(x)/2+f'(x)(x-1)/4-f''(x)(1+x^2-2x)/12$,
$\mathcal{L}_i(x)=\prod_{\substack{j=1\\ j\neq i}}^{m}\frac{x-x_j}{x_i-x_j}$: Lagrange polynomial,
$T_m(x)=\cos(m\arccos(x))$: Tchebytchev polynomial.

3. MAIN RESULTS

Our first result is the following proposition which sets forward the bias and the variance of $\widetilde{F}_{n,m}$.

Proposition 3.1 (Bias and variance of $\widetilde{F}_{n,m}$).
Under assumption $(\mathcal{A}_1)$, we have for $x\in[-1,1]$,

$$Bias(\widetilde{F}_{n,m}(x)) = \pi m^{-2}T_m(x)b(x)+o(m^{-2}), \tag{4}$$
$$Var(\widetilde{F}_{n,m}(x)) = n^{-1}\sigma^2(x)+O(n^{-1}m^{-1/2}). \tag{5}$$

Notice that for $x\in]0,1[$, the bias of the Bernstein estimator $\overline{F}_{n,\nu}$ and the bias of the kernel estimator $\tilde{F}_n$ are given respectively by

$$\begin{aligned} Bias(\overline{F}_{n,\nu}(x)) &= \nu^{-1}b(x)+o(\nu^{-1}),\\ Bias(\tilde{F}_n(x)) &= \frac{1}{2}h^2f'(x)\mu_2(K)+o(h^2),\end{aligned}$$

where $\mu_2(K)=\int_0^1 z^2K(z)dz$. The previous result implies that, in the case when $\nu=m$, the bias of the estimator $\widetilde{F}_{n,m}$ is $O(m^{-2})$ is smaller than the one obtained using the Bernstein polynomial, which has a bias of order $O(m^{-1})$. On the one hand, if we consider $h=m^{-1}$ and f' is bounded, we notice that the bias of $\widetilde{F}_{n,m}$ is $O(m^{-2})=O(h^2)$, which is asymptotically similar to the bias obtained using the kernel estimator $\tilde{F}_n$, that is generally $O(h^2)$ except near the boundaries. On the other hand, if f is bounded, it is well known that the variance of the Bernstein estimator and the variance of the kernel estimator are given respectively by

$$\begin{aligned} Var(\overline{F}_{n,\nu}(x)) &= n^{-1}\sigma^2(x)+O(\nu^{-1/2}n^{-1}),\\ Var(\tilde{F}_n(x)) &= n^{-1}\sigma^2(x)+O(hn^{-1}).\end{aligned}$$

In this respect, another consequence of the previous result is that in the case when $\nu=m$, the variance of $\widetilde{F}_{n,m}$ is asymptotically similar to the variance of the estimator obtained using Bernstein polynomial. On the other side, in order to compare the proposed estimator and the kernel estimator, we consider some classical choices, which are $m = n$ and $h = n^{-1/3}$, this choice is motivated by the optimal bandwidth based on the minimization of the MSE. We notice that in the case where f is bounded and $x \in]0,1[$, the variance of $\widetilde{F}_{n,m}$ is $n^{-1}\sigma^2(x)+O(n^{-3/2})$, which is asymptotically smaller than the variance obtained using kernel estimator, namely $n^{-1}\sigma^2(x) + O(n^{-4/3})$. In addition, it is well known that

$$MSE(\widehat{F}_n(x)) = Var(\widehat{F}_n(x)) = n^{-1}\sigma^2(x).$$

In conclusion, regarding the performance of the proposed estimator, we point out that

- The three considered estimators and the empirical distribution $\widehat{F}_n$ are asymptotically equivalent in terms of MSE up to the first order.
- The proposed estimator asymptotically dominates the Bernstein estimator $\overline{F}_{n,\nu}$ in terms of bias and in terms of MSE in the case when f is bounded.
- Under the assumption $(\mathcal{A}_2)$, the proposed estimator is asymptotically similar to the kernel estimator $\tilde{F}_n$ in terms of bias without any additional assumptions, and dominates the kernel estimator in terms of MSE under some classical conditions.

We complete our study with the following proposition which reveals that $\widetilde{F}_{n,m}$ is strongly consistent.

Proposition 3.2 (**Uniform convergence of $\widetilde{F}_{n,m}$**). *Under assumption $(\mathcal{A}_1)$, if $n, m \to \infty$, then*

$$\left\| \widetilde{F}_{n,m} - F \right\| \to 0 \quad a.s.,$$

where $\|K\| = \sup_{x\in[-1,1]} |K(x)|$ *for any bounded function K on* $[-1, 1]$.

In this paper, we prove that the estimator $\widetilde{F}_{n,m}$ satisfied the Chung-Smirnov property, which quantifies its extreme fluctuations about F, as $m \to \infty$, under certain regularity conditions on F. Let G_n be any estimator of the distribution function F. Therefore, G_n is said to satisfy the Chung-Smirnov property when

$$\limsup_{n\to\infty} \left(\frac{2n}{\log\log n}\right)^{1/2} \sup_{x\in[-1,1]} |G_n(x) - F(x)| \le 1, \tag{6}$$

a.s.

We know that the empirical distribution function $\widehat{F}_n$ satisfies the above property. To be more accurate, we have

$$\limsup_{n\to\infty} \left(\frac{2n}{\log\log n}\right)^{1/2} \sup_{x\in[-1,1]} \left|\widehat{F}_n(x) - F(x)\right| = 1. \tag{7}$$

This was proved by Chung [6] and [30]. The following proposition demonstrates that $\widetilde{F}_{n,m}$ satisfies this property under certain conditions.

Proposition 3.3 (**Chung Smirnov property for $\widetilde{F}_{n,m}$**). *Let $F \in Lip(\alpha, c)$ for some $c > 0$. If $m, n \to \infty$ and $\sqrt{n}m^{-\alpha/2} \to 0$, then $\widetilde{F}_{n,m}$ satisfies equation (6).*

Finally, the following proposition indicates the asymptotic normality of the estimator (3).

Proposition 3.4 (**Asymptotic normality of $\widetilde{F}_{n,m}$**). *Assume $(\mathcal{A}_1)$ holds and $m, n \to \infty$. For $x \in (-1, 1)$, we have*

$$n^{1/2}\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)).$$

Note that, under an appropriate choice of bandwidth, a result similar to proposition 3.4 was recorded by [34] for general kernel estimators, and by [17] for the Bernstein estimator of distribution functions.

