ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **10** (2021), no.12, 3515–3531
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.10.12.1

BIAS CORRECTION AT END POINTS IN KERNEL DENSITY ESTIMATION

Hind Bouredji and Abdallah Sayah[1]

ABSTRACT. In this paper, we propose a new approach of boundary correction for kernel density estimation with the support $[0,1]$, in particular at the right end-points and we derive the theoretical properties of this new estimator and show that it asymptotically reduce the order of bias at the boundary region, whereas the order of variance remains unchanged. Our Monte Carlo simulations demonstrate the good finite sample performance of our proposed estimator. Two examples with real data are provided.

1. INTRODUCTION

Suppose we observe n independent identically distributed aleatoire random variables, with unknown continuous density function f. The kernel density estimator which presented by Rosenblatt [8] then developed by Parzen [7], is defined as,

$$f_{n,R}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}, \tag{1.1}$$

[1]*Abdallah Sayah*

2020 *Mathematics Subject Classification.* 62G07.
Key words and phrases. Kernel density estimation, Bias correction, Transformation, Reflection, Mean squared error, Mean integrated squared error.
Submitted: 11.11.2021; *Accepted:* 27.11.2021; *Published:* 01.12.2021.

where h is a positive smoothing parameter, called the bandwidth, in which $h \to 0$ and $nh \to \infty$ when $n \to \infty$, and k is the kernel function with compact support $[-1, 1]$, satisfying the following conditions,

$$k(t) \geq 0, k(t) = k(-t), \int_{-1}^{1} k(t)\, dt = 1, 0 \neq \int_{-1}^{1} t^2 k(t)\, dt < \infty. \tag{1.2}$$

Let introduce the notation,

$$\int_{-1}^{1} t^j k(t)\, dt = \mu_j, j = 1, 2, 3, \tag{1.3}$$

to be more precise $\mu_1 = \mu_3 = 0$ since k is symmetric. Best reference in this area is Silverman [**?**] and Wand and Jones [15]. With appropriate choice of h, we can divided the support of the density onto regions, the intervals $[0, h)$ and $(1-h, 1]$ are called the left and the right boundary region respectively and the interior region formed by the interval $[h, 1-h]$. The performance of the kernel density estimator at least in one side of the support $(x \in [0, h) \cup (1-h, 1])$, differs from the interior points due to so-called boundary problems and the region formed by the points with boundary problems is called the boundary region.

To remove these boundary effects at the left region $(x \in [0, h))$, a diversity of methods have been developed during the past two decades. Among them the reflection method (Schuster [17]), the transformation method (Marron and Ruppert [6]), the boundary kernel method (Jones [3]), the pseudo-data method (Cowling and Hall [2]), the local linear method (Zhang and Karunamuni [12]).

As the boundary kernel density estimator could yield negative point estimates, Jones and Foster [16] propose much simpler nonnegative boundary corrected estimators which are analogues of the wide class of simple. Karunamuni and Alberts [4] proposed a new general method generates a class of boundary corrected estimators possess desirable properties such as local adaptivity and non-negativity, in addition to this work, Karunamuni and Alberts [5] constructed a new technique based on a data transformation that depends on the point of estimation. In a very exciting work, Zhang and al [14] expected a new method of boundary correction for kernel density estimation, their approach is an amount of generalized reflection method involving reflecting a transformation of the data.

In this paper, we focus on the boundary bias problem in the right side of the support $(1-h,1]$, when the true density supported with endpoints one, the kernel density estimator has the well-known boundary problem. More specifically, we assume that $f^{(j)}$, the j^{th} derivative of f, exists and is continuous on a neighborhood of x, $(j=0,1,2,3)$, with $f^{(0)}=f$, then for $x=1-ch, c\in[0,1[$,

$$\begin{aligned}
E\left(f_{n,R}\left(x\right)\right) &= \int_{-c}^{1} k\left(t\right) f(x-th)dt \\
&= f(x)\int_{-c}^{1} k\left(t\right)dt - hf^{(1)}(x)\int_{-c}^{1} tk\left(t\right)dt \\
&\quad + \frac{h^2}{2}f^{(2)}(x)\int_{-c}^{1} t^2k\left(t\right)dt + o\left(h^2\right) \\
&= f(x) - f(x)\int_{-1}^{-c} k\left(t\right)dt - hf^{(1)}(x)\int_{-c}^{1} tk\left(t\right)dt \\
&\quad + \frac{h^2}{2}f^{(2)}(x)\int_{-c}^{1} t^2k\left(t\right)dt + o\left(h^2\right),
\end{aligned}$$

therefore the value of bias of $f_{n,R}$ is

$$-f\left(x\right)\int_{-1}^{-c} k\left(t\right)dt + hf^{(1)}\left(x\right)\int_{-c}^{1} tk\left(t\right)dt + \frac{h^2}{2}f^{(2)}\left(x\right)\int_{-c}^{1} t^2k\left(t\right)dt + o\left(h^2\right). \tag{1.4}$$

Similar computations give the variance expression,

$$\frac{f\left(x\right)}{nh}\int_{-c}^{1} k^2\left(t\right)dt + o\left(\frac{1}{nh}\right). \tag{1.5}$$

