TWMS J. App. and Eng. Math. V.14, N.2, 2024, pp. 775-786

ON KERNEL DISTRIBUTION FUNCTION ESTIMATION NEAR END POINTS

N. ALMI[1], A. SAYAH[1*], §

Abstract. Under that so-called boundary problem on kernel cumulative distribution function estimation, the main objective of this paper is to introduce and investigate the properties of our new estimator in order to improve the boundary effects, we will restrict our attention to the right boundary. We turned out that the order of the Bias has been reduced to the second power of the bandwidth, simulation studies for three different types of bandwidth selecting methods were carried out to check these phenomena, we conclude that the proposed estimator is better than the existing boundary correction methods.

Keywords: Boundary effects, Bias reduction, Cumulative distribution function, Kernel estimator.

AMS Subject Classification: 62G05, 62G20.

1. Introduction

Let $X_1, ..., X_n$ be a random variables independent and identically distributed with density f having support $[0,1]$. The most commonly used non parametric estimator of the cumulative distribution function (CDF) F is the empirical distribution function (EDF) F_n, defined at some point x as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,x]}(X_i), \tag{1}$$

where the indicator function $\mathbb{1}_{]-\infty,x]}(X_i) = 1$ if $X_i \leq x$ and 0 otherwise.

Theoretical properties of $F_n(.)$ have been investigated by several authors among them, ([22], [14], [3]). It is well known that $F_n(.)$ is a step function even in case F is continuous and even when n is large, $F_n(.)$ is less smoothing, this fact leads to the effort to obtain an alternative estimator more smoothing.

[1] Department of Mathematics and Informatic, LMA Laboratory, Mohamed Khider University, Biskra, Algeria.
e-mail: almi.nassima@gmail.com; ORCID: https://orcid.org/0000-0001-8313-0537.
e-mail: abdallah.sayah@univ-biskra.dz; ORCID: https://orcid.org/0000-0003-2417-522X.
[*] Corresponding author.

§ Manuscript received: March 29, 2022; accepted: May 30, 2022.

Rosenblatt [15], Parsen [12] introduced the kernel density estimator of $f(.)$ defined at x by:

$$f_C\left(x\right)=\frac{1}{nh}\sum_{i=1}^{n}k\left(\frac{x-X_i}{h}\right),$$

then Nadaraya [11] proposed a classical kernel estimator of function F arises as an integral of kernel density estimator defined by:

$$F_C\left(x\right)=\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right), \tag{2}$$

where $h=h_n$ is the smoothing parameter (or the bandwidth) since it controls the amount of smoothness in the estimator, and satisfy $h:=h_n\rightarrow 0$ as $n\rightarrow\infty$, $k\left(.\right)$ is a kernel function which is a predetermined density function symmetric about 0.

The function K is defined from a kernel k as

$$K\left(x\right)=\int_{-\infty}^{x}k\left(t\right)dt.$$

If the support of f is a compact interval, from the continuity of F it is well known that the kernel estimator (2) is an asymptotically unbiased estimator of F if and only if $h\rightarrow 0$ as n goes to infinity (see Yamato [22], Lemma 1). However, if F is not smooth enough at the boundary points of support the previous kernel estimator F_C is a biased estimator near the boundary of its support, due to so-called boundary effects. In fact, the value of Bias and Variance of F_C at interior point provided by Azzalini [2] are respectively :
for $x\in[h,1-h]$

$$Bias(F_C(x))=\frac{1}{2}f^{(1)}(x)\mu_2(k)h^2+o(h^2), \tag{3}$$

and

$$Var(F_C(x))=\frac{F(x)(1-F(x))}{n}+\frac{h}{n}f(x)\left(\int_{-1}^{1}K^2(t)dt-1\right)+o\left(\frac{h}{n}\right), \tag{4}$$

where $\mu_2(k)=\int t^2k(t)dt$ and $f^{(1)}$ denote the first derivative of f.
While, for x in the right boundary $]1-h,1]$, we can write $x=1-ch$ where $0\leq c<1$, then the Bias and Variance of F_C at x are respectively:

$$-hf(1)\int_{-1}^{-c}K(t)dt+h^2f^{(1)}(1)\left(\frac{c^2}{2}-\int_{-1}^{c}tK(t)dt+c\int_{-1}^{-c}K(t)dt\right)+o(h^2), \tag{5}$$

and

$$\frac{F(x)(1-F(x))}{n}+\frac{h}{n}f(1)\left(-c-2\int_{-1}^{-c}K(t)dt+\int_{-1}^{c}K^2(t)dt\right)+o\left(\frac{h}{n}\right). \tag{6}$$

Proofs of (5) and (6) are provided in Almi & al [1].