4. NUMERICAL STUDIES

4.1 Comparison with estimators (1) and (2)

In this section, we investigate the performance of the proposed estimator in estimating different distributions by comparing it to the performances of Bernstein estimator and of the standard Gaussian kernel estimator. We can apply Bernstein estimator and the proposed estimator when the sample is concentrated on the intervals $[0, 1]$ and $[-1, 1]$, respectively. In order to enact the comparison between the estimators (1), (2) and (3), applicable in general, we list below suggested transformations in different cases:

(1) Suppose that X is concentrated on a finite support $[a, b]$, then we work with the sample values $Y_1, \ldots, Y_n$ where $Y_i = (X_i - a)/(b - a)$.
(2) For the distributions functions concentrated on $\mathbb{R}$, we can use the transformed sample $Y_i = 1/2 + \pi^{-1}\arctan(X_i)$ which transforms the range to the interval $[0, 1]$.
(3) For the support $\mathbb{R}_+$, we can use the transformed sample $Y_i = X_i/(1 + X_i)$ which transforms the range to the interval $[0, 1]$.

In our simulation study, six sample sizes are considered, $n = 10$, $n = 50$, $n = 100$, $n = 150$, $n = 200$, $n = 250$ and the following distribution functions:

1− The beta distribution $\mathcal{B}(3, 2)$,
2− The beta distribution $\mathcal{B}(2, 2)$,
3− The gamma distribution $\mathcal{G}(1, 6)$,
4− The mixture beta distribution $0.5\mathcal{B}(2.5, 6) + 0.5\mathcal{B}(9, 1)$.

For each distribution function and sample size n, we compute the Integrated Squared Error (ISE) of the estimator over $N = 500$ trials,

$$ISE[\widehat{F}] = \int_0^1 \left(\widehat{F}(x) - F(x)\right)^2 dx, \tag{8}$$

where $\widehat{F}$ is an estimator of the distribution F. To select the smoothing parameters m, ν and h, we consider the Monte

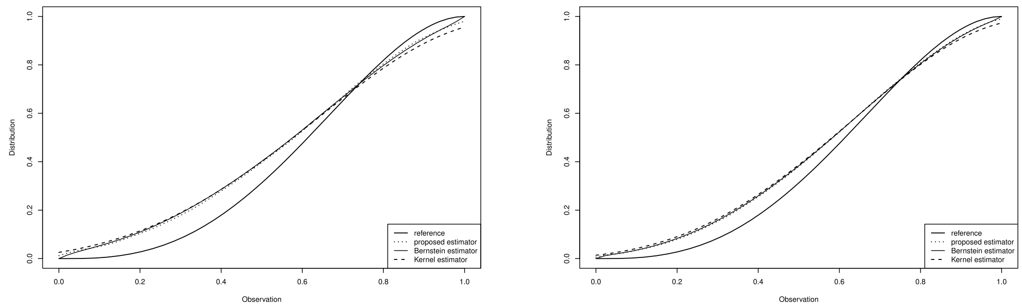

Figure 1. Qualitative comparison between the estimator $\overline{F}_{n,\nu}(x)$ *defined in (2),* $\tilde{F}_n$ *defined in (1) and the proposed distribution estimator* $\widetilde{F}_{n,m}$ *defined in (3), for* $N = 500$ *samples of size* $n = 50$ *(left panel) and of size* $n = 100$ *(right panel) for the beta distribution* $\mathcal{B}(3,2)$.

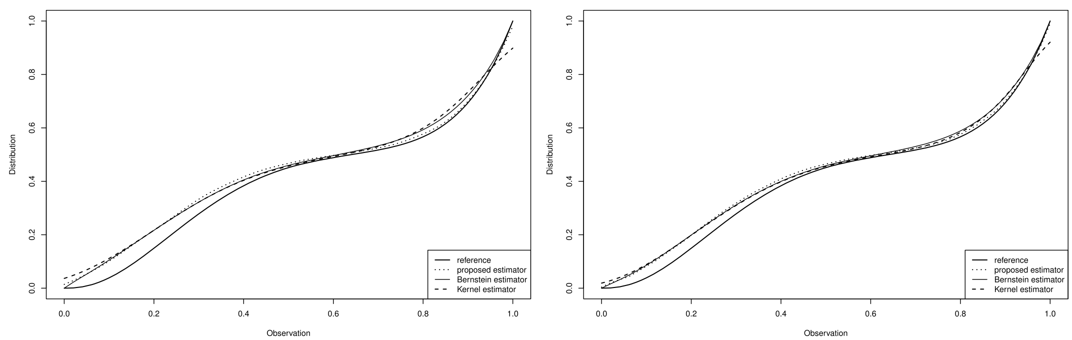

Figure 2. Qualitative comparison between the estimator $\overline{F}_{n,\nu}(x)$ *defined in (2),* $\tilde{F}_n$ *defined in (1) and the proposed distribution estimator* $\widetilde{F}_{n,m}$ *defined in (3), for* $N = 500$ *samples of size* $n = 50$ *(left panel) and of size* $n = 100$ *(right panel) for the exponential distribution* $0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$.

Carlo procedure for each point $x \in [0,1]$. We determine the parameters m (for $1 \leq m \leq 300$), ν (for $1 \leq \nu \leq 300$) and h (for $h = i/1000$ with $1 \leq i \leq 300$), which minimizes ISE, which is approximated by

$$\frac{1}{N}\sum_{i=1}^{N} ISE_i(\widehat{F}),$$

where $ISE_i(\widehat{F})$ is the value of ISE computed from the ith sample of size n and obtained from (8).

From figures 1–2 and tables 1–2, we conclude that

- In the considered distributions (1)–(4), by choosing the appropriate m, ν and h, the ISE of the distribution estimator (3) is smaller than that of Kernel estimator (1) and Bernstein estimator (2) even when the sample size is very large.
- The ISE decreases as the sample size increases.

4.2 Real dataset

We consider two examples that highlight the features of the proposed estimator $\widetilde{F}_{n,m}$:

1. At first time, the data show 50 alignments of a coding `DNA` sequence of the growth factor receptor of a Norwegian rat `EGFR` (Rattus norvegicus egfr gene, partial cds), which is available in the site https://www.ncbi.nlm.nih.gov/. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 50$ for our proposed estimator, $m = 35$ for the Bernstein estimator and $h = 0.636438$ for the kernel estimator.
2. At the second time, we used `Salvister` data which appear in `R` package `kerdiest` (Quintela-del-Río and Estévez-Pérez [20]). These data contain 85 observations of the annual peak instantaneous flow levels of the Salt River near Roosevelt, AZ, USA, for the period 1924–2009, obtained from the National Water Information System. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 85$ for our proposed estimator, $m = 80$ for the Bernstein estimator and $h = 0.06$ for the kernel estimator.
3. The third data show the failure time (breakdowns of electronic devices) in operating hours. These data contain 18 observations and are introduced by [33]. For con-

