

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Minimax properties of beta kernel estimators

Karine Bertin^{a,1}, Nicolas Klutchnikoff^{b,*}^a *Departamento Estadística, CIMFAV, Universidad de Valparaíso, Avenida Gran Bretaña 1091, Playa Ancha, Valparaíso, Chile*^b *Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67000 Strasbourg, France*

ARTICLE INFO

Article history:

Received 1 March 2010

Received in revised form

6 January 2011

Accepted 13 January 2011

Available online 31 January 2011

Keywords:

Beta kernel

Density

Minimax estimation

ABSTRACT

In this paper, we are interested in the study of beta kernel density estimators from an asymptotic minimax point of view. These estimators allows to estimate density functions with support in $[0,1]$. It is well-known that beta kernel estimators are – on the contrary of classical kernel estimators – “free of boundary effect” and thus are very useful in practice. The goal of this paper is to prove that there is a price to pay: for very regular density functions or for certain losses, these estimators are not minimax. Nevertheless they are minimax for classical regularities such as regularity of order two or less than two, supposed commonly in the practice and for some classical losses.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

This paper is devoted to the study of some properties concerning beta kernel density estimators.

These estimators were introduced in [Chen \(1999\)](#) for density estimation with support in $[0,1]$. Indeed, contrary to classical kernel estimators, they are (in the classical case of the estimation of a twice differentiable density) “free of boundary effect”: their bias tends to 0 (even at points 0 and 1) and their mean integrated square error is of order $n^{-4/5}$.

This property has contributed to popularize their use in many applied fields such as economy and finance. Number of papers deal with these applications. Among others let us point out [Bouezmarni and Rolin \(2003\)](#), [Bouezmarni and Van Belleghem \(2009\)](#) and [Charpentier and Oulidi \(2010\)](#).

In this article, we adopt a quite different point of view in order to study the performance of these beta kernel estimators. We put ourselves in an asymptotic minimax framework. In particular, our study will not be restricted to twice differentiable density or to mean integrated square error.

We study here the performance of beta kernel estimators in the density model for densities belonging to Hölder spaces. We observe n independent and identically distributed (i.i.d.) variables X_1, \dots, X_n which admit the unknown density f with respect to the Lebesgue measure on $[0,1]$. Moreover, we assume that f belongs to the class of functions $\Sigma(\beta, L)$ where $\beta > 0$ is a regularity parameter and $L > 0$ is a Lipschitz constant which are assumed to be known by the statistician. The class $\Sigma(\beta, L)$ is the set of all the density functions defined on $[0,1]$ which are m times differentiable such that for all $(x, y) \in [0,1]^2$:

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x - y|^{\beta - m},$$

where $m = \sup\{\ell \in \mathbb{N} : \ell < \beta\}$.

* Corresponding author. Tel.: +33 368 850186; fax: +33 368 850328.

E-mail addresses: karine.bertin@uv.cl (K. Bertin), klutchni@math.unistra.fr (N. Klutchnikoff).

¹ Tel./fax: +56322508268.

We measure the quality of the estimators of the unknown density function f with a risk in L^p loss. More precisely, if \tilde{f}_n is an arbitrary estimator, we define its risk over $\Sigma(\beta, L)$, for $p \geq 1$, by

$$R_n(\tilde{f}_n, \beta, L) = \sup_{f \in \Sigma(\beta, L)} R_n(\tilde{f}_n, f)$$

where

$$R_n(\tilde{f}_n, f) = (E_f^n(\|\tilde{f}_n - f\|_p^p))^{1/p},$$

and E_f^n is the expectation with respect to the law of (X_1, \dots, X_n) . The minimax rate of convergence on $\Sigma(\beta, L)$ is defined as $r_n(\beta, L) = \inf_{\tilde{f}_n} R_n(\tilde{f}_n, \beta, L)$ where the infimum is taken over all the estimators. The asymptotic of $r_n(\beta, L)$ is well-known up to a constant (see Ibragimov and Has'minskiĭ, 1981) and is of order $\varphi_n(\beta) = n^{-\beta/(2\beta+1)}$. We are then interested in knowing if beta kernel estimators are optimal estimators in minimax framework, in other words, if they converge at the rate $\varphi_n(\beta)$.