In order to improve the theoretical performance of the classical kernel distribution function estimator when the underlying distribution function F is not smooth enough at the extreme points of the distribution support, several methods have been proposed for kernel estimation in regression and density function estimation, among them, reflection of data [16], pseudo-data method [4] and also the boundary kernel method [6]. However, the boundary problem in kernel distribution estimation is less severe, some recent references

Koláček & al [9] considered the boundary problem in distribution function estimation in estimating ROC curves using the transformation method. Tenreiro[17] proposed a boundary kernel method for correcting the boundary. Tenreiro [18] and Zhang & al [23] introduced a new class of boundary kernels for distribution function estimation problem. Almi & al [1] proposed two estimators the first estimator based a self elimination between modify theoretical Bias term and the classical kernel estimator itself and the second estimator is kind of a generalized reflection method involving reflecting a transformation of the observed data.

The main subject of this paper is to propose a new estimator for kernel distribution function to improve the right boundary effects. The organization of the rest of this article is as follows. In the next section, we provide an explicit formulation and theoretical properties of the proposed estimator. In section 3, we support the theoretical results with Monte Carlo simulations to investigate the performance of the estimators in terms of Bias and MSE at the first, then comparing the graphical presentation of the behavior of the estimators. In addition, an application on real data as an example, when the data near the right boundary. Some concluding remarks are given in Section 4.

2. Main results

In order to derive our main results, we shall need the following assumptions.

- A_1 : F is twice continuously differentiable in a neighborhood of x.
- A_2 : The kernel k is a probability density, nonnegative, bounded, symmetric, and has compact support $[-1, 1]$.
- A_3 : $f(1) \neq 0$ and $h < \dfrac{1}{2}$.

If x is a point in the right boundary, we can write $x = 1 - ch$ where $c \in [0, 1[$, our proposed estimator has the form

$$F_G(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) + \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - 2 + g(X_i)}{h}\right), \tag{7}$$

where g is a transformation which is selected from a parametric family, we assume that verify:

- H_1 : g is a continuous and monotonically increasing function from $[0, 1]$ to $[0, 1]$.
- H_2 : g^{-1} exist and verify $g^{-1}(1) = 1$ and $g^{(1)}(1) = 1$ where g^{-1} and $g^{(1)}$ denoting respectively the inverse and the first derivative function of g.

It is clear that there are various possible choices available for the function g that satisfy the above assumptions. Based on extensive simulations, we choose the following transformation g which well adapts to various shapes of distributions and improve the Bias

$$g(t) = t - t(1-t)^2 \int_{-1}^{-c} K(t)dt, \quad c \in [0, 1[\,.$$

The following theorem provided the asymptotic properties of our estimator, it showing that the Bias reduced to order $o(h^2)$ while the Variance is of order $o\left(\dfrac{h}{n}\right)$.

Theorem 2.1 (Asymptotic properties)**.** *Assume that the above assumptions A_1, A_2, A_3, H_1 and H_2 hold, then the Bias and Variance of $F_G(.)$ at $x = 1 - ch$ respectively are*

$$h^2\left(\frac{c^2}{2}f^{(1)}(1)+\int_{-1}^{-c}K(t)\left(2cf^{(1)}(1)-f(1)g^{(2)}(1)(t+c)\right)dt-f^{(1)}(1)\int_{-c}^{c}tK(t)\,dt\right)+o\left(h^2\right), \tag{8}$$

and

$$\frac{F(x)(1-F(x))}{n} \quad +\frac{h}{n}f(1)\left(-c+\int_{-1}^{c}K^2(t)dt+\int_{-1}^{-c}\left(K^2(t)-2K(t)\right)dt + 2\int_{-1}^{-c}K(t)K(-2c-t)dt\right)+o\left(\frac{h}{n}\right). \tag{9}$$

Proof. Under above Assumptions, for $x \in]1-h;1]$, we have

$$\begin{aligned}
E\left(F_G(x)\right) &= E\left(K\left(\frac{x-X_i}{h}\right)\right)+E\left(K\left(\frac{x-2+g\left(X_i\right)}{h}\right)\right)\\
&= \int_0^1 K\left(\frac{x-z}{h}\right)f(z)dz+\int_0^1 K\left(\frac{x-2+g(z)}{h}\right)f(z)dz\\
&= I_1+I_2.
\end{aligned}$$

We will calculate each term separately.