However, the usual bias is

$$\frac{h^2}{2}f^{(2)}\left(x\right)\int_{-1}^{1} t^2k\left(t\right)dt + o\left(h^2\right), \tag{1.6}$$

for (1.4) and (1.6), we see that $f_{n,R}$ is not a consistent estimator of f and there exists an extra first order term of h. To correct this boundary problem, we construct a new approach, the basic technique of construction of the proposed estimator is kind of a generalized reflection method involving reflecting a transformation of the observed data. Then, a comparison of the boundary performance of our proposed estimator with the other kernel density estimators is carried out. It is well-known that a comparison between different methods is only meaningful with respect to their respective optimal performances. We have adopted this strategy in our comparison.

The rest of the paper is formulated as following. Section 2 introduces asymptotic properties of the proposed kernel estimator. Section 3 conducts Monte Carlo simulations and data analysis to compare the performance of our estimator, which is the main objective of this paper. The conclusion presents in Section 4.

2. Transformation-reflection Kernel Density Estimation

Using transformation and reflection method in kernel density estimations improved bias at the boundary, but unless the first derivative of the density is 0, the estimator with reflection can still be much more severely biased at the boundary than in the interior. Marron and Ruppert [6] propose to transform the data to a density that has its first derivative equal to 0 at both boundaries. The transformation is selected from a parametric family, which is allowed to be quite general in our theoretical study. Zhang and al [14] combine those two methods to construct a new approach which correct the boundary problem at the left side of the support. We use this technique to correct the boundary problem at the right side. The proposed estimator defined as follow,

$$f_{n,TR}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left\{k\left(\frac{x-X_i}{h}\right)+k\left(\frac{x-2+\psi(X_i)}{h}\right)\right\}. \tag{2.1}$$

The transformation ψ is stated in the theorem 2.1, which exhibits the explicit forms of the bias, variance and mean squared error (MSE), under certain conditions on ψ.

Theorem 2.1. *Assume that $\psi^{(3)}$ exist and is continuous, where $\psi^{(i)}$ denote the i^{th} derivative of ψ. Further assume that $\psi^{-1}(1)=1$ and $\psi^{(1)}(1)=1$, where ψ^{-1} is the*

inverse function of ψ. *Then for* $x = 1 - ch, 0 \leq c < 1$, *we have,*

$$
\begin{aligned}
Bias\left(f_{n,TR}\left(x\right)\right) = h \int_{-1}^{-c} (t+c) k\left(t\right) dt \left[2f^{(1)}\left(1\right) - f\left(1\right)\psi^{(2)}(1)\right] \\
+ \frac{h^2}{2}\mu_2 f^{(2)}(1) - \frac{h^2}{2}\int_{-1}^{-c}(t+c)^2 k\left(t\right) dt \\
\times \left[f\left(1\right)\psi^{(3)}(1) - 3\psi^{(2)}(1)\left[f^{(1)}\left(1\right) - f\left(1\right)\psi^{(2)}(1)\right]\right] \\
+ o\left(h^2\right),
\end{aligned}
\tag{2.2}
$$

and

$$
Var\left(f_{n,TR}\left(x\right)\right) = \frac{f(1)}{nh}\left[\int_{-1}^{1} k^2\left(t\right) dt + 2\int_{-c}^{1} k\left(t\right) k\left(-\left(2c+t\right)\right) dt\right] + o\left(\frac{1}{nh}\right).
$$

We shall choose the transformation ψ *so that the first order term in the bias expansions (2.2) is zero. Assume that* $f(1) > 0$, *it is enough to let,*

$$
\psi^{(2)}\left(1\right) = 2f^{(1)}\left(1\right)/f\left(1\right). \tag{2.3}
$$

So ψ *should satisfy the following three conditions:*

$C\left[1\right]$. ψ *is monotonically increasing.*
$C\left[2\right]$. $\psi^{(1)}\left(1\right) = 1$ *and* $\psi^{-1}\left(1\right) = 1$.
$C\left[3\right]$. $\psi^{(2)}\left(1\right) = 2f^{(1)}\left(1\right)/f\left(1\right)$.

The transformation function ψ, *verify the conditions* $C[1]$, $C[2]$ *and* $C[3]$, *has the form:*

$$
\psi\left(x\right) = M - BM^2 + \left(1 - 2M + 3BM^2\right)x + \left(M - 3BM^2\right)x^2 + BM^2x^3, \tag{2.4}
$$

where

$$
M = f^{(1)}\left(1\right)/f\left(1\right), \tag{2.5}
$$

and

$$
B > 1/3. \tag{2.6}
$$

For ψ be defined by (2.4) and for $x = 1 - ch, 0 \leq c < 1$, we have,

$$\begin{aligned} &Bias\left(f_{n,TR}\left(x\right)\right) \\ &= \frac{h^2}{2}\left\{\mu_2 f^{(2)}(1) - 6\left[B+1\right]\frac{\left[f^{(1)}\left(1\right)\right]^2}{f\left(1\right)} \times \int_{-1}^{-c}(t+c)^2 k\left(t\right)dt\right\} + o\left(h^2\right). \end{aligned} \tag{2.7}$$

