INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0108

Consistency of Beta Kernel Density Function Estimator

T.BOUEZMARNI and J.M.ROLIN

http://www.stat.ucl.ac.be

Consistency of Beta Kernel Density Function Estimator

T.Bouezmarni * and J.M.Rolin*

March 5, 2001

Abstract

We consider the Beta Kernel estimator for an unknown probability density function f with a compact support $[0,1]$:

$$f_b(x) = \frac{1}{n}\sum_{i=1}^{n} K(X_i, x/b+1, (1-x)/b+1) \tag{1}$$

where the kernel K is the beta density function with parameters x/b and $(1-x)/b$, b is the smoothing bandwidth and $X_1, \cdots, X_n$ is a random sample from a distribution F with an unknown probability density function f.

For such estimator, we obtain, for densities f in $C^2([0,1])$, the exact asymptotic behaviour of $I\!E(J_{bn})$, where

$$J_{bn} = \int_0^1 \mid f_b(x) - f(x) \mid dx$$

Further, we show that there exist a $B(f)$ depending only upon f such that

$$\inf_b n^{2/5} I\!E(J_{bn}) \leq C^* B(f) + o(1),$$

and $C^* = 1.3768102\cdots$ is a universal constant.

The uniform weak consistency on $[0,1]$ is proved when f is continuous on $[0,1]$. For $f \in C((0,1))$ and unbounded in 0 (or 1), we prove that the Beta kernel estimator converges in probability to infinity in 0 (or 1). By simulations, comparison in $L_1[0,1]$ of the Beta Kernel estimator with Bernstein estimator and another mixture of beta's will be shown.

Key words and phrases. Beta kernel, density estimation, weak convergence, unbounded density, L_1 error.

*Institute of Statistics, Catholic University of Louvain.

1 Introduction

Let $X_1, \cdots, X_n$ be a random sample from a distribution F with an unknown probability density function f. The Beta kernel estimator for f is

$$f_b(x) = \frac{1}{n} \sum_{i=1}^{n} K(X_i, x/b + 1, (1-x)/b + 1) \tag{2}$$

where the kernel K is the beta density function with parameter x/b and $(1-x)/b$ and b is the smoothing bandwidth.

By using the Beta density function as a kernel in the generalised kernel estimator introduced for densities with unbounded support, Chen(1999) introduces the Beta kernel estimator as an estimator of density function with known compact support $[0,1]$, to remove the boundary bias of the standard kernel introduced by Rosenblatt. The Beta kernel has three special features. The first is that the Beta kernel is easy in concept and in its implementation, Chen(1999) proves that it's free of boundary bias and has mean integrated squared error of $O(n^{-4/5})$. The second is that the shape of the Beta kernel changes according to the position where the density is estimated. We can see, in Figure (1), the special shape of the Beta kernel at $x = 0$ and $x = 1$. We will come back to the last remark in Section 4, where the density function is supposed unbounded in 0 or 1. The other feature is that the support of the Beta kernel matches the support of the density function.

In order to find some information about the smoothness of a density and best possible L_1 rate of convergence that can be attained by a Beta kernel, we examine the asymptotic expression for the mean integrated absolute deviation:

$$\mathbb{E}(J_{bn}) = \mathbb{E}\left(\int_0^1 \mid f_b(x) - f(x) \mid dx \right).$$

In Section 2, we prove that for $f \in C^2([0,1])$, and

$$\lim_{n \longrightarrow \infty} b = 0 \quad \text{and} \quad \lim_{n \longrightarrow \infty} n^{-1}\, b^{-1/2} = 0 \tag{3}$$

there exists a constant $B(f)$ depending only on f such that

$$\inf_b n^{2/5} \mathbb{E}(J_{bn}) \leq C^*\, B(f) + o(1),$$

where $C^* = 1.3768102 \cdots$ is a universal constant.

For such class of densities the exact asymptotic expression of $\mathbb{E}(J_{bn})$ will be presented. In Section 3, we prove the uniform weak convergence of the Beta kernel estimator for $f \in C([0,1])$. In Section 4, we consider a Beta kernel estimator for density functions which are unbounded in 0 (or 1). We prove its convergence in probability to infinity in 0 (or 1) for the density function$f$ under the two following conditions on the bandwidth parameter:

$$\lim_{n\longrightarrow\infty} b = 0 \quad \text{and} \quad \lim_{n\longrightarrow\infty} n\, b^2 = +\infty. \tag{4}$$

In Section 5, by simulations we compare in $L_1[0,1]$ the Beta Kernel estimator with the Bernstein estimator and a new mixture of beta's.

2 L_1- rates of Convergence of Beta Kernel

In this section, we obtain for $f \in C^2([0,1])$, the exact asymptotic behaviour for the expected $L_1[0,1]$-error of Beta kernel estimates:

$$\mathbb{E}(J_{bn}) = \mathbb{E}\left(\int_0^1 \mid f_b(x) - f(x) \mid dx\right)$$

First, we verify that for any density function f, his Beta kernel estimator f_b is an element of L_1.

Let $R(x) = \sqrt{2\pi} x^{x+1/2} e^{-x}/\Gamma(x+1), x > 0$. The R-function has two important properties:

$$R(x) < 1, \text{ for all } x > 0, \quad \text{and } R(x) \longrightarrow 1 \text{ as } \ x \longrightarrow \infty.$$

Using the Stirling's formula, properties of R-function and the fact that x is the mode of $K(t, x/b+1, (1-x)/b+1)$, we can show that:

$$K(t, x/b+1, (1-x)/b+1) \leq c_1\, b^{-1/2}\{x(1-x)\}^{-1/2} \tag{5}$$

where c_1 is a positive constant (see Chen(2000)).