By using the change of the variable and the property $K(t) = 1 - K(-t)$ on the first term, we have

$$\begin{aligned}
I_1 &= \int_0^1 K\left(\frac{x-z}{h}\right)f(z)dz\\
&= h\int_{-c}^{\frac{1}{h}-c}K(t)f(x-th)dt\\
&= h\int_{c}^{\frac{1}{h}-c}K(t)f(x-th)\,dt+h\int_{-c}^{c}K(t)f(x-th)\,dt\\
&= h\int_{c}^{\frac{1}{h}-c}(1-K(-t))f(x-th)\,dt+h\int_{-c}^{c}K(t)f(x-th)\,dt\\
&= h\int_{c}^{\frac{1}{h}-c}f(x-th)\,dt-h\int_{c}^{\frac{1}{h}-c}K(-t)f(x-th)\,dt+h\int_{-c}^{c}K(t)f(x-th)\,dt\\
&= F(1-2ch)-h\int_{\frac{-1}{h}+c}^{-c}K(t)f(x+th)\,dt+h\int_{-c}^{c}K(t)f(x-th)\,dt.
\end{aligned}$$

Depending on a Taylor expansion of the function $f(.)$ and $F(.)$ and at $x = 1$, we have

$$f(x+th) = f(1)-h(c-t)f^{(1)}(1)+o(h),$$

$$f(x-th) = f(1) - h(c+t)f^{(1)}(1) + o(h),$$

and

$$F\left(1-2ch\right) = F\left(1\right) - 2chf\left(1\right) + \frac{\left(2ch\right)^2}{2} f^{(1)}\left(1\right) + o\left(h^2\right). \tag{10}$$

By the existence and continuity of $F^{(2)}(.)$ near 1, we obtain for $x = 1 - ch$

$$F\left(1\right) = F\left(x\right) + chf\left(x\right) + \frac{\left(ch\right)^2}{2} f^{(1)}\left(x\right) + o\left(h^2\right),$$
$$f\left(x\right) = f\left(1\right) - chf^{(1)}\left(1\right) + o\left(h\right),$$
$$f^{(1)}\left(x\right) = f^{(1)}(1) + o\left(1\right).$$

By substitution in (10), we find

$$F\left(1-2ch\right) = F\left(x\right) - chf\left(1\right) + \frac{3}{2}\left(ch\right)^2 f^{(1)}\left(1\right) + o(h^2). \tag{11}$$

Therefore

$$I_1 = F(x) - hf(1)\int_{-1}^{-c} K(t)dt + h^2 f^{(1)}(1)\left(\frac{3}{2}c^2 + \int_{-1}^{c}(c-t)K(t)dt - \int_{-c}^{c}(c+t)K(t)dt\right) + o(h^2).$$

By the same procedure, we calculate the second term I_2

$$\begin{aligned} I_2 &= \int_0^1 K\left(\frac{x-2+g\left(z\right)}{h}\right) f(z)dz. \\ &= h\int_{-\frac{1}{h}-c}^{-c} K\left(t\right)\frac{f(g^{-1}\left(1+h\left(t+c\right)\right))}{g^{(1)}\left(g^{-1}\left(1+h\left(t+c\right)\right)\right)}dt, \end{aligned}$$

we use a Taylor expansion of the function $\dfrac{f(g^{-1}(.))}{g^{(1)}(g^{-1}(.))}$, we obtain

$$I_2 = hf(1)\int_{-1}^{-c} K(t)dt + h^2\int_{-1}^{-c} K(t)\left((c+t)(f^{(1)}(1) - f(1)g^{(2)}(1)\right)dt + o(h^2).$$

At last, we combine I_1 and I_2 terms to get the Bias of F_G .