Then the approximate form of mean squared error (MSE) is,

$$\begin{aligned} &MSE\left(f_{n,TR}\left(x\right)\right) \\ &\sim \frac{h^4}{4}\left(\mu_2 f^{(2)}(1) - 6\left[B+1\right]\frac{\left[f^{(1)}\left(1\right)\right]^2}{f\left(1\right)} \times \int_{-1}^{-c}(t+c)^2 k\left(t\right)dt\right)^2 \\ &+ \frac{f(1)}{nh}\left[\int_{-1}^{1} k^2\left(t\right)dt + 2\int_{-c}^{1} k\left(t\right)k\left(-\left(2c+t\right)\right)dt\right]. \end{aligned} \tag{2.8}$$

The mean integrated squared error $(MISE)$ of $f_{n,TR}\left(x\right)$ can be expressed as the sum of the integrated squared bias and the integrated variance for it,

$$MISE\left(f_{n,TR}\left(x\right)\right) = \int Bias^2\left(f_{n,TR}\left(x\right)\right)dx + \int Var\left(f_{n,TR}\left(x\right)\right)dx. \tag{2.9}$$

Estimation of ψ. In practice, the transformation ψ given by (2.4) is not available because it defined by unknown term M (2.5). We must replace M with a pilot estimator. Our proposed estimator (2.1) is not very sensitive to the accurate details of the pilot estimate of M, and therefore any appropriate estimate can be used. Note that M can be written as the derivative of $\log f(x)$ evaluated at $x = 1$, so M can be estimated by,

$$M_n = \frac{\log f_{n,R}\left(1\right) - \log f_{n,R}\left(1-h\right)}{h}, \tag{2.10}$$

we now define,

$$\psi_n\left(x\right) = M_n - BM_n^2 + \left(1 - 2M_n + 3BM_n^2\right)x + \left(M_n - 3BM_n^2\right)x^2 + BM_n^2 x^3, \tag{2.11}$$

as the estimator of $\psi\left(x\right)$.

The proposed new estimator. Our proposed new estimator of $f(x)$ is defined as, for $x = 1 - ch, 0 \leq c < 1$,

$$f_{n,TR,new}\left(x\right) = \frac{1}{nh}\sum_{i=1}^{n}\left\{k\left(\frac{x - X_i}{h}\right) + k\left(\frac{x - 2 + \psi_n(X_i)}{h}\right)\right\},$$

where ψ_n is given by (2.11) with M replaced by M_n of (2.10).

Proof. This proof starts by proving the bias of $f_{n,TR}$. We have

$$E(f_{n,TR}(x)) = \frac{1}{h}E\left\{k\left(\frac{x-X_1}{h}\right) + k\left(\frac{x-2+\psi(X_1)}{h}\right)\right\} = I_1 + I_2$$

By using change of variable and Taylor expansion for $x = 1 - ch, 0 \leq c < 1$, we can write,

$$\begin{aligned} I_1 &= \frac{1}{h}\int_0^1 k\left(\frac{x-y}{h}\right) f(y)\, dy \\ &= f(x)\int_{-c}^{1} k(t)\, dt - hf^{(1)}(x)\int_{-c}^{1} tk(t)\, dt + \frac{h^2}{2}f^{(2)}(x)\int_{-c}^{1} t^2 k(t)\, dt \\ &+ o\left(h^2\right). \end{aligned}$$

Therefore,

$$\begin{aligned} I_2 &= \int_{-1}^{-c} k(t)\frac{f\left(\psi^{-1}(th-x+2)\right)}{\psi^{(1)}(\psi^{-1}(th-x+2))}dt \\ &= \int_{-1}^{-c} k(t)\left[\frac{f\left(\psi^{-1}(1)\right)}{\psi^{(1)}(\psi^{-1}(1))}\right. \\ &+ (t+c)h\left[\frac{f^{(1)}\left(\psi^{-1}(1)\right)\psi^{(1)}(\psi^{-1}(1)) - f\left(\psi^{-1}(1)\right)\psi^{(2)}(\psi^{-1}(1))}{\left[\psi^{(1)}(\psi^{-1}(1))\right]^3}\right] \\ &+ (t+c)^2\frac{h^2}{2}\left[\frac{\psi^{(1)}(\psi^{-1}(1))f^{(2)}\left(\psi^{-1}(1)\right) - f\left(\psi^{-1}(1)\right)\psi^{(3)}(\psi^{-1}(1))}{\left[\psi^{(1)}(\psi^{-1}(1))\right]^4}\right. \\ &- \frac{3\psi^{(2)}(\psi^{-1}(1))\left[f^{(1)}\left(\psi^{-1}(1)\right)\psi^{(1)}(\psi^{-1}(1))\right.}{\left[\psi^{(1)}(\psi^{-1}(1))\right]^5} \\ &\left.\left.- \frac{f\left(\psi^{-1}(1)\right)\psi^{(2)}(\psi^{-1}(1))\right]}{\left[\psi^{(1)}(\psi^{-1}(1))\right]^5}\right] dt \\ &+ o\left(h^2\right). \end{aligned}$$