Now we show that f_b is in L_1, using the inequality (5),

$$\begin{aligned}
\int_0^1 f_b(x)\, dx &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 K(X_i, x/b+1, (1-x)/b+1)\, dx \\
&\leq c_1\, b^{-1/2}\int_0^1 \{x(1-x)\}^{-1/2}\, dx \\
&\leq c_1\, b^{-1/2}\, \pi \, < \, \infty.
\end{aligned}$$

■

For the remainder of the paper we will need some notations:

$$\alpha(x) = \sqrt{\frac{f(x)}{2\sqrt{\pi x(1-x)}}}$$

$$\gamma(x) = \frac{f'(x)(1-2x)}{2} \quad \text{and} \quad \xi(x) = \frac{f''(x)x(1-x)}{2}.$$

The following quantity is closely linked to the rate of convergence of $\mathbb{E}(J_{bn})$,

$$B(f) = \left(\int_0^1 \alpha(x)dx\right)^{4/5}\left(\int_0^1 (\mid \gamma(x) \mid + \mid \xi(x) \mid)\, dx\right)^{1/5}. \tag{6}$$

The quantity $B(f)$ will be used for all densities f belonging to $C^2([0,1])$.

Finally, to describe the exact asymptotic behaviour of $\mathbb{E}(J_{bn})$, we introduce the function

$$\psi(u) = \sqrt{\frac{2}{\pi}}\left(u\int_0^u e^{-x^2/2}dx + e^{-u^2/2}\right), \qquad u \geq 0.$$

Note that for $u \geq 0$,

$$\psi(u) = \mathbb{E}(\mid Z-u \mid) \quad \text{where} \quad Z \sim N(0,1),$$

and

$$\max\left\{u, \sqrt{\frac{2}{\pi}}\right\} \leq \psi(u) \leq u + \sqrt{\frac{2}{\pi}}$$

We state now, our main theorem.

Theorem 2.1 *(Exact asymptotic behaviour of $\mathbb{E}(J_{bn})$). Let f be a density function in $C^2([0,1])$, f_b the Beta kernel estimator of f, under (3),*

$$\mathbb{E}(J_{bn}) = J(b,n) + o\left(b + n^{-1/2}b^{-1/4}\right)$$

where

$$J(b,n) = \int_0^1 (n^{1/2}b^{1/4})^{-1}\,\alpha(x)\,\psi\left((n^{1/2}b^{5/4})\frac{\mid \gamma(x) \mid + \mid \xi(x) \mid}{\alpha(x)}\right) dx$$

Also

$$J(b,n) \leq \sqrt{\frac{2}{\pi}} \int_0^1 (n^{1/2} b^{1/4})^{-1} \, \alpha(x) dx + b \int_0^1 (\mid \gamma(x) \mid + \xi(x) \mid) \, dx.$$

In particular,

$$\limsup_{n\to\infty} \inf_{b>0} n^{2/5} \mathbb{E}(J_{bn}) \leq C^* \, B(f),$$

where $C^* = 1.3768102\cdots$ *is a universal constant.*

■

The upper bound is not exceeded for the following choice of b, when f is a density on $[0,1]$ and $B(f) < \infty$:

$$b_n = \left[\frac{\sqrt{\frac{2}{\pi}} \int_0^1 \alpha(x) \, dx}{\int_0^1 4 \left(\mid \gamma(x) \mid \; + \mid \xi(x) \mid\right) \, dx} \right]^{4/5} n^{-2/5}$$

Remark 1 *For a large value of* b*, the Beta kernel becomes very smooth and close to the uniform* $[0,1]$*. We can note this from the expression of the Beta kernel estimator and the presence in the above expression of* b*, of the component*

$$\int_0^1 4 \left(\mid \gamma(x) \mid \; + \mid \xi(x) \mid\right) \, dx$$

which measures the oscillation of the density f*.*

■

Theorem (2.1) states that for densities f on $[0,1]$ with finite $B(f)$, $\mathbb{E}(J_{bn})$ decreases at rate $n^{-2/5}$ if b is chosen in proportion to $n^{-2/5}$. A similar upper bound for standard kernel estimates was obtained by Rosenblatt(1956) and Devroye and Penrod(1982). Also a similar upper bound for Bernstein estimates was obtained by Tenbusch(1998).
For the remainder of the paper we will need some more notations:
$B(x) = \mathbb{E}(f_b(x)) - f(x)$ is the bias of the Beta kernel estimator at x and the variance of the Beta kernel at x is $\sigma_n^2(x) = \mathbb{E}[(f_b(x) - \mathbb{E}(f_b(x)))^2]$.

To prove theorem(2.1), we use the five following lemmas.

Lemma 2.1 *(Devroye and Penrod 1982). Let $X_1, \cdots, X_n$ be independent random variables with common distribution. Let $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) = \sigma^2 > 0$ and $\mathbb{E}(\mid X_i \mid^3) = \rho < \infty$. Then*

$$\sup_{a \in \mathbb{R}} \left| \mathbb{E}\left(\left| (\sigma\sqrt{n})^{-1} \sum_{i=1}^{n} X_i - a \right| \right) - \mathbb{E}(\mid Z - a \mid) \right| \leq \frac{c\rho\sigma^{-3}}{\sqrt{n}},$$

where c is an universal positive constant and Z is the standard normal random variable. Note that,

$$\psi(\mid u \mid) = \mathbb{E}(\mid Z - u \mid).$$

■

Lemma 2.2 *For all f defined on $[0, 1]$,*

$$\left| \mathbb{E}\left(\mid f_b(x) - f(x) \mid\right) - \sigma_n(x)\psi\left(\frac{\mid B(x) \mid}{\sigma_n(x)}\right) \right| \leq \frac{cc_1 b^{-1/2}}{n\sqrt{x(1-x)}}$$

for some positive constant c_1.