Table 1. ISE for $N = 500$ trials of Bernstein estimator, standard Gaussian Kernel estimator and the proposed estimator $\widetilde{F}_{n,m}$, for $n = 10$, $n = 50$ and $n = 100$. The bold values indicates the smallest values of ISE

	n	Proposed estimator	Bernstein estimator	Kernel estimator
$\mathcal{B}(3,2)$	10	0.032331	**0.013258**	0.019944
	50	**0.003819**	0.004411	0.005014
	100	**0.002198**	0.002431	0.002999
$\mathcal{B}(2,2)$	10	0.009598	**0.006958**	0.012854
	50	**0.001302**	0.001717	0.002420
	100	$\mathbf{0.564e^{-3}}$	$0.802e^{-3}$	0.001132
$\mathcal{G}(1,6)$	10	**0.037654**	0.038798	0.040357
	50	**0.005205**	0.006879	0.006393
	100	**0.001780**	0.002236	0.002052
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	10	0.005359	**0.003579**	0.007807
	50	**0.001326**	0.001515	0.001767
	100	$\mathbf{0.699e^{-3}}$	$0.727e^{-3}$	$0.820e^{-3}$

Table 2. ISE for $N = 500$ trials of Bernstein estimator $\overline{F}_{n,\nu}(x)$, standard Gaussian kernel estimator $\tilde{F}_n$ and the proposed estimator $\widetilde{F}_{n,m}$, for $n = 150$, $n = 200$ and $n = 250$. The bold values indicates the smallest values of ISE

	n	Proposed estimator	Bernstein estimator	Kernel estimator
$\mathcal{B}(3,2)$	150	**0.001799**	0.002023	0.002342
	200	**0.001596**	0.001782	0.001763
	250	**0.001258**	0.001447	0.001462
$\mathcal{B}(2,2)$	150	$\mathbf{0.377e^{-3}}$	$0.489e^{-3}$	$0.718e^{-3}$
	200	$\mathbf{0.264e^{-3}}$	$0.327e^{-3}$	$0.522e^{-3}$
	250	$\mathbf{0.229e^{-3}}$	$0.289e^{-3}$	$0.392e^{-3}$
$\mathcal{G}(1,6)$	150	$\mathbf{0.540e^{-3}}$	$0.896e^{-3}$	$0.676e^{-3}$
	200	$\mathbf{0.107e^{-3}}$	$0.200e^{-3}$	$0.115e^{-3}$
	250	$\mathbf{2.921e^{-5}}$	$4.996e^{-5}$	$5.429e^{-5}$
$0.5\mathcal{B}(2.5,6) + 0.5\mathcal{B}(9,1)$	150	$0.503e^{-3}$	$\mathbf{0.501e^{-3}}$	$0.568e^{-3}$
	200	$\mathbf{0.379e^{-3}}$	$0.380e^{-3}$	$0.486e^{-3}$
	250	$\mathbf{0.309e^{-3}}$	$\mathbf{0.309e^{-3}}$	$0.354e^{-3}$

venience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 18$ for our proposed estimator, $m = 15$ for the Bernstein estimator and $h = 0.20559$ for the kernel estimator.

4. Moreover, we used `attenu` data which appear in `R` package `datasets` ([13]). These data contain 182 observations of the numeric moment magnitude at various stations for 23 earthquakes in California. For convenience, we analyzed the original data rescaled to the unit interval. Finally, we used the Monte Carlo method to obtain $m = 182$ for our proposed estimator, $m = 180$ for the Bernstein estimator and $h = 0.0305$ for the kernel estimator.

In the real examples, the three estimators are compared with the empirical distribution $\widehat{F}_n$. Then, for any considered estimator $\widehat{F}$ of the distribution function F, we propose to compute the ISE defined as:

$$ISE(\widehat{F}) = \int_0^1 \left(\widehat{F}(x) - \widehat{F}_n(x)\right)^2 dx.$$

Departing from Tables 3–4 and figures 3–6, we infer that the ISE of the proposed estimator is smaller than the ISE of the Bernstein estimator and the ISE of the kernel estimator, thus demonstrating the effectiveness of our considered estimator.

5. CONCLUSION

The central focus of this paper is upon suggesting an estimator of the distribution function using Lagrange polynomials and Tchebytchev-Gauss points. We showed that a few important properties contributing to the popularity of kernel estimator and Bernstein estimator of distribution function are also satisfied by the proposed estimator. The asymptotic laws of the proposed estimator are established under general conditions. We also argued that the proposed estimator asymptotically dominates the Bernstein estimator in terms of bias. Through a simulation study and a simple data set examples, we have demonstrated how the proposed estimator can lead to satisfactory estimates of the distribution function. To sum up, our simulations also suggest that the proposed estimator is quite promising and

Table 3. ISE of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $\widetilde{F}_{n,m}$, for S A DNA (Score Alignments DNA) data and Saltriver data. The bold values indicates the smallest values of ISE

Data set	Proposed estimator	Bernstein estimator	Kernel estimator
S A DNA	$\mathbf{0.334860e^{-3}}$	$0.647956e^{-3}$	$0.466513e^{-3}$
Saltriver	$\mathbf{4.586125e^{-5}}$	$0.112049e^{-3}$	$0.756750e^{-3}$

Table 4. ISE of Bernstein estimator, standard Gaussian kernel estimator and the proposed estimator $\widetilde{F}_{n,m}$, for magnitude data and failure time data. The bold values indicates the smallest values of ISE

Data set	Proposed estimator	Bernstein estimator	Kernel estimator
Failure time	**0.000549**	0.000760	0.001088
Magnitude	**0.000223**	0.000929	0.000695

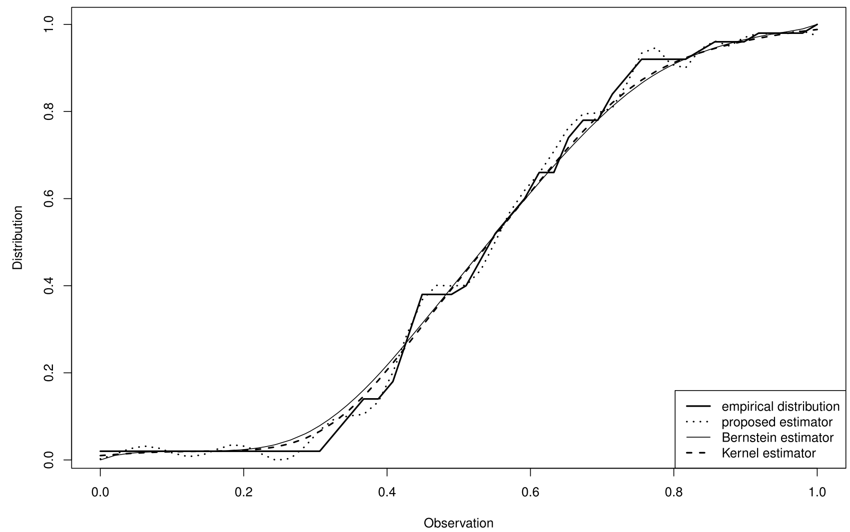

Figure 3. Qualitative comparison between the estimator $\overline{F}_{n,\nu}$ defined in (2), $\tilde{F}_n$ defined in (1) and the proposed distribution estimator $\widetilde{F}_{n,m}$ defined in (3), for S A DNA data.