The advantage of this approach with respect to the classical one (second order regularity and mean integrated square error) is that our study is quite precise concerning the understanding of the expression “free of boundary effect”. Our first result, Theorem 1, illustrates that, for regularities less or equal than 2 and for L^p losses with $p < 4$, it is possible to construct an optimal beta kernel estimator. Of course, this is linked to the fact that, even on the boundary of $[0, 1]$, the bias term tends to 0. Nevertheless, our second result, Theorem 2, shows that for higher regularities ($\beta > 2$), even if this bias term tends to 0, the order of convergence is not optimal. This leads to the impossibility to construct an optimal beta kernel estimator in that case. To conclude our study, we show, in Theorem 3, that even if the bias term is of the good order (for $\beta \leq 2$), for L^p losses such that $p \geq 4$ the variance term is not optimal. This leads again to the impossibility to construct an optimal beta kernel estimator.

Finally, let us point out that our results clarify the conditions for using beta kernel estimator by showing some intrinsic limitations of these estimators. Nevertheless they have good properties of convergence for small regularities as it is often supposed for practical purposes. This work is a first step in the study of beta kernel estimators from an asymptotic minimax point of view. There is still a lot to do in order to complete this study such as finding data-driven methods to chose the bandwidth of these estimators (cross validation methods or Lepski-type procedures).

In Section 2, we introduce our estimators and give our main results. Section 3 is devoted to the proofs.

2. Main results

We recall that we observe X_1, \dots, X_n an i.i.d. sample with density f that belongs to $\Sigma(\beta, L)$ and our goal is to estimate f from a minimax point of view.

Before stating our main results, let us recall the definition of beta kernel estimators.

Definition 1. For all $b \in (0, 1)$ and $t \in [0, 1]$, let us introduce the following density:

$$K_{t,b}(x) = \frac{x^{t/b}(1-x)^{(1-t)/b}}{B\left(\frac{t}{b} + 1, \frac{1-t}{b} + 1\right)} I_{[0,1]}(x) \tag{1}$$

where $B(\cdot, \cdot)$ is the standard beta function and I denotes the characteristic function. Following Chen (1999) let us introduce the associated beta kernel estimator:

$$\hat{f}_b(t) = \frac{1}{n} \sum_{k=1}^n K_{t,b}(X_k), \quad t \in [0, 1].$$

Remark 1. Let us notice that $(\hat{f}_b)_{b \in (0,1)}$ defines a one-parameter family of estimators. Note also that the density given by Eq. (1) corresponds to a beta distribution of parameters $t/b + 1$ and $(1-t)/b + 1$.

Fig. 1 represents some beta kernels $K_{t,b}$ drawn for different values of t and b . From left to right, b – which plays the role of a bandwidth – decreases and one can observe a concentration of the kernel in a neighborhood of t which is the mode.

We will give three results about beta kernel estimators. The first one is a positive result: for $0 < \beta \leq 2$ and $1 \leq p < 4$ there exists a beta kernel estimator with properly chosen *bandwidth* which achieves the minimax rate of convergence on $\Sigma(\beta, L)$ in L^p -loss.

Theorem 1. Set $1 \leq p < 4$ and $0 < \beta \leq 2$. Set $b_n = cn^{-2/(2\beta+1)}$, where c is a positive constant. Then the estimator \hat{f}_{b_n} achieves the rate $r_n(\beta, L)$. More precisely:

$$\limsup_{n \rightarrow \infty} \frac{R_n(\hat{f}_{b_n}, \beta, L)}{r_n(\beta, L)} < \infty.$$

The following theorem shows that, for regularities larger than 2 and most of all losses, beta kernel density estimators are not minimax.