■

Proof. We apply lemma(2.1) to the random variables

$$Y_i = K(X_i, x/b + 1, (1-x)/b + 1) - \mathbb{E}(K(X_i, x/b + 1, (1-x)/b + 1))$$

and $a = B(x)/\sigma_n(x)$, we obtain

$$\left| \mathbb{E}\left(\left| (\sigma_Y\sqrt{n})^{-1} \sum_{i=1}^{n} Y_i - B(x)/\sigma_n(x) \right| \right) - \mathbb{E}(\mid Z - B(x)/\sigma_n(x) \mid) \right| \leq \frac{c\rho_Y\sigma_Y^{-3}}{\sqrt{n}}$$

where

$$\sigma_Y^2 = var(Y_i) = n\sigma_n^2(x)$$

and

$$\rho_Y = \mathbb{E}(\mid Y_i \mid^3).$$

With some computations and using the inequality(5) we get,

$$\begin{aligned} \left| \mathbb{E}\left(\mid f_b(x) - f(x) \mid\right) - \sigma_n(x)\psi\left(\frac{\mid B(x) \mid}{\sigma_n(x)}\right) \right| &\leq \frac{c\rho_Y\sigma_Y^{-3}\sigma_n(x)}{\sqrt{n}} \\ &= c\frac{\rho_Y}{n\sigma_Y^2} \\ &\leq \frac{cc_1 b^{-1/2}}{n\sqrt{x(1-x)}}. \end{aligned}$$

■

2.1 Bias and Variance of the Beta Kernel Estimator

For the proof of theorem(2.1) we need further results about the bias and variance of the Beta kernel. Since the pointwise results are due to Chen(1999) we only sketch the proofs of lemma 2.3 and lemma2.4.

Lemma 2.3 *for all $f \in C^2([0,1])$, we have*

$$\int_0^1 || B(x) | - b\,(| \gamma(x) | + | \xi(x) |)| \, dx = o(b)$$

$\blacksquare$

Proof From Chen(1997),

$$B(x) = b \left\{ \frac{f'(x)(1-2x) + f''(x)x(1-x)}{2} \right\} (1 + o(1)).$$

Then,

$$\int_0^1 \left| | B(x) | - b \frac{| f'(x)(1-2x) | + | f''(x)x(1-x) |}{2} \right| dx$$

$$\leq \int_0^1 \left| B(x) - b \frac{f'(x)(1-2x) + f''(x)x(1-x)}{2} \right| dx$$

$$\leq \frac{b}{2} \left\{ \int_0^1 | f'(x) || 1 - 2x | \, dx + \int_0^1 | f''(x) || x(1-x) | \, dx \right\} o(1)$$

$$= o(b).$$

$\blacksquare$

Lemma 2.4 *for all $f \in C^2([0,1])$, we have*

$$\int_0^1 \left| \sigma_n(x) - \alpha(x) \sqrt{n^{-1} b^{-1/2}} \right| dx = o\left(n^{-1/2} b^{-1/4} \right)$$

$\blacksquare$

Proof From Chen(1997),

$$\sigma_n(x) = n^{-1}b^{-1/2}\frac{f(x)}{2\sqrt{\pi x(1-x)}}(1+o(1)).$$

Then,

$$\begin{aligned}
&\int_0^1 \left|\sigma_n(x) - \alpha(x)\sqrt{n^{-1}b^{-1/2}}\right| dx \\
&= \left(\int_0^1 \alpha(x)\sqrt{n^{-1}b^{-1/2}}dx\right) o(1) \\
&\leq \left(n^{-1/2}b^{-1/4}\int_0^1 \alpha(x)\,dx\right) o(1) \\
&= o\left((n^{-1/2}b^{-1/4})\right).
\end{aligned}$$

■

Before proving theorem(2.1), we need one last lemma due to Devroye and Penrod(1982).

Lemma 2.5 *(Devroye and Penrod(1982)) For nonnegative numbers u, v, w, z we have*

$$\left|u\psi\left(\frac{v}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| \leq \mid v - z \mid + \sqrt{\frac{2}{\pi}} \mid u - w \mid .$$

■

2.2 Proof of theorem (2.1)

We begin the proof of the theorem by using lemma(2.2).

$$\begin{aligned}
\mid \mathbb{E}(J_{bn}) - J(b,n) \mid \;\leq\; & \left|\int_0^1 \mathbb{E} \mid f_{bn}(x) - f(x) \mid dx - \int_0^1 \sigma_n(x)\,\psi\left(\frac{\mid B(x) \mid}{\sigma_n(x)}\right) dx\right| \\
& + \left|\int_0^1 \sigma_n(x)\,\psi\left(\frac{\mid B(x) \mid}{\sigma_n(x)}\right) dx - J(b,n)\right| \\
\leq\; & c\,c_1\frac{\pi}{nb^{1/2}} + \left|\int_0^1 \sigma_n(x)\,\psi\left(\frac{\mid B(x) \mid}{\sigma_n(x)}\right) dx - J(b,n)\right|.
\end{aligned}$$