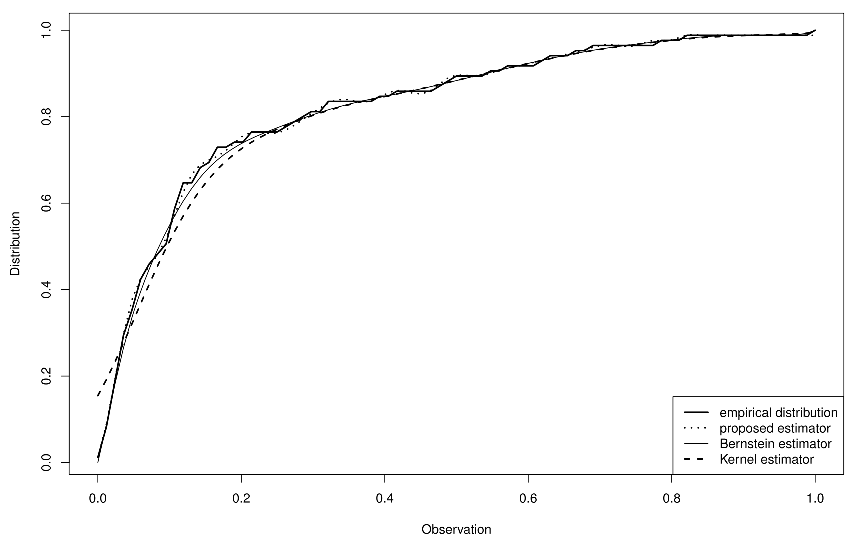

Figure 4. Qualitative comparison between the estimator $\overline{F}_{n,\nu}$ defined in (2), $\tilde{F}_n$ defined in (1) and the proposed distribution estimator $\widetilde{F}_{n,m}$ defined in (3), for Saltriver data.

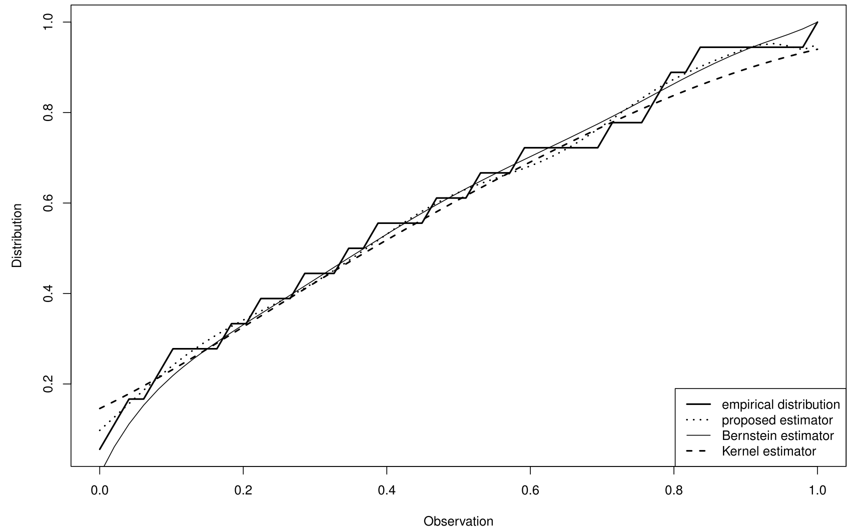

Figure 5. Qualitative comparison between the estimator $\overline{F}_{n,\nu}$ defined in (2), $\tilde{F}_n$ defined in (1) and the proposed distribution estimator $\widetilde{F}_{n,m}$ defined in (3), for failure time data.

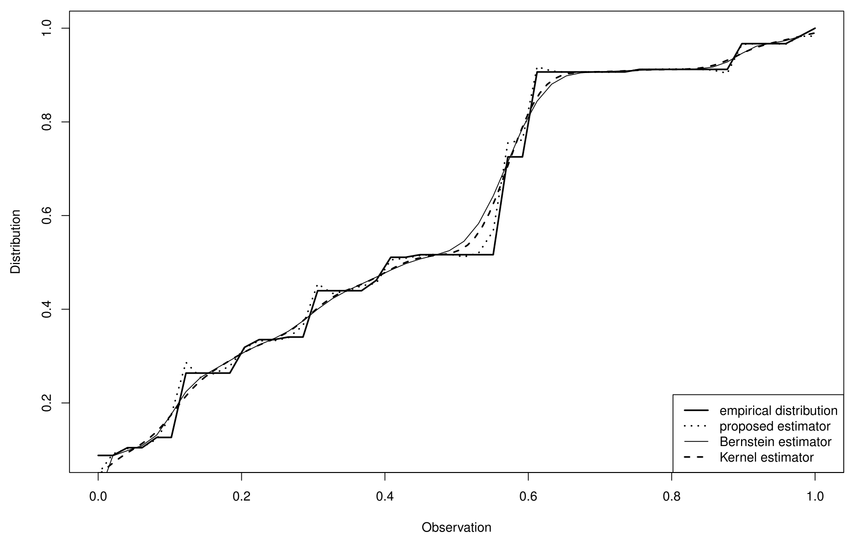

Figure 6. Qualitative comparison between the estimator $\overline{F}_{n,\nu}$ defined in (2), $\tilde{F}_n$ defined in (1) and the proposed distribution estimator $\widetilde{F}_{n,m}$ defined in (3), for magnitude data.

interesting as it behaves well when compared with both the Bernstein estimator and the Gaussian kernel estimator.

To this extent, we could simply assert that our work is a step may be taken further as it lays the ground and offers new perspectives for future works to extend this investigation by considering a recursive version and compare the obtained estimators to the one adopted by [27] and [12]. We plan also to consider the estimation of a density function in a recursive framework and then the estimation of a regression function in a recursive framework by using Lagrange polynomials (see [28], [29]).

APPENDIX A. PROOFS

Throughout the proofs, we use the following notations:

$$R_m^{(1)} = \sum_{k=1}^{m} \frac{\sin\theta_k}{\sin(m\theta_k)},$$

$$R_m^{(2)} = \sum_{k=1}^{m} \frac{x_k \sin\theta_k}{\sin(m\theta_k)},$$

$$R_m^{(3)} = \sum_{k=1}^{m} \frac{\cos(2\theta_k)\sin\theta_k}{\sin(m\theta_k)},$$

$$J_m(x) = \sum_{k=1}^{m} \mid x_k - x \mid \mathcal{L}_k^2(x),$$
$$S_m(x) = \sum_{k=1}^{m} \mathcal{L}_k^2(x),$$
$$\lambda_m(x) = \sum_{k=1}^{m} |\mathcal{L}_k(x)| \text{ Lebesgue function},$$
$$\Lambda_m = \max_{x\in[-1,1]} \lambda_m(x) \text{ Lebesgue constant, for } j \in \{0,1,2\},$$
$$P_{j,m}(x) = \sum_{\substack{k=1\\k<l}}^{m} (x_k - x)^j \mathcal{L}_k(x)\mathcal{L}_l(x).$$

In order to prove Theorems 3.1–3.4, we establish the following technical lemmas A.1 and A.2 stated below.