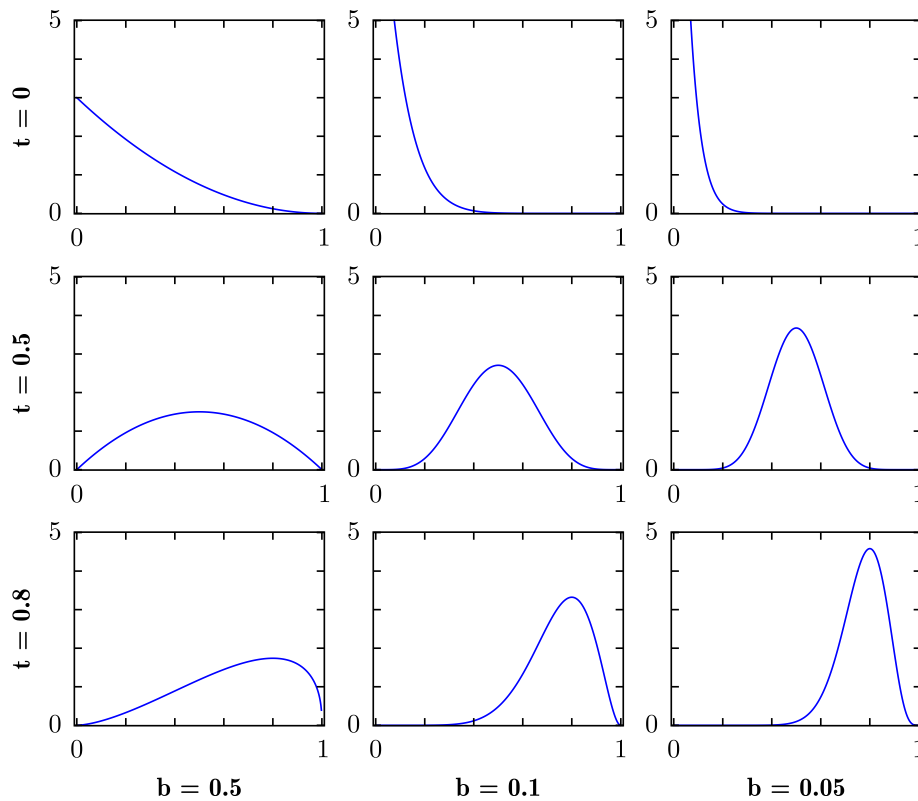


Fig. 1. Some beta kernels for different values of b and t .

Theorem 2. Set $p \geq 2$ and $\beta > 2$. Then the family of estimators \hat{f}_{b_n} satisfies, for all sequence $(b_n)_{n \in \mathbb{N}}$ in $(0,1)$,

$$\liminf_{n \rightarrow \infty} \frac{R_n(\hat{f}_{b_n}, \beta, L)}{r_n(\beta, L)} = \infty.$$

The next theorem proves, for regularities less or equal than 2, contrary to what happens in Theorem 1, that the beta kernel density estimator is not minimax if the risk is measured with L_p losses with $p \geq 4$.

Theorem 3. Set $p \geq 4$ and $0 < \beta \leq 2$. Then the family of estimators \hat{f}_{b_n} satisfies, for all sequence $(b_n)_{n \in \mathbb{N}}$ in $(0,1)$,

$$\liminf_{n \rightarrow \infty} \frac{R_n(\hat{f}_{b_n}, \beta, L)}{r_n(\beta, L)} = \infty.$$

In the framework of Theorem 1, the bias term of beta kernel estimators is of order $b^{\beta/2}$ and variance term $(nb^{1/2})^{-1/2}$. These orders are the same as for classical kernel estimators with $h=b^{1/2}$. A trade off between bias and variance leads to choose b of the form $cn^{-2/(2\beta+1)}$ and the associated beta kernel estimator converges to f at rate $\varphi_n(\beta)$. Let us remark that this theorem can be viewed as a generalization of the “classical case” which corresponds to $\beta = 2$ and $p=2$.

When $\beta > 2$ and $p \geq 2$, the impossibility to construct an optimal beta kernel estimator is linked with the construction of two functions which belong to $\Sigma(\beta, L)$. The first one has its bias term that cannot go to 0 faster than the rate b while the variance term of the second one is lower bounded by $(nb^{1/2})^{-1/2}$. This implies that beta kernel estimators cannot converge to f at a rate faster than $\varphi_n(2)$ in this framework.