In the last inequality we use the fact that

$$\int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = \pi.$$

Now we use lemma(2.5) with:

$$u = \sigma_n(x), \qquad v = \mid B(x) \mid$$

and

$$w = n^{-1/2} b^{-1/4} \alpha(x), \qquad z = b(\mid \gamma(x) \mid + \mid \xi(x) \mid).$$

Note that

$$\int_0^1 \mid v - z \mid dx = o(b) \quad \text{by lemma(2.3)}$$

and

$$\int_0^1 \mid u - w \mid = o(n^{-1/2} b^{-1/4}) \quad \text{by lemma(2.4)}.$$

Then the last term in the upper bound of $\mid I\!\!E(J_{bn}) - J(b,n) \mid$ is bounded by

$$\begin{aligned} c\, c_1 \frac{\pi}{n b^{1/2}} \quad &+ \quad \int_0^1 \mid\mid B(x) \mid - b(\mid \gamma(x) \mid + \mid \xi(x) \mid) \mid dx \\ &+ \quad \sqrt{\frac{2}{\pi}} \int_0^1 \left| \sigma(x) - \sqrt{\frac{1}{n\, b^{1/2}} \alpha(x)} \right| dx \\ &= \quad c\, c_1 \frac{\pi}{n b^{1/2}} + o(b) + o(b^{-1/4} n^{-1/2}) \\ &= \quad o(b + b^{-1/4} n^{-1/2}). \end{aligned}$$

Thus we have proved that

$$I\!\!E(J_{bn}) = J(b,n) + o(b + b^{-1/4} n^{-1/2}) \quad \text{for all} \quad f \in C^2[0,1].$$

Now using the inequality $\psi(u) \le u + \sqrt{\frac{2}{\pi}}$,

$$J(b,n) \le b \left(\int_0^1 \mid \gamma(x) \mid + \mid \xi(x) \mid dx \right) + \sqrt{\frac{2}{\pi}} b^{-1/4} n^{-1/2} \int_0^1 \alpha(x) dx.$$

To find the upper bound for densities $f \in C^2[0,1]$, we take the value of b which minimizes the upper bound of $J(b,n)$, and then we obtain

$$J(b,n) \leq C^* \frac{B(f)}{n^{2/5}}$$

where $C^* = (4^{1/5} + 4^{-4/5}) \left(\sqrt{\frac{2}{\pi}}\right)^{4/5} = 5(8\pi)^{-2/5} = 1.3768102 \cdots$ is a universal constant.

■

3 Uniform Weak Convergence of the Beta Kernel Estimator.

In this section, we suppose that the density function f is continuous on $[0,1]$. We show the uniform weak convergence on $[0,1]$ of the Beta kernel estimator f_b.

Theorem 3.1 *(Uniform Weak Convergence of f_b). Let f be a probability density function on $[0,1]$, $f \in C([0,1])$, and f_b be the Beta kernel estimator. We assume that $b \longrightarrow 0$, as $n \longrightarrow \infty$, such that $nb^2 \longrightarrow \infty$.*
Then, we have

$$\sup_{x \in [0,1]} \mid f_b(x) - f(x) \mid \xrightarrow{P} 0 \quad as \quad n \longrightarrow \infty.$$

■

To prove theorem (3.1) we need the following lemma, which state the uniform convergence of the bias of the Beta kernel estimator on $[0,1]$.

Lemma 3.1 *Let f be a probability density function on $[0,1]$, $f \in C([0,1])$, and f_b be the Beta kernel estimator. We assume that $b \longrightarrow 0$, as $n \longrightarrow \infty$*
Then, we have

$$\sup_{x \in [0,1]} \mid \mathbb{E}\left(f_b(x)\right) - f(x) \mid \longrightarrow 0 \quad as \quad n \longrightarrow \infty.$$

■

Proof
We note by μ_x and σ_x^2 the mean and the variance of the random variable X where $X \sim \beta(t, x/b{+}1, (1{-}x)/b{+}1)$. From Johnson, Kotz and Balakrishnan (1994) it may be shown that there exists a constant M such that:

$$\mu_x = x + b(1-2x) + \Delta_1(x), \quad \text{and} \tag{7}$$

$$\sigma_x^2 = bx(1-x) + \Delta_2(x) \tag{8}$$

where $\Delta_j(x) \leq Mb^2$ for j=1 and 2. (we can take $M = 2$)

Since f is uniformly continuous on [0,1], for $\epsilon > 0$, there exists a $\delta > 0$ such that

$$| f(t) - f(x) | < \epsilon, \qquad \text{for} \qquad | x - t | < \delta \tag{9}$$

for all x in $[0,1]$ and from equation(7), there exists a n_0 such that

$$| \mu_x - x | < \delta/2, \qquad \text{for} \quad n \geq n_0 \quad \text{for all } x \text{ in } [0,1]. \tag{10}$$

Now,

$$\begin{aligned}
| \mathbb{E}\,(f_b(x)) - f(x) | &= \left| \int_0^1 (f(t) - f(x))\, K(t, x/b+1, (1-x)/b+1) dt \right| \\
&\leq \int_{|t-\mu_x|<\delta/2} |f(t) - f(x)|\, K(t, x/b+1, (1-x)/b+1) dt \\
&\quad + \int_{|t-\mu_x|>\delta/2} |f(t) - f(x)|\, K(t, x/b+1, (1-x)/b+1) dt \\
&\leq \int_{|t-\mu_x|<\delta/2} |f(t) - f(x)|\, K(t, x/b+1, (1-x)/b+1) dt \\
&\quad + 2 \sup_{t\in[0,1]} | f(t) | \int_{|t-\mu_x|>\delta/2} K(t, x/b+1, (1-x)/b+1) dt \\
&= I + II.
\end{aligned}$$