For the Variance term, we have

$$\begin{aligned} nVar(F_G\left(x\right)) &= Var\left(K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x-2+g(X_i)}{h}\right)\right) \\ &= E\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x-2+g(X_i)}{h}\right)\right]^2 \\ &- \left[E\left\{K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x-2+g(X_i)}{h}\right)\right\}\right]^2 \\ &= J_1 - J_2, \end{aligned}$$

where

$$
\begin{aligned}
J_1 &= E\left[K\left(\frac{x-X_i}{h}\right)+K\left(\frac{x-2+g(X_i)}{h}\right)\right]^2 \\
&= E\left[K\left(\frac{x-X_i}{h}\right)\right]^2+E\left[K\left(\frac{x-2+g(X_i)}{h}\right)\right]^2 \\
&+ 2E\left[K\left(\frac{x-X_i}{h}\right)\times K\left(\frac{x-2+g(X_i)}{h}\right)\right], \\
&= \int_0^1 K^2\left(\frac{x-z}{h}\right)f(z)dz+\int_0^1 K^2\left(\frac{x-2+g(z)}{h}\right)f(z)dz \\
&+ 2\int_0^1 K\left(\frac{x-z}{h}\right)K\left(\frac{x-2+g(z)}{h}\right)f(z)dz \\
&= J_{11}+J_{12}+J_{13}.
\end{aligned}
$$

We will calculate each term separately.

$$
\begin{aligned}
J_{11} &= \int_0^1 K^2\left(\frac{x-z}{h}\right)f(z)dz \\
&= h\int_{-c}^{\frac{1}{h}-c} K^2(t)\,f(x-th)\,dt \\
&= h\int_{-c}^{c} K^2(t)\,f(x-th)\,dt+h\int_{c}^{\frac{1}{h}-c} K^2(t)\,f(x-th)\,dt \\
&= h\int_{-c}^{c} K^2(t)\,f(x-th)\,dt+h\int_{c}^{\frac{1}{h}-c} (1-K(-t))^2\,f(x-th)\,dt, \\
&= F(1-2ch)+h\int_{\frac{-1}{h}+c}^{-c} K^2(t)\,f(x+th)\,dt-2h\int_{\frac{-1}{h}+c}^{-c} K(t)\,f(x+th)\,dt \\
&+ h\int_{-c}^{c} K^2(t)\,f(x-th)\,dt.
\end{aligned}
$$

Depending on a Taylor expansion (11) we have

$$
J_{11}=F(x)-chf(1)+hf(1)\int_{-1}^{c} K^2(t)\,dt-2hf(1)\int_{-1}^{-c} K(t)\,dt+o(h).
$$

Similar computation give

$$\begin{aligned}
J_{12} &= \int_0^1 K^2\left(\frac{x-2+g(z)}{h}\right) f(z)dz \\
&= h \int_{-\frac{1}{h}-c}^{-c} K^2(t) \frac{f(g^{-1}(1+h(t+c)))}{g^{(1)}(g^{-1}(1+h(t+c)))} dt \\
&= hf(1) \int_{-1}^{-c} K^2(t)\, dt + o(h).
\end{aligned}$$

$$\begin{aligned}
J_{13} &= \int_0^1 K\left(\frac{x-z}{h}\right) K\left(\frac{x-2+g(z)}{h}\right) f(z)dz \\
&= \int_{-1}^{-c} K(-2c-t)\, K(t) \frac{f(g^{-1}(1+h(c+t)))}{g^{(1)}(g^{-1}(1+ch+th))} dt \\
&= 2hf(1) \int_{-1}^{-c} K(t)\, K(-2c-t)\, dt.
\end{aligned}$$

We combine J_{11}, J_{12} and J_{13} to obtain J_1.

With the expression of the $Bias(F_G)$, we find

$$J_2 = F^2(x) + o(h).$$

□

Corollary 2.1. *Under the assumptions of Theorem (2.1), the asymptotically optimal bandwidth in the sense of minimising the leading terms in the expansion of the MSE is*

$$\left(\frac{f(1)\left(c - \int_{-1}^{c} K^2(t)\, dt - \int_{-1}^{-c} K^2(t)\, dt + 2\int_{-1}^{-c} K(t)\, dt - 2\int_{-c}^{1} K(t)\, K(-2c-t))dt \right)}{4n\left(\frac{1}{2}c^2 f^{(1)}(1) + \int_{-1}^{-c} K(t)\left(2cf^{(1)}(1) - f(1)\, g^{(2)}(1)\,(t+c)\right) dt - f^{(1)}(1)\int_{-c}^{c} tK(t)\, dt \right)^2} \right)^{1/3}$$