In the case of Theorem 3, the non-optimality of beta kernel estimators is linked with the construction of a function, which belongs to $\Sigma(\beta, L)$, such its variance term is not of good order (by at least an extra $|\log b|$ factor). We exhibit a second function in $\Sigma(\beta, L)$ with a bias term of order $b^{\beta/2}$ that allows us to conclude that beta kernel estimators cannot converge at the minimax rate $\varphi_n(\beta)$.

Remark 2. A classical example of densities with support in $[0,1]$ is the Beta distributions. The density function of the Beta distribution of parameter α and γ is given by

$$f_{\alpha, \gamma}(x) = \frac{x^{\alpha-1}(1-x)^{\gamma-1}}{B(\alpha, \gamma)} I_{[0,1]}(x).$$

For $\alpha = 1$ and $\gamma = 1$, the density function is of regularity β for all $\beta > 0$. For $\alpha > 1$ and $\gamma > 1$, the density function is of regularity β for all $\beta \leq 1$. For $\alpha \geq 2$ and $\gamma \geq 2$, the density function is of regularity β for all $\beta \leq 2$. More generally, for $\alpha \geq p$ and $\gamma \geq p$ with $p \in \mathbb{N}$, the density function is of regularity β for all $\beta \leq p$.

Other examples are the following density functions. Let us consider $g=I_{[0.2,0.8]}$ and, for all $x \in [0,1]$ the following density functions:

$$g_1(x) = c_1 \int_0^x g(t) dt, \quad g_2(x) = c_2 \int_0^x g_1(t) dt$$

where c_1 and c_2 are renormalizing constants and

$$g_3(x) = \frac{\pi}{2} \cos(\pi x/2).$$

The function g_1 is of regularity β for all $\beta \leq 1$. The function g_2 is of regularity β for all $\beta \leq 2$. The function g_3 is of regularity β for all $\beta > 0$.

Proofs of these three theorems are given in Section 3.

3. Proofs

In all the proofs, C denotes a positive constant that can change of values from line to line. In the following, $(a_n) \asymp (b_n)$ stands for $0 < \liminf_{n \rightarrow \infty} a_n/b_n \leq \limsup_{n \rightarrow \infty} a_n/b_n < +\infty$.

3.1. Proof of Theorem 1

Firstly, set $1 \leq p < 4$, $0 < \beta \leq 2$, $L > 0$ and $f \in \Sigma(\beta, L)$. As the function f (and the number of observations n) is always fixed in this proof, we will denote, for simplicity, E instead of E_f^n . We have

$$R_n(\hat{f}_b, f) = (E(\|\hat{f}_b - f\|_p^p))^{1/p} = \left(E \left(\int_0^1 |\hat{f}_b(t) - f(t)|^p dt \right) \right)^{1/p} \leq C \left(\int_0^1 |B_t|^p dt + E \left(\int_0^1 |Z_t|^p dt \right) \right)^{1/p},$$

where $B_t = E(\hat{f}_b(t)) - f(t)$ and $Z_t = \hat{f}_b(t) - E(\hat{f}_b(t))$. Thus, as $p \geq 1$, we obtain

$$R_n(\hat{f}_b, f) \leq C \left\{ \left(\int_0^1 |B_t|^p dt \right)^{1/p} + \left(\int_0^1 E(|Z_t|^p) dt \right)^{1/p} \right\}. \tag{2}$$

The proof of our theorem will be derived from two lemmas. The first one is used to control the integrated bias term and the second one to control the integrated moment of the centered stochastic term Z_t .

Lemma 1. Set p and β as in Theorem 1. Then, for all $0 < b < 1$:

$$\left(\int_0^1 |B_t|^p dt \right)^{1/p} \leq CLb^{\beta/2}. \tag{3}$$

Lemma 2. Set p and β as in Theorem 1. Then, for all $0 < b < 1$:

$$\left(\int_0^1 E(|Z_t|^p) dt \right)^{1/p} \leq C \left(\frac{1}{nb^{1/2}} \right)^{1/2}.$$