From equation (9) and (10) we have

$$I \leq \epsilon, \qquad \text{for all} \qquad n \geq n_0. \tag{11}$$

On the other hand, by using Tchebyshev inequality and equation (8)

$$
\begin{aligned}
II &\leq \left\{ 8 \sup_{t\in[0,1]} \mid f(t) \mid \sigma_x^2 \right\} / \delta^2 \\
&\leq \left\{ 2 \sup_{t\in[0,1]} \mid f(t) \mid \ (b + Mb^2) \right\} / \delta^2 \\
&\leq \epsilon, \qquad \text{for all} \qquad n \geq n_1. \qquad (12)
\end{aligned}
$$

By combining equations(11) and (12), we have

$$
\sup_{x\in[0,1]} \mid \mathbb{E}\left(f_b(x)\right) - f(x) \mid < 2\epsilon, \qquad \text{for all} \qquad n \geq \sup(n_0, n_1).
$$

■

Proof of theorem (3.1):

We begin with the usual inequality

$$
\begin{aligned}
\sup_{x\in[0,1]} \mid f_b(x) - f(x) \mid &\leq \sup_{x\in[0,1]} \mid f_b(x) - \mathbb{E}(f_b(x)) \mid \\
&+ \sup_{x\in[0,1]} \mid \mathbb{E}(f_b(x)) - f(x) \mid
\end{aligned}
$$

Since the second term is non-stochastic and converges to zero by lemma(3.1), we only need to prove that

$$
\sup_{x\in[0,1]} \mid f_b(x) - \mathbb{E}(f_b(x)) \mid \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty. \qquad (13)
$$

Note that,

$$
\begin{aligned}
f_b(x) &= \frac{1}{n}\sum_{i=1}^{n} K(X_i, x/b + 1, (1-x)/b + 1) \\
&= \int_0^1 K(t, x/b + 1, (1-x)/b + 1) dF_n(t),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(f_b(x)) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(K(X_i, x/b + 1, (1-x)/b + 1)) \\
&= \mathbb{E}(K(Y, x/b + 1, (1-x)/b + 1)) \\
&= \int_0^1 K(t, x/b + 1, (1-x)/b + 1) dF(t),
\end{aligned}
$$

where Y is a random variable with distribution F with an unknown probability density function f.

For all x in $(0,1)$,

$$\begin{aligned}
| f_b(x) - I\!E(f_b(x)) | &= \left|\int_0^1 K(t, x/b+1, (1-x)/b+1)\, d[F_n(t) - F(t)]\right| \\
&= \left|\int_0^1 (F_n(t) - F(t))\, dK(t, x/b+1, (1-x)/b+1)\right| \\
&\leq \int_0^1 |F_n(t) - F(t)|\, |\, dK(t, x/b+1, (1-x)/b+1)| \\
&\leq \sup_{t\in[0,1]} |F_n(t) - F(t)| \int_0^1 |dK(t, x/b+1, (1-x)/b+1)| \\
&\leq b^{-1}(1+b) \sup_{t\in[0,1]} |F_n(t) - F(t)| \\
&\quad \times \int_0^1 |K(t, x/b+1, (1-x)/b) - K(t, x/b, (1-x)/b+1)|\, dt \\
&\leq 2\, b^{-1}(1+b) \sup_{t\in[0,1]} |F_n(t) - F(t)|\, .
\end{aligned}$$

And for $x = 0$ or 1,

$$\begin{aligned}
| f_b(x) - I\!E(f_b(x)) | &= \left|\int_0^1 K(t, x/b+1, (1-x)/b+1)\, d[F_n(t) - F(t)]\right| \\
&= \left|\int_0^1 (F_n(t) - F(t))\, dK(t, x/b+1, (1-x)/b+1)\right| \\
&\leq \int_0^1 |F_n(t) - F(t)|\, |\, dK(t, x/b+1, (1-x)/b+1)| \\
&\leq \sup_{t\in[0,1]} |F_n(t) - F(t)| \int_0^1 |dK(t, x/b+1, (1-x)/b+1)| \\
&\leq b^{-1}(1+b) \sup_{t\in[0,1]} |F_n(t) - F(t)| \times \int_0^1 K(t, x/b+1, (1-x)/b)\, dt \\
&\leq b^{-1}(1+b) \sup_{t\in[0,1]} |F_n(t) - F(t)|\, .
\end{aligned}$$

Then for all x in $[0,1]$,

$$| f_b(x) - I\!E(f_b(x)) | \leq 2\, b^{-1}(1+b) \sup_{t\in[0,1]} |F_n(t) - F(t)|\, .$$

Applying the Dvoretzky-Kiefer-Wolfowitz(1956) inequality, we obtain

$$\begin{aligned} \mathbb{P}\left(\sup_{x\in[0,1]} |f_b(x) - \mathbb{E}(f_b(x))| \geq \epsilon\right) &\leq \mathbb{P}\left(\sup_{t\in[0,1]} |F_n(t) - F(t)| \geq 2\,\epsilon\, b(1+b)^{-1}\right) \\ &\leq 2\exp\left(-8\,\epsilon^2\, nb^2(1+b)^{-2}\right). \end{aligned}$$

wich concludes the proof of theorem(3.1).