Using the condition $C\,[2]$ we have,

$$\begin{aligned}
&\frac{1}{h}E\left[k\left(\frac{x-2+\psi(X_1)}{h}\right)\right]\\
&=\int_{-1}^{-c} f\,(1)\,k\,(t)\,dt+h\left[f^{(1)}\,(1)-f\,(1)\,\psi^{(2)}(1)\right]\int_{-1}^{-c}(t+c)k\,(t)\,dt\\
&+\frac{h^2}{2}\left[f^{(2)}\,(1)-f\,(1)\,\psi^{(3)}(1)-3\psi^{(2)}(1)\left[f^{(1)}\,(1)-f\,(1)\,\psi^{(2)}(1)\right]\right]\\
&\times\int_{-1}^{-c}(t+c)^2k\,(t)\,dt\\
&+o\left(h^2\right).
\end{aligned}\tag{2.12}$$

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

$$\begin{aligned}
f\,(1)=&f(x)+chf^{(1)}(x)+\frac{(ch)^2}{2}f^{(2)}(x)+o(h^2).\\
f^{(1)}(x)=&f^{(1)}(1)-chf^{(2)}(1)+o(h).\\
f^{(2)}(x)=&f^{(2)}(1)+o(1).
\end{aligned}$$

So,

$$f\,(1)=f(x)+chf^{(1)}(1)-\frac{(ch)^2}{2}f^{(2)}(1)+o(h^2).\tag{2.13}$$

Now combining (I_1) and (I_2) and using the formula (2.13), we get,

$$\begin{aligned}
&Bias\left(f_{n,TR}\,(x)\right)\\
&=h\int_{-1}^{-c}(t+c)k\,(t)\,dt\left[2f^{(1)}\,(1)-f\,(1)\,\psi^{(2)}(1)\right]+\frac{h^2}{2}\mu_2 f^{(2)}(1)\\
&-\frac{h^2}{2}\int_{-1}^{-c}(t+c)^2k\,(t)\,dt\left[f\,(1)\,\psi^{(3)}(1)-3\psi^{(2)}(1)\left[f^{(1)}\,(1)-f\,(1)\,\psi^{(2)}(1)\right]\right]\\
&+o\left(h^2\right).
\end{aligned}$$

The task now is to prove the variance of $f_{n,TR}$: observe that for $x=1-ch, 0\leq c<1$, we have,

$$Var\left(f_{n,TR}\,(x)\right)=\frac{1}{nh^2}Var\left\{k\left(\frac{x-X_1}{h}\right)+k\left(\frac{x-2+\psi(X_1)}{h}\right)\right\}=J_1+J_2,$$

where,

$$\begin{aligned} J_1 &= \frac{1}{nh^2} E\left[k\left(\frac{x-X_1}{h}\right) + k\left(\frac{x-2+\psi(X_1)}{h}\right)\right]^2 \\ &= \frac{1}{nh^2}\left[\int_0^1 k^2\left(\frac{x-y}{h}\right) f(y)dy + \int_0^1 k^2\left(\frac{x-2+\psi(y)}{h}\right) f(y)dy\right] \\ &+ \frac{2}{nh^2}\int_0^1 k\left(\frac{x-y}{h}\right) k\left(\frac{x-2+\psi(y)}{h}\right) f(y)dy \\ &= J_{11} + J_{12}. \end{aligned}$$

Using a Taylor expansion, it can be shown that,

$$\begin{aligned} J_{11} &= \frac{1}{nh}\left[\int_{-c}^1 k^2(t)\, f(x-th)dt + \int_{-1}^{-c} k^2(t)\, \frac{f(\psi^{-1}(th-x+2))}{\psi^{(1)}(\psi^{-1}(th-x+2))}dt\right] \\ &= \frac{1}{nh}\left[\int_{-c}^1 k^2(t)\,[f(1)+o(1)]\,dt + \int_{-1}^{-c} k^2(t)\left(\frac{f(\psi^{-1}(1))}{\psi^{(1)}(\psi^{-1}(1))} + o(1)\right)dt\right] \\ &= \frac{f(1)}{nh}\mu_2 + o\left(\frac{1}{nh}\right), \end{aligned}$$

and,

$$\begin{aligned} J_{12} &= \frac{2}{nh}\int_{-c}^1 k(t)\, k\left(\frac{x-2+\psi(x-th)}{h}\right) f(x-th)dt \\ &= \frac{2}{nh}\int_{-c}^1 k(t)\, k\left(\frac{1-ch-2+1-(c+t)h)+o(h)}{h}\right) f(1-(c+t)h)dt \\ &= \frac{2f(1)}{nh}\int_{-c}^1 k(t)\, k(-(2c+t))\,dt + o(\frac{1}{nh}). \end{aligned}$$

Similarly as in the proof of J_1, we get

$$J_2 = -\frac{1}{nh^2}\left[E^2\left(k\left(\frac{x-X_1}{h}\right) + k\left(\frac{x-2+\psi(X_1)}{h}\right)\right)\right] = o\left(\frac{1}{nh}\right).$$

By adding up J_1 and J_2, we have the desired result for the variance. □

3. SIMULATION STUDY

To compare the performance of our proposed estimator against the other well known estimators, we divided this section into two parts, in the first, we introduce the results of simulated data and in the second part, we present some examples of real data. All computations were done by utilizing R software.