Lemma A.1. *For $m \geq 1$, we have*

$$R_m^{(1)} = \sin(\pi/2m),\ R_m^{(2)} = \sin(\pi/m)/2,$$
$$R_m^{(3)} = (\sin(3\pi/2m) - \sin(\pi/2m))/2.$$

Lemma A.2. *For $x \in [-1,1]$, we have*

i)

$$\sum_{k=1}^{m}(x_k - x)\mathcal{L}_k(x) = -\frac{T_m(x)}{m}R_m^{(1)} = -\frac{\pi}{2m^2}T_m(x) + o(m^{-2}),$$

ii)

$$\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x) = \frac{T_m(x)}{m}\left(xR_m^{(1)} - R_m^{(2)}\right) = \frac{\pi}{2m^2}T_m(x)(x-1) + o(m^{-2}),$$

iii)

$$\sum_{k=1}^{m}(x_k - x)^3\mathcal{L}_k(x) = -\frac{T_m(x)}{m}\left(\frac{R_m^{(1)} + R_m^{(3)}}{2} + x^2R_m^{(1)} - 2xR_m^{(2)}\right) = -\frac{\pi}{2m^2}T_m(x)(1 + x^2 - 2x) + o(m^{-2}).$$

Proof of Lemma 1. We first note that $R_m^{(1)}$, $R_m^{(2)}$ and $R_m^{(3)}$ can be rewritten as:

$$R_m^{(1)} = \frac{-\cos(\pi/2m)}{2}\sum_{k=1}^{m}\left[\sin\left(k\left(\pi/m + \pi\right)\right) + \sin\left(k\left(\pi/m - \pi\right)\right)\right] + \sin(\pi/2m)/2\sum_{k=1}^{m}\left[\cos\left(k\left(\pi/m - \pi\right)\right) + \cos\left(k\left(\pi/m + m\right)\right)\right].$$

$$R_m^{(2)} = \frac{-1}{4}\cos(\pi/m)\sum_{k=1}^{m}\sin(k(2\pi/m + \pi)) - \frac{1}{4}\cos(\pi/m)\sum_{k=1}^{m}\sin(k(2\pi/m - \pi)) + \frac{1}{4}\sin(\pi/m)\sum_{k=1}^{m}\cos(k(2\pi/m - \pi)) + \frac{1}{4}\sin(\pi/m)\sum_{k=1}^{m}\cos(k(2\pi/m + \pi)).$$

$$R_m^{(3)} = -\frac{1}{4}\sum_{k=1}^{m}\sin(3\theta_k + k\pi) + \sin(3\theta_k - k\pi) - \frac{1}{4}\sum_{k=1}^{m}\sin(-\theta_k + k\pi) + \sin(-\theta_k - k\pi).$$

Using for $t \in \mathbb{R}$, $\sum_{k=1}^{m}\cos(kt) = \frac{\cos(mt/2)\sin((m+1)t/2)}{\sin(t/2)}$ and $\sum_{k=1}^{m}\sin(kt) = \frac{\sin(mt/2)\sin((m+1)t/2)}{\sin(t/2)}$, some classical computations provide

$$R_m^{(1)} = \sin(\pi/2m) = \frac{\pi}{2m} + o(m^{-1}),$$
$$R_m^{(2)} = \frac{1}{2}\sin(\pi/m) = \frac{\pi}{2m} + o(m^{-1}),$$
$$R_m^{(3)} = \frac{1}{2}(\sin(3\pi/2m) - \sin(\pi/2m)) = \frac{\pi}{2m} + o(m^{-1}).$$

□

Proof of Lemma 2. First, we have (see [8] page 10)

$$\mathcal{L}_k(x) = \frac{T_m(x)}{T'_m(x_k)(x - x_k)} \tag{9}$$

and $-\sin\theta_k T'_m(\cos\theta_k) = -m\sin(m\theta_k)$. It follows that

$$\sum_{k=1}^{m}(x_k - x)\mathcal{L}_k(x) = -\frac{T_m(x)}{m}R_m^{(1)},$$
$$\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x) = \frac{xT_m(x)}{m}R_m^{(1)} - \frac{T_m(x)}{m}R_m^{(2)},$$
$$\sum_{k=1}^{m}(x_k - x)^3\mathcal{L}_k(x)] = \frac{-T_m(x)}{2m}\left(R_m^{(1)} + R_m^{(3)}\right) - \frac{T_m(x)x^2}{m}R_m^{(1)} + 2\frac{T_m(x)}{m}R_m^{(2)}.$$

□

A.1 Proof of Proposition 3.1

Clearly, we have

$$\mathbb{E}(\widetilde{F}_{n,m}(x)) = A_m(x).$$

The expansion of Taylor-Young ensures that for $1 \leq k \leq m$,

$$\begin{aligned}\mathbb{E}(\widetilde{F}_{n,m}(x)) &= F(x) + f(x)\sum_{k=1}^{m}(x_k - x)\mathcal{L}_k(x) \\ &+ \frac{f'(x)}{2}\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x) \\ &+ \frac{f''(x)}{6}\sum_{k=1}^{m}(x_k - x)^3\mathcal{L}_k(x) \\ &+ o\left(\sum_{k=1}^{m}(x - x_k)^3\mathcal{L}_k(x)\right).\end{aligned}$$

The application of Lemma A.1 together with Lemma A.2 yield the equation (4). Let's now focus on calculating the variance of our estimator. First, we set $(\eta_i)_{1\leq i\leq n} = \left(\sum_{k=1}^{m}\left(\mathbb{1}_{\{X_i\leq x_k\}} - F(x_k)\right)\mathcal{L}_k(x)\right)_{1\leq i\leq n}$, it comes that

$$\begin{aligned}&\widetilde{F}_{n,m}(x) - A_m(x) \\ &= \sum_{k=1}^{m}\left(\widehat{F}_n(x_k) - F(x_k)\right)\mathcal{L}_k(x) \\ &= \frac{1}{n}\sum_{k=1}^{m}\left(\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leq x_k\}} - F(x_k)\right)\mathcal{L}_k(x) \\ &= \frac{1}{n}\sum_{i=1}^{n}\eta_i.\end{aligned}$$