3. Simulation study

In this section, we compare the performance of our proposed estimator with the classical kernel estimator and three existing estimators based on boundary modified kernel distribution function method summarized in the coming subsection. It is well known that the choice of the kernel function k is less important than the choice of the bandwidth h, in this section we use three different bandwidth selection methods, the first one consist on the plug on approach of Polansky & al denoted by h_{Bp}, the second one consist on the use of the Altman & al denoted by h_{Al}, the third one consists on the use of Cross-validation approach of Bowman denoted by h_{Cv} for Epanechnikov kernel.

The comparison was made by calculating the Bias and MSE through generating a sample size of $n = 200$ from distribution with support $[0, 1]$ listed in table (1) and we did thousand replication for each estimator. Let $\hat{\theta}_i$ be estimator of θ based on the ith generated random numbers of size n. Then the Monte Carlo estimator of the Bias and MSE are

$$Bias(\hat{\theta}) = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\theta}_i(x) - \theta(x)\right), \tag{12}$$

$$MSE(\hat{\theta}) = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\theta}_i(x) - \theta(x)\right)^2. \tag{13}$$

TABLE 1. Distributions used in the simulation study

	Description	Density for $x \in [0,1]$
D_1	Truncated Normal(0,1)	$exp(-x^2/2)/\int_0^1 exp(-x^2/2)dx$
D_2	Truncated Exponential(2)	$2exp(-2x)/(1-exp(-2))$
D_3	Kumaraswamy(4,2)	$8x^3(1-x^4)$
D_4	Beta$(3,1)$	$3x^2$

3.1. **Existing estimators used in comparison.** In this subsection, we mentioned existing boundary modified kernel distribution function estimator given by

$$F_{B_j}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{B_j}\left(\frac{x - X_i}{h}\right), \qquad j = 1, 2, 3.$$

where K_{B_j} is a kernel distribution function.

Tenreiro [18] and Zhang & al [23] found that k_{B_j} must satisfying

$$\int_{-c}^{1} \frac{c+x}{c} k_{B_j}(x)dx = 1,$$

we choice three boundary kernel which are

$$k_{B_1}(t) = 6(1-t)(c+t)\frac{1}{(1+c)^3}\left(1 + 5\left(\frac{1-c}{1+c}\right)^2 - 10\frac{1-c}{(1+c)^2}t\right), \qquad -c \le t \le 1.$$

$$k_{B_2}(t) = 12\frac{1-t}{(1+c)^4}\left(\frac{3c^2-2c+1}{2} - t(1-2c)\right), \qquad -c \le t \le 1.$$

$$k_{B_3}(t) = \frac{1}{(1+c)^3}(3(1+c^2) - 6t^2), \qquad -c \le t \le 1.$$

Each boundary kernel gives a different estimator, which we denote by F_{B_j} where $j = 1, 2, 3$.

Remark 3.1. *If $c = 1$, we have*

$$k_{B_1}(t) = k_{B_2}(t) = k_{B_3}(t) = \frac{3}{4}(1-t^2)\mathbb{1}_{[-1,1]}(t).$$

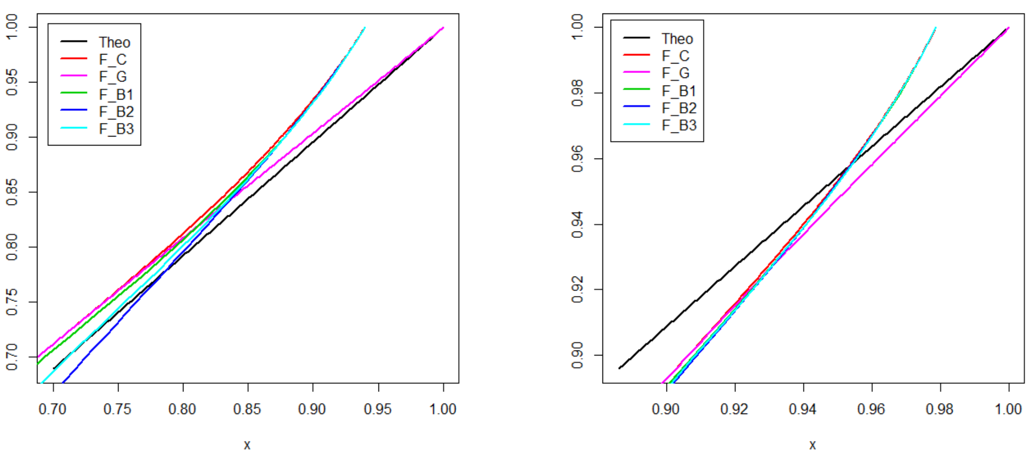
The simulation results measuring the performance of the different estimators for each distribution are summarized in the following table (2).