3.1. **Simulated Data.** In our simulation study reported in this part, we introduced the issue of potential quality of our proposed estimator per se form that of bandwidth selection. Throughout our study we use Epanechnikov kernel $k(t) = (3/4)(1-t^2)\mathbb{I}(-1 \leq t \leq 1)$, where $\mathbb{I}$ denote the indicatrice function.

3.1.1. *Smoothing Parameter Selection.* It is well known that the kernel estimation of the density depends crucially on the bandwidths. In our study, we used two methods of smoothing parameter selection which are the optimal bandwidth and the cross validation method.

Optimal Bandwidth. The popular bandwidth selector in kernel density estimation is due to Sheather and Jones [18]. This method adopts the asymptotic $MISE$ as criterion, defined by

$$AMISE \sim \frac{h^4}{4}\mu_2^2 \int \left[f^{(2)}(x)\right]^2 dx + \frac{1}{nh}\int k^2(t)dt, \tag{3.1}$$

the optimal bandwidth minimizing (3.1) is,

$$h_{opt} = \left\{ \int k^2(t)dt / n\mu_2^2 \int \left[f^{(2)}(x)\right]^2 dx \right\}^{1/5}.$$

Cross Validation Method. Rudemo [9] and Bowman [1] suggested known as unbiased cross-validation (UCV) in kernel density estimator, is surely the most popular and exceed studied one. The basic thought of this strategie, it purpose to estimate h the minimizer of $ISE(h)$. The minimisation measure is characterized by,

$$h_{ucv} = \arg\min_h UCV(h), \tag{3.2}$$

where

$$UCV(h) = \int f_{n,R}^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n} f_{n,R}(x_i). \tag{3.3}$$

3.1.2. *Compared Estimators.* We compare the performance of the kernel density estimator $f_{n,R}$, the transformation-reflection kernel density estimation $f_{n,TR}$, the boundary kernel estimator $f_{n,B}$ and the Jones and Foster estimator $f_{n,JF}$. The comparison is carried out with respect to the different densities.

The boundary kernel estimator is the general boundary corrected estimators define by Jones [3], which replace the standard kernel function by the modified version. The modified kernel function gave at the right boundary region based on the Epanechnikov kernel, by

$$k_B(t) = 12\frac{1-t}{(1+c)^4}\left(\frac{3c^2-2c+1}{2} - t(1-2c)\right)\mathbb{I}(-c \le t \le 1), \tag{3.4}$$

this kernel satisfies the following conditions,

$$\int_{-1}^{-c} k_B(t)\,dt = 0,\ \int_{-c}^{1} k_B(t)\,dt = 1,\ \int_{-c}^{1} t k_B(t)\,dt = 0,\ \int_{-c}^{1} t^2 k_B(t)\,dt < \infty, \tag{3.5}$$

the boundary kernel estimator is defined as,

$$f_{n,B}(x) = \frac{1}{nh}\sum_{i=1}^{n} k_B\left(\frac{x-X_i}{h}\right). \tag{3.6}$$

The Jones and Foster estimator that corrects for the possible negativity of the boundary kernel estimates has the following form,

$$f_{n,JF}(x) = f_{n,CN}(x)\exp\left(\frac{f_{n,B}(x) - f_{n,CN}(x)}{f_{n,CN}(x)}\right), \tag{3.7}$$

where,

$$f_{n,CN}(x) = \frac{1}{nh}\sum_{i=1}^{n} k_{CN}\left(\frac{x-X_i}{h}\right), \tag{3.8}$$

denote the cut-and-normalized density estimator introduced by Gasser and Muller [11] and by using the kernel function k_{CN} for the right boundary region truncated and normalized, ensuring integration to unity. For Epanechnikov kernel, it given by

$$k_{CN}(t) = \frac{1-t^2}{\int_{-c}^{1}\left(1-t^2\right)dt}\mathbb{I}(-c \le t \le 1) \tag{3.9}$$

3.1.3. *Simulation Steps.* We can compile the steps of simulation as follow,

Step 1. We simulate sample of size n with R repetition from the probability density f.

Step 2. We calculate h using the optimal bandwidth and the cross validation method.

Step 3. We estimate f by $f_{n,R}$, $f_{n,TR}$, $f_{n,B}$ and $f_{n,JF}$.

Step 4. We compute the $SBIS$, VAR and MSE of each estimator.

Step 5. We graph the $MISE$ in the form of a boxplot.

For evaluating the performance of estimators at the boundaries, we tend to limit our attention to $n = 200$. We consider six distributions with bounded support $[0, 1]$. This set of distributions illustrated in Table 1, and for each distrubution we simulate $R = 1000$.