■

4 Beta Kernel Estimator For Unbounded Density.

In this section, we consider a very special case of densities. We suppose that the density function f is unbounded in 0 (or 1), i.e. there are many observations concentrated in the boundary. The standard kernel estimator introduced by Rosenblatt(1975) is unsatisfactory for this class of density functions. Indeed, there is a problem of boundary bias and the density function is unbounded in the boundary. For such class of densities we consider the Beta kernel as an estimator. We remark, in Figure (1), the special shape of the Beta kernel used to estimate the density function at $x = 0$ and $x = 1$. Figure (2) shows that the Beta kernel and the true density have the same behaviour for the boundary and the interior points, even for unbounded densities. The data of size $n = 200$ are drawn from a beta density $B(\alpha, \beta)$ with $\alpha = 0.3$ and $\beta = 2$.

The main result of this section, is that the Beta kernel estimator converges in probability to infinity in $x = 0$, if the underlying density is unbounded in $x = 0$. We get the same result if f is unbounded in 1.

Theorem 4.1 . *Let f be a probability density function on $[0, 1]$, $f \in C((0, 1))$ and unbounded in 0. f_b is the Beta kernel estimator. We assume that $b \longrightarrow 0$, as $n \longrightarrow \infty$, such that $nb^2 \longrightarrow \infty$.*
Then, we have

$$f_b(0) \xrightarrow{P} +\infty \quad as \qquad n \longrightarrow \infty.$$

■

Proof
We prove in the first step that:

$$\mathbb{E}\left(f_b(0)\right) \longrightarrow +\infty \quad \text{as} \qquad n \longrightarrow \infty. \tag{14}$$

In fact, let $C > 0$, we have

$$f(t) \longrightarrow +\infty \quad \text{as} \qquad t \longrightarrow 0.$$

then there exist $\delta(2C) > 0$ such that $f(t) > 2C$ for all $0 < t < \delta < 1$.

$$\begin{aligned}
\mathbb{E}\left(f_b(0)\right) &= b^{-1}(b+1)\int_0^1 f(t)(1-t)^{1/b}\,dt \\
&\geq 2Cb^{-1}(b+1)\int_0^\delta (1-t)^{1/b}\,dt \\
&\geq 2C\left(1-(1-\delta)^{b^{-1}(b+1)}\right).
\end{aligned}$$

but we know that $b \longrightarrow 0$, then there exist n_0 such that

$$(1-\delta)^{b^{-1}(b+1)} < 1/2 \qquad \text{for all} \qquad n \geq n_0$$

Then, for $C > 0$, there exist n_0 such that

$$\mathbb{E}\left(f_b(0)\right) > C, \qquad \text{for all} \qquad n \geq n_0.$$

From the proof of theorem(3.1), we have

$$\sup_{x\in[0,1]} \mid f_b(x) - \mathbb{E}(f_b(x)) \mid \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty. \tag{15}$$

Then

$$\mid f_b(0) - \mathbb{E}(f_b(0)) \mid \xrightarrow{P} 0 \quad \text{as} \quad n \longrightarrow \infty, \tag{16}$$

wich concludes the proof of theorem(4.1).

■

5 SIMULATION RESULTS

In this section we present results of a simulations designed to investigate performance, in $L_1[0,1]$, of Beta kernel estimator for unbounded density function. For comparison purposes Bernstein polynomial estimator

$$\hat{f}_{kn}(x) = \sum_{i=0}^{k} \omega_{i,k}\beta(x, i+1, k-i+1)$$

where k is the smoothing bandwidth, The weight $\omega_{i,k}$ are defined by

$$\omega_{i,k} \;=\; F_n\left(\frac{i+1}{k+1}\right) - F_n\left(\frac{i}{k+1}\right)$$

with F_n the empirical distribution, and $\beta(x, \alpha_1, \alpha_2)$ represent the value of the Beta density in x. The Bernstein estimator $\hat{f}_1$ is then a mixture of $(k+1)$ Beta densities with random weights $\omega_{i,k}$

We consider also a new estimator $\hat{f}_{new}$ for unbounded density function

$$\hat{f}\text{new}(x) \;=\; \frac{1}{n}\sum_{i=1}^{n}\beta(x, X_i/h+1, (1-X_i)/h+1)$$

Where h is the smoothing bandwidth. The new estimator $\hat{f}_{new}$ is a mixture of n Beta density functions with random parameters $(X_i/h+1, (1-X_i)/h+1)$.

Remark 2 *Note that the new estimator is a density, i.e. integrated to one. By simulation we found that the new estimator is a good estimator, as shown in Figure (3), the data of size=200 from the beta density function* $\beta(0.3, 2)$ *and the smoothing bandwidth* $h = 0.05$.

Remark 3 *If we permute the role of* x *and* X_i, *the new estimator* $\hat{f}_{new}$ *is the Beta kernel estimator* f_b.

Two density function were considered. One is the beta density

$$f_1(x) = \frac{1}{\beta(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1).$$

Another density function is,

$$f_2(x) = \left(\frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1}e^{-\lambda x}\right)\{\gamma(1,\alpha,\lambda)\}^{-1}, \quad x \in (0,1).$$

Where $\gamma(x, \alpha, \lambda)$ is the value in x of gamma distribution with parameter (α, λ).
f_2 is the density of the truncated gamma distribution $\Gamma(\alpha, \lambda)$ random variable in $[0, 1]$.

Random samples were generated from the two distributions. We choose $\alpha = 3$ and $\beta = 0.3$ for the Beta density f_1, and $\lambda = 2$ and $\alpha = 0.3$ for the truncated gamma f_2.