Moreover, since $\mathbb{E}\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) = 0$, it follows that

$$\begin{aligned}Var\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) &= \mathbb{E}\left[\left(\widetilde{F}_{n,m}(x) - A_m(x)\right)^2\right] \\ &= Var\left(\widetilde{F}_{n,m}(x)\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}Var(\eta_i) \\ &= \frac{1}{n}\mathbb{E}(\eta_1^2).\end{aligned}$$

Now, we define for any $x \in [-1,1]$ and for $i \geq 1$, $\varphi_i(x) = \mathbb{1}_{\{X_i\leq x\}} - F(x)$. We infer that

$$\begin{aligned}\mathbb{E}(\eta_i^2) &= \mathbb{E}\left[\left(\sum_{k=1}^{m}\varphi_i(x_k)\mathcal{L}_k(x)\right)^2\right] \\ &= \mathbb{E}\left[\sum_{k,l=1}^{m}\varphi_i(x_k)\mathcal{L}_k(x)\varphi_i(x_l)\mathcal{L}_l(x)\right] \\ &= \sum_{k,l=1}^{m}\mathbb{E}\left[\varphi_i(x_k)\varphi_i(x_l)\right]\mathcal{L}_k(x)\mathcal{L}_l(x). \qquad (10)\end{aligned}$$

Moreover, we have

$$\begin{aligned}&\mathbb{E}\left[\varphi_i(x_k)\varphi_i(x_l)\right] \\ &= \mathbb{E}\left[\left(\mathbb{1}_{\{X_i\leq x_k\}} - F(x_k)\right)\left(\mathbb{1}_{\{X_i\leq x_l\}} - F(x_l)\right)\right] \\ &= \mathbb{E}\left(\mathbb{1}_{\{X_i\leq x_k\}}\mathbb{1}_{\{X_i\leq x_l\}}\right) - F(x_k)F(x_l) \\ &= \mathbb{E}\left(\mathbb{1}_{\{X_i\leq \min(x_k,x_l)\}}\right) - F(x_k)F(x_l) \\ &= F(\min(x_k,x_l)) - F(x_k)F(x_l) \\ &= \min(F(x_k),F(x_l)) - F(x_k)F(x_l).\end{aligned}$$

Substituting this result for (10) leads to

$$\begin{aligned}&\mathbb{E}(\eta_i^2) \\ &= \sum_{k,l=1}^{m}\left[\min(F(x_k),F(x_l)) - F(x_k)F(x_l)\right]\mathcal{L}_k(x)\mathcal{L}_l(x) \\ &= \sum_{k=1}^{m}F(x_k)\mathcal{L}_k^2(x) + 2\sum_{\substack{k=1\\k<l}}^{m}F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x) \\ &-A_m(x)^2. \qquad (11)\end{aligned}$$

We need now to find an asymptotic expression for (11). For this reason, we first expand $F(x_k)$ about x to state that for all $0 \leq k \leq m$, $F(x_k) = F(x) + O(|x_k - x|)$. This allows us to write the first term of (11) as

$$\begin{aligned}&\sum_{k=1}^{m}F(x_k)\mathcal{L}_k^2(x) \\ &= \sum_{k=1}^{m}\left[F(x) + O(|x_k - x|)\right]\mathcal{L}_k^2(x) \\ &= \sum_{k=1}^{m}F(x)\mathcal{L}_k^2(x) + \sum_{k=1}^{m}O(|x_k - x|\,\mathcal{L}_k^2(x)) \\ &= F(x)S_m(x) + O(J_m(x)),\end{aligned}$$

where $J_m(x) = \sum_{k=1}^{m} \mid x_k - x \mid \mathcal{L}_k^2(x)$.

For the second term of (11), we instead write $F(x_k)$ as

$$F(x_k) = F(x) + (x_k - x)f(x) + O((x_k - x)^2).$$

Moreover, we have

$$2P_{0,m}(x) + S_m(x) = \sum_{k,l=1}^{m}\mathcal{L}_k(x)\mathcal{L}_l(x).$$

Since

$$\sum_{k,l=1}^{m}\mathcal{L}_k(x)\mathcal{L}_l(x) = 1,$$

it comes that

$$P_{0,m}(x) = \frac{1}{2}(1 - S_m(x)).$$

Then

$$\begin{aligned}
&\sum_{\substack{k=1\\k<l}}^{m} F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x)\\
&= \sum_{\substack{k=1\\k<l}}^{m} \left(F(x) + (x_k - x)f(x) + O((x_k - x)^2)\right)\mathcal{L}_k(x)\mathcal{L}_l(x)\\
&= \sum_{\substack{k=1\\k<l}}^{m} F(x)\mathcal{L}_k(x)\mathcal{L}_l(x) + \sum_{\substack{k=1\\k<l}}^{m}(x_k - x)f(x)\mathcal{L}_k(x)\mathcal{L}_l(x)\\
&\quad + \sum_{\substack{k=1\\k<l}}^{m} O\left((x_k - x)^2\mathcal{L}_k(x)\mathcal{L}_l(x)\right)\\
&= F(x)P_{0,m}(x) + f(x)P_{1,m}(x) + O(P_{2,m}(x))\\
&= \frac{1}{2}F(x)(1 - S_m(x)) + f(x)P_{1,m}(x) + O(P_{2,m}(x)).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&2\sum_{\substack{k=1\\k<l}}^{m} F(x_k)\mathcal{L}_k(x)\mathcal{L}_l(x)\\
&\quad = F(x)(1 - S_m(x)) + 2f(x)P_{1,m}(x) + O(m^{-4}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}(\eta_i^2)\\
&= F(x) + 2f(x)P_{1,m}(x) + O(J_m(x)) + O(P_{2,m}(x))\\
&\quad - A_m^2(x)\\
&= F(x)(1 - F(x)) + 2f(x)P_{1,m}(x) + O(J_m(x))\\
&\quad + O(m^{-4})\\
&= \sigma^2(x) + 2f(x)P_{1,m}(x) + O(J_m(x)) + O(P_{2,m}(x)). \qquad (12)
\end{aligned}$$

Now, using Cauchy-Schwartz's inequality combined with the fact that $|\mathcal{L}_k(x)| \leq 1$, we get

$$\begin{aligned}
|J_m(x)| &= \left|\sum_{k=1}^{m} |x_k - x|\,\mathcal{L}_k^2(x)\right|\\
&\leq \left|\left(\sum_{k=1}^{m}(x_k - x)^2\mathcal{L}_k(x)\right)\right|^{1/2}\left|\left(\sum_{k=1}^{m}\mathcal{L}_k^3(x)\right)\right|^{1/2}\\
&\leq \left[\left(\frac{\pi}{m^2} + o\left(m^{-2}\right)\right)S_m(x)\right]^{1/2}.
\end{aligned}$$