TABLE 1. Densities used in the simulation

Distribution		Density function
D_1	Beta $(3/2, 1)$	$(3/2)x^{1/2}$
D_2	Truncated Gamma $(2, 1)$	$x\exp(-x)/1 - 2\exp(-1)$
D_3	Truncated Normal $(0, 1)$	$\exp(-x^2/2)/\int_0^1 \exp(-t^2/2)dt$
D_4	$(1/2)$Beta $(3/2, 1)$ +$(1/2)$Beta $(1, 3/2)$	$0.75x^{1/2} + 0.75(1-x)^{1/2}$
D_5	Truncated Exponential (1)	$\exp(-x)/1 - \exp(-1)$
D_6	Truncated Beta $(5, 1)_{[1/2,1]}$	$160x^4/31$

3.1.4. *Results and discussions.* For each density, we have calculated the squared bias ($SBIS$), variance (VAR) and mean squared error (MSE) of the estimators at the endpoint $x = 1$ using the two methods of smoothing parameter selection. The results are presented in Table 2 and Table 3.

Comparing between the estimators, we can see from the table 2, which includes simulated values of $SBIS$, VAR and MSE calculated by using the optimal bandwidth, that $f_{n,TR}$ had the smallest values of $SBIS$, VAR and MSE among the other estimators for all the cases considered, followed by $f_{n,B}$ and $f_{n,JF}$ estimators, while the $f_{n,R}$ estimator is the worst among them, as to the Beta, Truncated

TABLE 2. The squared bias, variance and MSE values computed using the optimal bandwidth

Density		D_1	D_2	D_3	D_4	D_5	D_6
Estimator		$h = 0.2122$	$h = 0.1884$	$h = 0.2259$	$h = 0.2341$	$h = 0.2304$	$h = 0.2378$
$f_{n,R}$	$SBIS$	0.0954	0.0832	0.0221	0.0315	0.0148	0.0434
	VAR	0.0584	0.0466	0.0109	0.0132	0.0078	0.0248
	MSE	0.1538	0.1298	0.0331	0.0447	0.0226	0.0683
$f_{n,TR}$	$SBIS$	0.0205	0.0220	0.0096	0.0119	0.0078	0.0133
	VAR	0.0094	0.0077	0.0025	0.0030	0.0022	0.0048
	MSE	0.0299	0.0297	0.0121	0.0149	0.0100	0.0182
$f_{n,B}$	$SBIS$	0.0222	0.0262	0.0123	0.0149	0.0108	0.0163
	VAR	0.0156	0.0172	0.0073	0.0093	0.0065	0.0107
	MSE	0.0378	0.0433	0.0196	0.0242	0.0173	0.0270
$f_{n,JF}$	$SBIS$	0.0228	0.0267	0.0121	0.0143	0.0106	0.0165
	VAR	0.0173	0.0183	0.0069	0.0088	0.0062	0.0116
	MSE	0.0402	0.0451	0.0190	0.0231	0.0167	0.0282

TABLE 3. The squared bias, variance and MSE values computed using the cross validation method

Density		D_1	D_2	D_3	D_4	D_5	D_6
Estimator		$h = 0.1324$	$h = 0.0865$	$h = 0.0945$	$h = 0.2256$	$h = 0.1297$	$h = 0.1729$
$f_{n,R}$	$SBIS$	0.0946	0.0803	0.0292	0.0353	0.0209	0.0502
	VAR	0.0687	0.0583	0.0141	0.0162	0.0083	0.0289
	MSE	0.1633	0.1385	0.0433	0.0516	0.0292	0.0791
$f_{n,TR}$	$SBIS$	0.0463	0.0415	0.0223	0.0240	0.0146	0.0349
	VAR	0.0189	0.0157	0.0054	0.0055	0.0032	0.0092
	MSE	0.0652	0.0572	0.0278	0.0295	0.0178	0.0440
$f_{n,B}$	$SBIS$	0.0566	0.0511	0.0284	0.0298	0.0191	0.0444
	VAR	0.0331	0.0298	0.0144	0.0153	0.0102	0.0217
	MSE	0.0897	0.0809	0.0429	0.0452	0.0293	0.0661
$f_{n,JF}$	$SBIS$	0.0575	0.0518	0.0286	0.0299	0.0194	0.0449
	VAR	0.0320	0.0280	0.0146	0.0152	0.0114	0.0224
	MSE	0.0895	0.0798	0.0432	0.0451	0.0308	0.0674

Gamma and the last one Truncated Beta, but as to Truncated Normal, Mixture Beta and Truncated Exponontiel, the $f_{n,B}$ changed your position to the third place and $f_{n,JF}$ came in the second place.

From the table 3, which includes simulated values of $SBIS$, VAR and MSE calculated by using the unbasied cross validation method, it can be observe that $f_{n,TR}$ has good performance among the others. Moreover, the ranking from best to worst concerning the $SBIS$ is $f_{n,TR}$, $f_{n,B}$, $f_{n,JF}$, $f_{n,R}$ for all densities, but concerning the MSE we can observe that $f_{n,B}$ change it's order to the third place and $f_{n,JF}$ came to the second place for the Beta, Truncated Gamma and Mixture Beta densities, but for the other densities Truncated Normal, Truncated Exponentiel and Truncated Beta $f_{n,B}$ and $f_{n,JF}$ came in the second, third place respectively.