The sample sizes used in the simulation ranged from 25 to 1000. For each sample size, N = 100 random samples were generated. For each simulated sample and for each estimator, the smoothings bandwidth was chosen by directly minimising the integrated absoluted Deviation, called IAD, of the density estimate produced by the estimator.
For the Beta kernel estimator

$$IAD(b) = \int_0^1 \mid \hat{f}_b(t) - f(t) \mid \, dt$$

For the Bernstein estimator

$$IAD(k) = \int_0^1 \mid \hat{f}_{kn}(t) - f(t) \mid \, dt$$

For the new estimator

$$IAD(h) = \int_0^1 \mid \hat{f}_{new}(t) - f(t) \mid \, dt.$$

By substituting the optimal bandwidth, the median integrated absoluted deviations as measures of performance for each of the estimators for the density function f_1 and the mean integrated absoluted deviations as measures of performance for each of the estimators for the density functionf_2 are calculated.

The results of simultions are in Figure (4) for f_1 and in Figure (5) for f_2. We remark in the two Figures that, for both densities f_1 and f_2, the Beta kernel estimator is better for all ranges of size of data than the Berstein estimator and the new estimator.

Acknowledgements Comments by Ingrid Van Keilegom are gratefully acknowledged; they have been most useful for the improvement of the exposition. Support from the contract"Projet d'Actions de Recherche Concertées" n^0 98/03 − 217 of the Belgian Government is also gratefully acknowledged.

REFERENCES

1. Brown, B.M and Chen, S.X (1999). *Beta-Bernstein smoothing for regression curves with compact supports*. Scand. J. Statist, **26**: 47-59.

2. Chen, S.X (1999). *A beta Kernel estimation for density functions* . Computational Statistics and Data analysis. **31**: 131-145.

3. Chen, S.X (2000). *Beta Kernel for regression curves*. Statistica Sinica **10**: 73-92.

4. Deheuvels, P. (1977). *Estimation non paramtrique de la densité par histogrammes généralisés*. Revus de Statistique Appliquée. **25**: 4-42.

5. Devroye, L.P. and Penrod, C.S. (1982). *Distribution-Free Lower Bounds in Density Estimation*. Technical Report, Applied Research Laboraories, The University of Texas at Austin.

6. Devroye, L.P . (1983a). *The equivalence of weak, strong and complete convrgence in L_1 for kernel density estimates*. Annals of Statistics. **11**, 896-904.

7. Devroye, (1983b). *On arbitrarly slow rates of global convergence in density estimation*. Z. Wahrsch. Verw. Gebiete **62**, 475-483.

8. Devroye, L. P. and Györfi, L. (1985). *Nonparametrique Density estimation: The $L_1 View$* Wiley, New York.

9. Devroye, L. P. (1987). *A course in Density estimation.* Birkhuser, Boston.

10. Dvoretzky, A., Kiefer, J. and Wolfowitz, J (1956). *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*. The annals of Mathematical Statistics, **1**: 444-453.

11. Grawronski, N, Stadtmüller, U. (1981). *Smoothing histograms by lattice and continuos distributions.* Metrika **28**: 155-164.

12. Johnson, N. L, Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. New York: Wiley.

13. Massart, P. (1990). *The tight constant in Dvoretzky-Kiefer-Woolfowitz inequality.* The annals of Probability, **18**, 1269-1283.

14. Parzen, E. (1962). *On estimation of the a probability density function and mode.* The Annals of Mathematical Statistics, **33**: 1065-1076.

15. Rosenblatt, M. (1956). *Remarks on some nonparametric estimates of a density function.* Ann. Math. Statist. **27** 832-837.

16. Scott, D. W Tapia, R.A and Thompson, J.R (1977). *Kenel density estimation revised.* Nonlinear anal. Theory Meth. Applications, **1**, 339-372.

17. Serfling, R. (1980). *Approximation Theorems and Mathematical Statistics.* Wiley, New York.

18. Silverman, B.W. (1986).*Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

19. Stadtmüller, U. (1983). *Asymptotic distribution of smoothed histograms.* Metrika **28**: 145-158.

20. Tapia, R.A. and Thompson, J.R (1978). *Nonparametric probability density estimation.* Johns Hopkins University Press. Baltimore.

21. Tenbusch, A. (1998). *Nonparametric Curve Estimation Whith Bernstein Polynomial.* Universitätsverlag Rasch, Osnabrück.

22. Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman and Hall, Lonndon.