On the other side, using the fact that $\Lambda_m \leq \frac{2}{\pi}\ln(m+1) + 1$ (see [4]), we obtain

$$\begin{aligned}
S_m(x) &\leq \left(\sum_{k=1}^{m} |\mathcal{L}_k(x)|\right)^2\\
&\leq \Lambda_m^2 \leq \frac{4}{\pi^2}\ln(m+1)^2 + 1 + \frac{4}{\pi}\ln(m+1).
\end{aligned}$$

As a matter of fact, we infer that $J_m(x) = O(m^{-1/2})$. Now, it follows from (9), that

$$\begin{aligned}
P_{2,m}(x) &= [(1-x)^2 + O(m^{-2})]\sum_{\substack{k=1\\k<l}}^{m}\mathcal{L}_k(x)\mathcal{L}_l(x)\\
&= [(1-x)^2 + O(m^{-2})]\sum_{\substack{k=1\\k<l}}^{m} O(m^{-2}).
\end{aligned}$$

It follows that $P_{2,m}(x) = O(m^{-1})$. Moreover, we have

$$\begin{aligned}
P_{1,m}(x) &= \sum_{k=1}^{m}(x_k - x)\mathcal{L}_k(x)\sum_{l=k+1}^{m}\mathcal{L}_l(x)\\
&= [1 - x + O(m^{-2})]\sum_{\substack{k=1\\k<l}}^{m} O(m^{-2}).
\end{aligned}$$

Hence, we obtain $P_{1,m}(x) = O(m^{-1})$ and equation (5) follows.

A.2 Proof of Proposition 3.2

We first use the fact that

$$\left\|\widetilde{F}_{n,m} - F\right\| \leq \left\|\widetilde{F}_{n,m} - A_m\right\| + \|A_m - F\|.$$

The use of Jackson's theorem, ensures that

$$\lim_{m\to\infty}\|A_m - F\| = 0.$$

Moreover, we have

$$\widetilde{F}_{n,m}(x) - A_m(x) = \sum_{k=1}^{m}\left(\widehat{F}_n(x_k) - F(x_k)\right)\mathcal{L}_k(x),$$

it comes that

$$\left\|\widetilde{F}_{n,m} - A_m\right\| \leq \max_{1\leq k\leq m}\left|\widehat{F}_n(x_k) - F(x_k)\right|.$$

In addition, the application of Clivenco-Cantelli's theorem, ensures that

$$\lim_{n\to\infty}\left\|\widehat{F}_n - F\right\| = 0,$$

which conclude the proof.

A.3 Proof of Proposition 3.3

First, we note that for all $m \geq 1$,

$$\left\|\widetilde{F}_{n,m} - F\right\| \leq \left\|\widehat{F}_n - F\right\| + \|A_m - F\|.$$

Moreover, as $F \in Lip(\alpha, c)$, section 1.3.2 of De Dyn Nira et al. [7] and Jackson [10, 11] implies that

$$\|A_m - F\| = O\left(\frac{\log(m)}{m^{\alpha/2}}\right).$$

It follows that

$$\begin{aligned} &\limsup_{n\to\infty} u_n \left\|\widetilde{F}_{n,m} - F\right\| \\ &\quad\le \limsup_{n\to\infty} u_n \left\|\widehat{F}_n - F\right\| + \limsup_{n\to\infty} u_n \|A_m - F\|, \end{aligned}$$

where $u_n = (2n/\log\log n)^{1/2}$, for all $n \ge 1$. Now, using equation (7), we obtain

$$\limsup_{n\to\infty} u_n \left\|\widehat{F}_n - F\right\| = 1 \text{ a.s.}$$

Moreover, since $n^{1/2}m^{-\alpha/2} \to 0$ when $n,\ m \to \infty$, we have

$$\begin{aligned} \limsup_{n\to\infty} u_n \|A_m - F\| &= \limsup_{n\to\infty} \frac{(2n)^{1/2}}{(\log\log n)^{1/2}} \frac{\log m}{m^{\alpha/2}} \\ &= \limsup_{n\to\infty} \frac{\sqrt{n}}{m^{\alpha/2}} \\ &= 0. \end{aligned}$$

It comes that, $\limsup_{n\to\infty} u_n \left\|\widetilde{F}_{n,m} - F\right\| \le 1$. This completes the proof of proposition 3.3.

A.4 Proof of Proposition 3.4

Since we have

$$\widetilde{F}_{n,m}(x) - A_m(x) = \frac{1}{n}\sum_{i=1}^n \eta_i.$$

It follows that,

$$n^{1/2}\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) = \sum_{i=1}^n \frac{\eta_i}{n^{1/2}}.$$

Now, in order to check the Lindeberg condition, we notice for all $n \ge 1$ and for $i = 1, \dots, n$

$$X_{i,n} = \frac{\eta_i}{n^{1/2}} \quad \text{and} \quad s_n^2 = \sum_{i=1}^n \mathbb{E}(X_{i,n}^2).$$

We have

$$n^{1/2}\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) = \sum_{i=1}^n X_{i,n},$$

with $(X_{i,n})_{i\ge1}$ is a sequence of i.i.d. random variables such that $\mathbb{E}(X_{i,n}) = 0$. Further, we have for $n \ge 1$,

$$s_n^2 = \sum_{i=1}^n \mathbb{E}(X_{i,n}^2) = \sum_{i=1}^n \frac{1}{n}\mathbb{E}(\eta_i^2) = \mathbb{E}(\eta_1^2).$$

However, in the light of (12), we have $\lim_{n\to\infty} s_n^2 = \sigma^2(x)$. Indeed, using the Cauchy-Schwarz inequality, $\mathcal{L}_k(x) \le 1$ and by inferring the proof of proposition 3.1, we get $\lim_{n\to\infty} J_m(x) = 0$ and $\lim_{n\to\infty} P_{1,m}(x) = 0$. Moreover, since $\sum_{k=0}^m \mathcal{L}_k(x) = 1$, we have

$$\begin{aligned} |\eta_1| &= \left|\sum_{k=1}^m \left(\mathbb{1}_{\{X_i \le x_k\}} - F(x_k)\right)\mathcal{L}_k(x)\right| \\ &\le \sum_{k=1}^m \left|\mathbb{1}_{\{X_i \le x_k\}} - F(x_k)\right| \mathcal{L}_k(x) \\ &\le \sum_{k=1}^m (1+1)\mathcal{L}_k(x) = 2. \end{aligned}$$

It comes that

$$\begin{aligned} X_{1,n}^2 \mathbb{1}_{\frac{|X_{1,n}|}{s_n} > \varepsilon} &= \frac{\eta_1^2}{n}\mathbb{1}_{\{|\eta_1| > s_n n^{1/2}\varepsilon\}} \\ &\le \frac{4}{n}\mathbb{1}_{\{|\eta_1| > s_n n^{1/2}\varepsilon\}}. \end{aligned}$$

Hence,

$$\sum_{i=1}^n \mathbb{E}\left[X_{i,m}^2 \mathbb{1}_{\frac{|X_{i,m}|}{s_n} > \varepsilon}\right] \le \frac{4}{n}\sum_{i=1}^n \mathbb{1}_{\{|\eta_i| > s_n n^{1/2}\varepsilon\}}.$$

Moreover, we have $s_n^2 \to \sigma^2(x)$ when $n \to \infty$, then Lindeberg's condition

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}\left[X_{i,n}^2 \mathbb{1}_{\frac{|X_{i,n}|}{s_n} > \varepsilon}\right] = 0,$$

is fulfilled. Thus by Lindeberg-Feller's central limit theorem, we get

$$n^{1/2}\left(\widetilde{F}_{n,m}(x) - A_m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)),$$

which concludes the proof.