Comparing two smoothing parameter for a given estimators, we also find that, in general, the optimal bandwidth tends to perform better.

3.2. **Real Data.** In this section, we apply our proposed estimators over two data sets. The densities of our data sets are assumed to have a compact support $S = [a, b]$. In our study, we mapped the original observation $X_i \in S$ onto the unit interval by the transformation $Y_i = (X_i - a)/(b - a)$.

The natural stands of the seedlings and saplings of Japanese black pines. The first data set consists the data were collected on the position, height (cm) and age $(years)$ of the natural stands of the seedlings and saplings of 204 Japanese black pines in a $100m^2$ region. The data set can be found in the paper of Ogata Y. and Tanemura M. (1985). Table 4 shows the descriptive statistics of the data. We used bandwidth $h = 0.14$, which we choose subjectively. We have graphed the performance of our proposed estimator and the histogram of unknown real density function in figure 1.

TABLE 4. Descriptive statistics of the natural stands of the seedlings and saplings of Japanese black pines data.

Min	1st Qu	Median	Mean	3rd Qu	Max	kurtosis	skewness
3.10	10.20	15.40	31.92	46.52	150.2	6.70	1.91

From the figure 1 alone, one can see that the $f_{n,TR}$ is a good estimator of the true density removes a large part of the boundary effect and when we move to the interior, we remark that all the estimators close to the kernel density estimator. We can conclude that, $f_{n,TR}$ yield the best estimator of natural stands of the seedlings and saplings of Japanese black pines data and hence can be adequat for estimation these data.

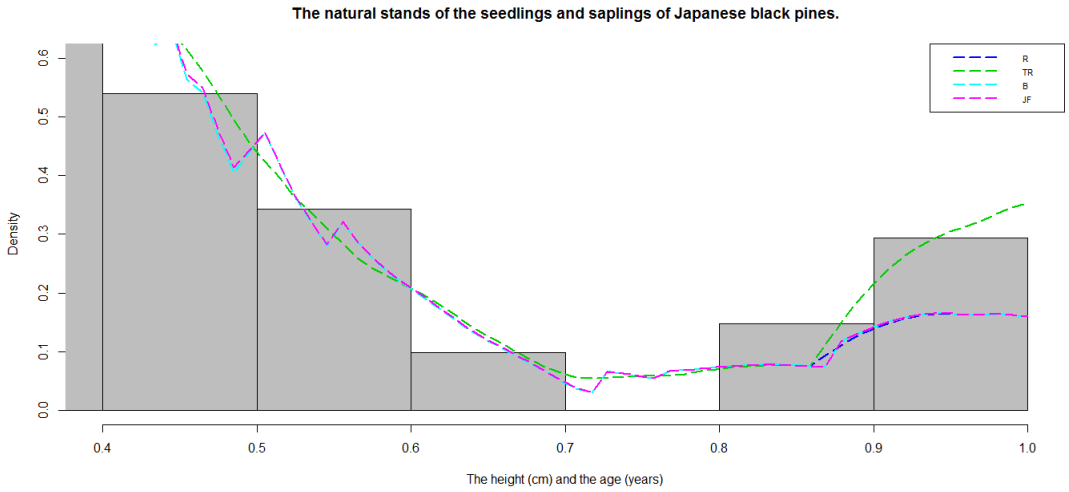

FIGURE 1. Density estimates of the natural stands of the seedlings and saplings of 204 Japanese black pines.

The measure of the motor cortex neuron interspike of unstimulated monkey. The second data set is the measres intervals of motor cortex neuron interspike (in ms) for an unstimulated monkey. The objects of the analysis were to estimate the firing rate prior to Stimulation and to characterize the time dependence. The data set can be found in the paper of Zeger, S.L. and Bahjat Qaqish (1988). The descriptive statistics of the data is given in Table 5. The bandwidth is chosen subjectively to be $h = 0.16$. The proposed estimators are plotted in figure 2, superimposed on the histogram of the data.

TABLE 5. Descriptive statistics of the measrues intervals of motor cortex neuron interspike for an unstimulated monkey data.

Min	1st Qu	Median	Mean	3rd Qu	Max	kurtosis	skewness
2.00	20.00	29.50	36.49	49.25	104	1.08	3.87

From figure 2 we can see that the $f_{n,TR}$ is closer to the empirical histogram of the density. That indicates, $f_{n,TR}$ is well covers the density of the measure of the motor cortex neuron interspike of unstimulated monkey data.