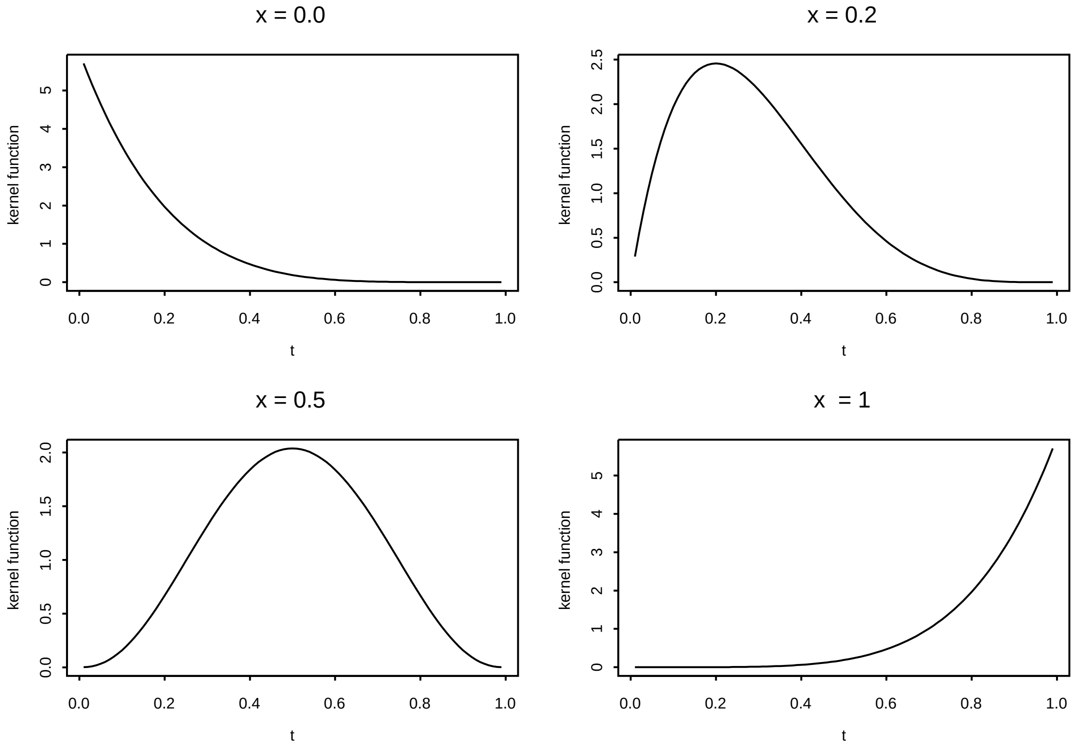

Figure 1: Beta kernels $K(t, x/b + 1, (1 - x)/b + 1)$, for $b = 0.2$.

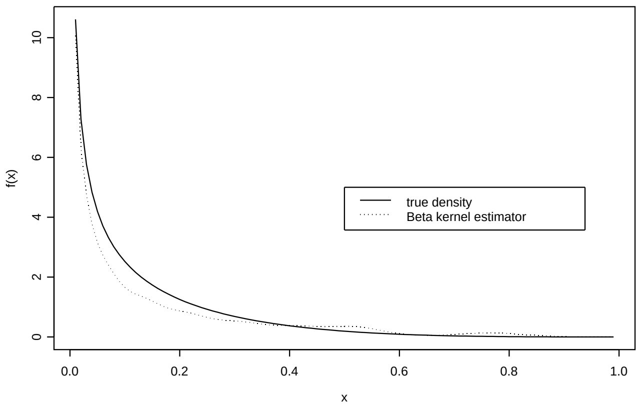

Figure 2: The Beta kernel estimator of the density $f(x) = \beta(0.2, 3)(n = 200)$, with $b = 0.6$.

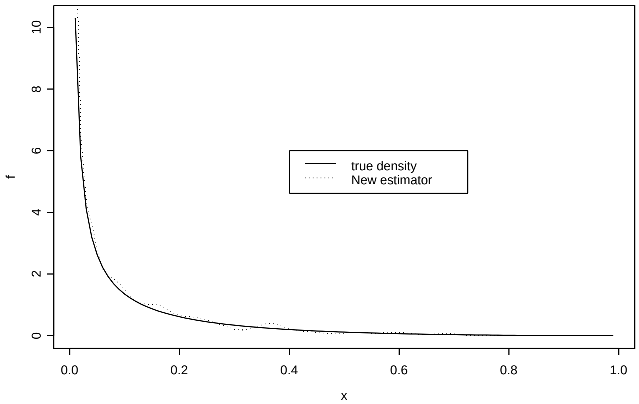

Figure 3: The New estimator of the density $f(x) = \beta(0.2, 3)(n = 200)$.

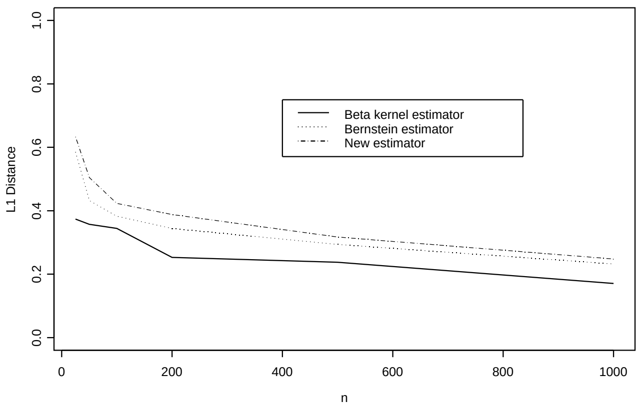

Figure 4: Comparaison in L_1 of Beta kernel estimate, Bernstein estimate and the new estimate, for different size of data. The data are from beta density function $\beta(0.3, 3)$.

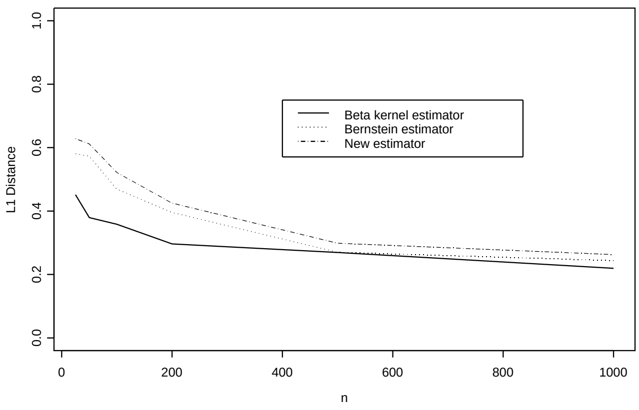

Figure 5: Comparaison in L_1 of Beta kernel estimate, Bernstein estimate and the new estimate, for different size of data. The data are from truncated gamma density function $\gamma(2, 0.3)$.