TABLE 2. Performance Bias(MSE) of distributions values at $x = 1$, Results are re-scaled by the factor 0.001.

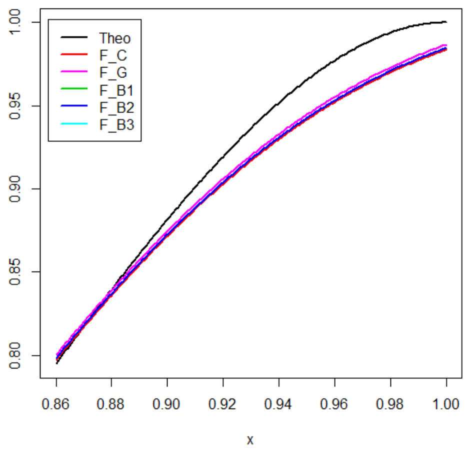
		F_C	F_{B_1}	F_{B_2}	F_{B_3}	F_G
D_1	hCv	6.452(6.919)	5.844 (6.867)	5.772 (6.866)	5.623 (6.636)	**3. 812 (4.034)**
	hAl	6.419(5.548)	6.533 (6.547)	6.436 (7.478)	6.224 (6.154)	4.234 (4.255)
	hBp	6.371(5.825)	5.971(6.725)	5.971(4.725)	5.971(4.725)	4.055(4.123)
D_2	hCv	2.106(2.170)	1.102(2.104)	1.101(2.124)	1.096(2.100)	**0.470(1.527)**
	hAl	2.938(2.963)	1.464(2.549)	1.457(2.108)	1.128(2.328)	0.826(1.937)
	hBp	2.671(2.225)	1.119(2.108)	1.131(2.106)	1.106(2.057)	0.529(1.734)
D_3	hCv	0.480(0.712)	0.4706(0.693)	0.470(0.693)	0.472(0.692)	**0.469(0.691)**
	hAl	0.707(0.478)	0.473(0.698)	0.472(0.696)	0.478(0.696)	0.472(0.695)
	hBp	0.480(0.712)	0.471(0.694)	0.470(0.693)	0.474(0.693)	0.470(0.692)
D_4	hCv	6.725(5.039)	5.325(4.687)	4.728(3.697)	4.283(3.423)	**3.754**(3.245)
	hAl	7.154(5.259)	6.325(4.921)	4.733(3.782)	4.326(3.457)	3.897(**3.012**)
	hBp	6.892(5.156)	6.012(4.698)	4.732(3.795)	4.301(3.424)	3.757(3.015)

Depending on the table, we can see that our proposed estimator is well performed when comparing with each considered estimators mentioned previously in the sens of Bias and MSE also, followed by boundary modified kernel estimators mentioned respectively F_{B_3}, F_{B_2} and F_{B_1}, even when we change the bandwidth value. the effect of the h_{Cv} method is more efficiency than h_{Bp} which better than h_{Al}. Also, we can see for density D_3 which takes the value zero at the right endpoints $f(1) = 0$ is free of boundary problem in such a case.