ACKNOWLEDGMENTS

This work was performed while the first author was at the Laboratory of Mathematics and Applications of Poitiers as a visiting scientist, partly funded by the University of Sfax. The authors would like to thank the Editor-in-Chief Prof. Yuedong Wang of Journal of Statistics and Its Interface and the reviewer for their helpful comments, which helped us to focus on improving the original version of the paper.

FUNDING

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

Received 7 November 2019

REFERENCES

[1] Austin, A. P. (2016). Some New Results On, and Applications Of, Interpolation in Numerical Computation. *Doctoral dissertation, University of Oxford*.

[2] Babu, G. J., Canty, A. J., and Chaubey, Y. P. (2002). Application of Bernstein Polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference* **105** 377–392. MR1910059

[3] Babu, G. J. and Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on a hyper cube using Bernstein polynomials for dependent random vectors. *Statist. Probab. Lett.* **76** 959–969. MR2270097

[4] Brutman, L. (1978). On the Lebesgue function for polynomial interpolation. *SIAM J. Numer. Anal.* **15** 694–704. MR0510554

[5] Cheney, E. W. (1966). Introduction to approximation theory. *Mc Graw-Hill, New York*. MR0222517

[6] Chung, K. L. (1949). An estimate concerning the kolmogroff limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50. MR0034552

[7] Dyn, N., Farkhi, E., and Mokhov, A. (2014). Approximation of set-valued functions: Adaptation of classical approximation operators. MR3309445

[8] Feldheim, E. (1939). Théorie de la convergence des procédés d'interpolation et de quadrature mécanique. *Mem. Sci. math.*

[9] Galambos, J. and Seneta, E. (1973). Regularly Varying Sequences. *Proc. Amer. Math. Soc.* **41** 110–116. MR0323963

[10] Jackson, D. (1911). Über die Genauigkeit der Annäherung stetiger Funktionen durch ganze rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung. Dieterich.

[11] Jackson, D. (1912). On approximation by trigonometric sums and polynomials. *Trans. Amer. Math. Soc.* **13** 491–515. MR1500930

[12] Jmaei, A., Slaoui, Y., and Dellagi, W. (2017). Recursive distribution estimators defined by stochastic approximation method using Bernstein polynomials. *J. Nonparametr. Stat.* **29** 792–805. MR3740720

[13] Joyner, W. B. and Boore, D. M. (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the seismological Society of America* **71** 2011–2038.

[14] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *J. Nonparametr. Stat.* **16** 709–729. MR2068610

[15] Leblanc, A. (2009). Chung–Smirnov property for Bernstein estimators of distribution functions. *J. Nonparametr. Stat.* **21** 133–142. MR2488150

[16] Leblanc, A. (2010). A bias-reduced approach to density estimation using Bernstein polynomials. *J. Nonparametr. Stat.* **22** 459–475. MR2662607

[17] Leblanc, A. (2012). On estimationg distribution function using Bernstein polynomials. *Ann. Inst. Statist. Math.* **64** 919–943. MR2960952

[18] Mokkadem, A., Pelletier, M., and Slaoui, Y. (2009). The stochastic approximation method for the estimation of a multivariate probability density. *J. Statist. Plann. Inference* **139** 2459–2478. MR2508006

[19] Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory Probab. Appl.* **9** 497–500. MR0166862

[20] Quintela-del-Río, A. and Estévez-Pérez, G. (2012). Non parametric kernel distribution function estimation with kerdiest: an R package for bandwidth choice and applications. *J. Statist. Softw* **50** 1–21.

[21] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282

[22] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.* **27** 832–837. MR0079873

[23] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference* **35** 65–75. MR1220406

[24] Salzer, H. E. (1972). Lagrangian interpolation at the Chebyshev points X_n, $\nu \equiv \cos(\nu\pi/n)$, $\nu = o(1)n$; some unnoted advantages. *Comput. J* **15** 156–159. MR0315865

[25] Serfling, R. J. (2009). Approximation theorems of mathematical statistics (Vol. 162). *John Wiley and Sons* MR0595165

[26] Singh, R. S., Gasser, T., and Prasad, B. (1983). Nonparametric estimates of distribution functions. *Comm. Statist. Theory Methods* **12** 2095–2108. MR0710562

[27] Slaoui, Y. (2014). The stochastic approximation method for the estimation of a distribution function. *Math. Methods Statist.* **23** 306–325. MR3295061

[28] Slaoui, Y. (2015). Plug-In Bandwidth selector for recursive kernel regression estimators defined by stochastic approximation method. *Stat. Neerl.* **69** 483–509. MR3414708

[29] Slaoui, Y. (2016). Optimal bandwidth selection for semi-recursive kernel regression estimators. *Stat. Interface* **9** 375–388. MR3457504

[30] Smirnov, N. V. E. (1944). Approximate laws of distribution of random variables from empirical data. *Uspekhi Mat. Nauk* **10** 179–206. MR0012387

[31] Tenbusch, A. (1994). Two-dimensional Bernstein polynomial density estimators. *Metrika* **41** 233–253. MR1293514

[32] Vitale, R. A. (1975). A Bernstein Polynomial Approach to Density Function Estimation. *Stat. Inference Relat. Topics* **2** 87–99. MR0397977

[33] Wang, F. K. (2000). A new model with bathtub-shaped failure rate using an additive Burr XII distribution. *Reliability Engineering and System Safety* **70** 305–312.

[34] Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis II. *Sankhya A* **26** 101–116. MR0184336

[35] Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.* **1** 247–253. MR0395026

[36] Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions. *J. Appl. Probab.* **16** 163–173. MR0520946

[37] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.* **15** 69–78. MR0329113

Salima Helali
Laboratoire de Probabilité et Statistiques
Sfax
Tunisie
E-mail address: helali.salima@gmail.com

Yousri Slaoui
Laboratoire de Mathématiques et Applications
Poitiers
France
E-mail address: Yousri.Slaoui@math.univ-poitiers.fr