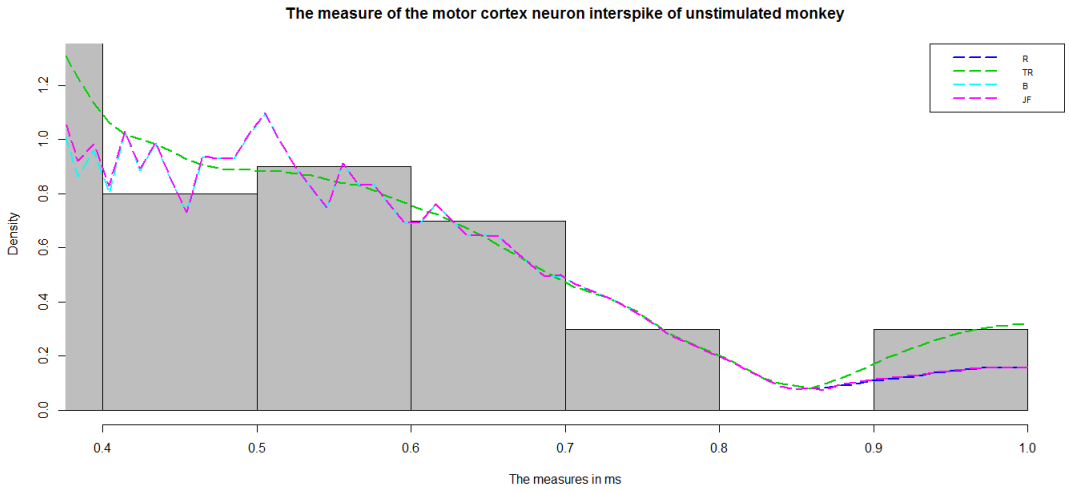

FIGURE 2. Density estimates of the measres intervals of motor cortex neuron interspike for an unstimulated monkey.

4. CONCLUSION

In this paper, we proposed a new approach reducing the bias of the kernel density estimator near the right bord of the support $[0, 1]$. Our estimator has a good comparable performances in approximating the true density. It perform better than the other well known kernel density estimators. We also noticed the consistent property of our estimator for estimating the density at the endpoint $x = 1$. So, we show that the new proposed estimator here improves the bias in such point.

REFERENCES

[1] A.W. BOWMAN: *An alternative method of cross-validation for the smoothing of density estimates*, Biometrika, **71**(1984), 353–360.

[2] A. COWLING, P. HALL: *On pseudodata methods for removing boundary effects in kernel density estimation*, Journal of the Royal Statistical Society: Series B (Methodological), **58**(3), (1996), 551–563.

[3] M.C. JONES: *Simple boundary correction for kernel density estimation*, Statistics and computing. **3**(3) (1993), 135–146.

[4] R.J. KARUNAMUNI, T. ALBERTS: *On boundary correction in kernel density estimation*, Statistical Methodology, **2**(3) (2005), 191–212.

[5] R.J. KARUNAMUNI, T. ALBERTS: *A locally adaptive transformation method of boundary correction in kernel density estimation*, Journal of Statistical Planning and Inference, **136**(9) (2006), 2936–2960.

[6] J.S. MARRON, D. RUPPERT: *Transformations to reduce boundary bias in kernel density estimation*, Journal of the Royal Statistical Society: Series B (Methodological), **56**(4) (1994), 653–671.

[7] E. PARZEN: *On estimation of a probability density function and modeOn estimation of a probability density function and mode*, The annals of mathematical statistics. **33**(3) (1962), 1065–1076.

[8] M. ROSENBLATT: *Remarks on some nonparametric estimates of a density function*, The Annals of Mathematical Statistics, (1956), 832–837.

[9] M. RUDEMO: *Empirical choice of histograms and kernel density estimators*, Scandinavian Journal of Statistics, (1982), pages 65–78.

[10] B.W. SILVERMAN: *Density estimation for statistics and data analysis*, Chapman and Hall, 1986.

[11] T. GASSER, H.-G. MÜLLER: *Kernel estimation of regression functions*, In Smoothing techniques for curve estimation, 23—68. Springer, 1979.

[12] S. ZHANG, R.J. KARUNAMUNI: *On kernel density estimation near endpoints*, Journal of statistical Planning and inference, **70**(2) (1998), 301–316.

[13] S. ZHANG, R.J. KARUNAMUNI: *On nonparametric density estimation at the boundary*, Journal of nonparametric statistics. **12**(2) (2000), 197–221.

[14] S. ZHANG, R.J. KARUNAMUNI, M.C. JONES: *An improved estimator of the density function at the boundary*, Journal of the American Statistical Association. **94**(448) (1999), 1231–1240.

[15] M.P. WAND, M.C. JONES: *Kernel smoothing*, CRC Press, 1994.

[16] M.C. JONES, P.J. FOSTER: *A simple nonnegative boundary correction method for kernel density estimation*, Statistica Sinica, 1005–1013.

[17] E.F. SCHUSTER: *Incorporating support constraints into nonparametric estimators of densities*, Communications in Statistics-Theory and methods. **14**(5) (1985), 1123–1136.

[18] S.J. SHEATHER, M.C. JONES: *A reliable data-based bandwidth selection method for kernel density*, ournal of the Royal Statistical Society. Series B (Methodological), **53**(3) (1991), 683–690.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MOHAMED KHIDER, BISKRA, ALGERIA.
Email address: bourdjistat23@gmail.com

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MOHAMED KHIDER, BISKRA, ALGERIA.
Email address: sayahabdel@yahoo.fr