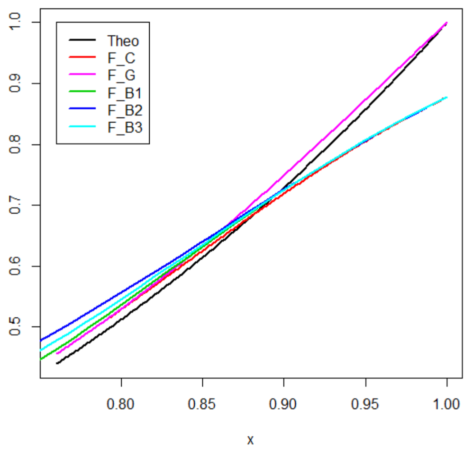
To support our numerical results, we present the graphical representation of the estimators when the data near the right boundary of the support



(a) Performance of truncated normal distribution (b) Performance of truncated exponential distribution



(c) Performance of Kumaraswamy distribution



(d) Performance of Beta distribution

FIGURE 1. Performance of considering estimators at the right boundary of the support

3.2. **Real Data Application.** The aim of our application is to compare the performance of the previously kernel estimators respectively F_n, F_C, F_G and F_{B_j} where $j = 1, 2, 3$, by using the cross-validation bandwidth selection method for two real data sets. The first data X consists of the number of deaths due to COVID-19 recorded from February 29, 2020 to December 31,2020 in 50 states of the United States of America which gives 305 observation taken from www.nytimes.com, where $X_i \in [0, 3808]$. The second data sets take of [5], which rises from the 1989 total charges in thousands of dollars for 33 patients at a Wisconsin hospital, is used. Each patient was female, aged 33–49, and admitted to the hospital for circulatory disorders where $X_i \in [2.337, 3.041]$. The support can be mapped onto the unit interval by the transformation $Z_i = (X_i - X_{(1)})/(X_n - X_{(1)})$, where $X_{(1)}$ and $X_{(n)}$ are repectively the minimum and the maximum of the data.

The results reveal that the proposed kernel estimator of the distribution function performs well when compared with the estimators previously mentioned.

4. Conclusions

In conclusion, the proposed estimator allowed us to reducing the Bias of kernel distribution function estimator and obtain better results when comparing with boundary modified kernel methods, it has a smaller Bias in the sense of convergence rate, specifically when the data near the right boundary, these results similar for each bandwidth selecting method. In generally, we can say that the cross-validation bandwidth reveals a very good performance fellowed by the plog in approach of Polansky and Baker. A simulation study achieves these results as well.

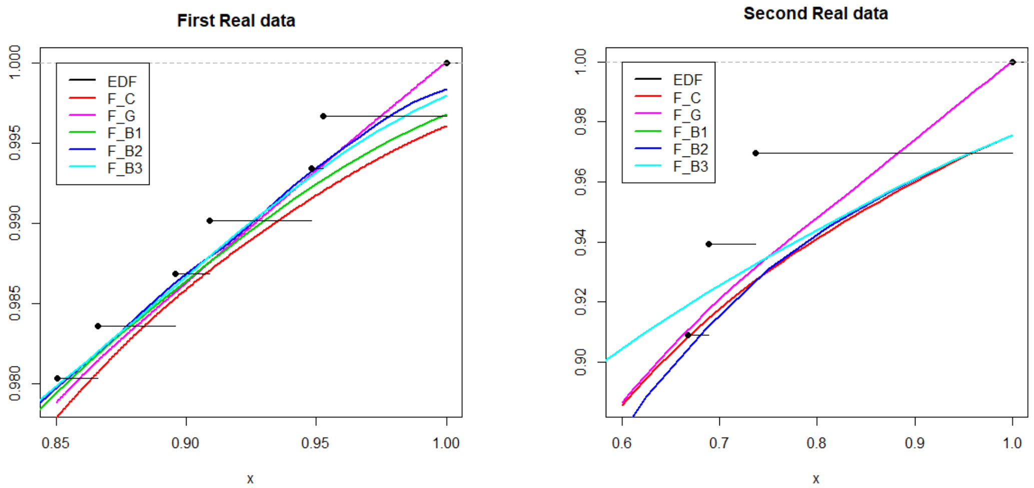


FIGURE 2. Performance of considering estimators at the right boundary of the support for real data sets

REFERENCES

[1] Almi, N., Sayah, A., (2021), nonparametric kernel distribution function estimation near endpoints, Advances in Mathematics: Scientific Journal, 10, pp. 3679-3697.

[2] Azzalini, A., (1981), A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method, Biometrika, 68, pp. 326-328.

[3] Altman, N., Leger, C., (1995), Bandwidth selection for kernel distribution function estimation , Journal of Statistical Planning and Inference, 46, pp. 195–214.

[4] Cowling, A., Hall, P., (1996), On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation, Journal of the Royal Statistical Society: Series B, 58, pp. 551-563.

[5] Frees, E., (1994), Estimating densities of functions of observations, Journal of the American Statistical Association, 89, pp. 517–525.

[6] Gasser, T., Muller, H., Mammitzsch, V., (1985), Kernels for Nonparametric Curve Estimation, Journal of the Royal Statistical Society. Series B (Methodological), pp. 238-252.

[7] Karunamuni, RJ., Alberts, T., (2005), A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation, Canadian Journal of Statistics, 33, pp. 497-509.

[8] Karunamuni, RJ., Zhang, S., (2008), Some Improvements on a Boundary Corrected Kernel Density Estimator, Statistics and Probability Letters, 78, pp. 497-507.

[9] Koláček, J., Karunamuni, RJ., (2009), On Boundary Correction in Kernel Estimation of ROC Curves, Austrian Journal of Statistics, 38, pp. 17-32.

[10] Kolácek, J., Karunamuni, RJ.,(2011), A Generalized Reflection Method for Kernel Distribution and Hazard Functions Estimation, Journal of Applied Probability and Statistics, 6, pp. 73-85.

[11] Nadaraya, EA., (1964), Some New Estimates for Distribution Functions, Theory of Probability and Its Applications, 9, pp. 497-500.

[12] Parzen, E.,(1962), On Estimation of a Probability Density Function and Mode, The annals of mathematical statistics, 33, pp. 1065-1076.

[13] Quintela-del-Río, A,. Estévez-Pérez, G., (2012), Nonparametric Kernel Distribution Function Estimation with kerdiest: an R Package For Bandwidth Choice and Applications, Journal of Statistical Software, 50, pp. 1-21 .

[14] Reiss, R-D., (1981), Nonparametric Estimation of Smooth Distribution Functions, Scandinavian Journal of Statistics, 8, pp. 116-119.

[15] Rosenblatt, M., (1956), Remarks on Some Nonparametric Estimates of a Density Function, The Annals of Mathematical Statistics, 27, pp. 832-837.

[16] Silverman, WR., (1986), Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.
[17] Tenreiro, C,. (2013), Boundary Kernels for Distribution Function Estimation, REVSTAT Statistical Journal, 11, pp. 169-190.
[18] Tenreiro, C,. (2017), A New Class of Boundary Kernels for Distribution Function Estimation, Communications in Statistics-Theory and Methods, 47, pp. 5319-5332.
[19] Tour, M., Sayah, A., Djebrane, Y., (2017), A Modified Champernowne Transformation to Improve Boundary Effect in Kernel Distribution Estimation, Afrika Statistika, 12, pp. 1219-1233.
[20] Watson, GS., Leadbetter, MR., (1964), Hazard Analysis II, Sankhyā: The Indian Journal of Statistics, Series A, 26, pp. 101-116.
[21] Winter, BB., (1979), Convergence Rate of Perturbed Empirical Distribution Functions, Journal of Applied Probability, 16, pp. 163-173.
[22] Yamato, H., (1973), Uniform Convergence of an Estimator of a Distribution Function, Bulletin of Mathematical Statistics, 15, pp. 69-78.
[23] Zhang, S., Zhong, L., Zhang, Z., (2020), Estimating a Distribution Function at the Boundary, Austrian Journal of Statistics, 49, pp. 1-23.

Nassima Almi is a PhD student in the Faculty of Mathematical Sciences, affiliated to the Applied Mathematics Laboratory of Mathematics Department, Mohamed Khider University, Biskra, Algeria. She has a master's degree in Mathematics, on the option of Statistic (2016) and a Bachelor's degree in applied mathematics (2014) from the same university. She is a statistician research scholar and her research interests include Nonparametric Estimation and Applied Mathematics.

Abdallah Sayah is a faculty member in the department of Mathematics and a member in the Laboratory of Applied Mathematics, Mohamed Khider University, Biskra, Algeria. He received his Diploma of Advanced Studies, Fundamental Mathematics and Applications from the University of Rennes 1 in France. He received his Msc in Applied Mahtematics from the University of Constantine 1, Constantine, Algeria and his Ph.D. in Applied Mathematics (Statistics and probability) from Mohamed Khider University, Biskra, Algeria. His research areas are in Nonparametric estimation, Kernel distribution, density, regression and quantile estimation.