Proofs of these lemmas are in Subsection 3.3 and 3.4. Let us complete the proof of Theorem 1. From Eq. (2) and Lemmas 1 and 2 we derive

$$R_n(\hat{f}_b, f) \leq C \left\{ Lb^{\beta/2} + \left(\frac{1}{nb^{1/2}} \right)^{1/2} \right\}.$$

As this equation is valid for all $0 < b < 1$ we deduce that, if b_n denotes the quantity $cn^{-2/(2\beta+1)} \asymp \operatorname{argmin}_{0 < b < 1} (Lb^{\beta/2} + (nb^{1/2})^{-(1/2)})$ with c a positive constant, we have

$$R_n(\hat{f}_{b_n}, f) \leq C \left\{ Lb_n^{\beta/2} + \left(\frac{1}{nb_n^{1/2}} \right)^{1/2} \right\} \leq Cr_n(\beta, L),$$

that gives the result of the theorem.

3.2. Proofs of Theorems 2 and 3

In all these proofs we will use auxiliary functions. Let us define the first one:

$$f_0(x) = I_{[0,1]}(x).$$

Next, for $0 < \beta \leq 2$, we define

$$f_\beta(x) = 1 + L_\beta \sum_{k=1}^{2N} (-1)^{k+1} \left(\frac{1}{(4N)^\beta} - \left| x - \frac{2k-1}{4N} \right|^\beta \right) I_{[(k-1)/2N, k/2N]}(x),$$

where N is the integer part of $b^{-1/2}/20$ and $L_\beta = L/2\min(1, 1/\beta)$. Fig. 2 shows the function f_β for two values of β . Finally we consider

$$f_3(x) = 2xI_{[0,1]}(x).$$

For simplicity, we will use the notation E_α instead of $E_{f_x}^n$. These functions satisfy the following lemma:

Lemma 3. For all $p \geq 2$ and $b \in (0,1)$, we have:

(i) The function f_0 belongs to $\Sigma(\beta, L)$ for all $\beta > 0$ and $L > 0$ and satisfies

$$E_0(\|\hat{f}_b - f_0\|_p^p) \geq Cd_n(b, p)(nb^{1/2})^{-p/2}, \tag{4}$$

where

$$d_n(b, p) = \int_{1/2}^{1-b} \left(\frac{1}{1-t} \right)^{p/4} dt.$$

(ii) Set $0 < \beta \leq 2$ and $L > 0$. The function f_β (which also depends on b and L) belongs to $\Sigma(\beta, L)$ and satisfies

$$E_\beta(\|\hat{f}_b - f_\beta\|_p^p) \geq Cb^{p\beta/2}, \tag{5}$$

(iii) The function f_3 belongs to $\Sigma(\beta, L)$ for all $\beta > 2$ and $L > 0$ and satisfies

$$E_3(\|\hat{f}_b - f_3\|_p^p) \geq Cb^p \tag{6}$$

Let $\beta > 2$, $L > 0$ and $p \geq 2$. We have for all $b \in (0,1)$, using Eqs. (4) and (6):

$$\sup_{f \in \Sigma(\beta, L)} R_n(\hat{f}_b, f) \geq \frac{1}{2}(R_n(\hat{f}_b, f_0) + R_n(\hat{f}_b, f_3)) \geq C \inf_{b \in (0,1)} \left(b + \left(\frac{1}{nb^{1/2}} \right)^{1/2} \right) \asymp \varphi_n(2).$$

Since $\varphi_n(\beta) = o(\varphi_n(2))$, this implies Theorem 2.

Let $0 < \beta \leq 2$, $L > 0$ and $p \geq 4$. We have for all $b \in (0,1)$:

$$\sup_{f \in \Sigma(\beta, L)} R_n(\hat{f}_b, f) \geq \frac{1}{2}(R_n(\hat{f}_b, f_0) + R_n(\hat{f}_b, f_\beta)).$$

Then we have, using Eqs. (4) and (5):

$$\sup_{f \in \Sigma(\beta, L)} R_n(\hat{f}_b, f) \geq C \left(b^{\beta/2} + \frac{d_n(b, p)}{(nb^{1/2})^{1/2}} \right) \geq C \left(b^{\beta/2} + \frac{|\log b|}{(nb^{1/2})^{1/2}} \right).$$

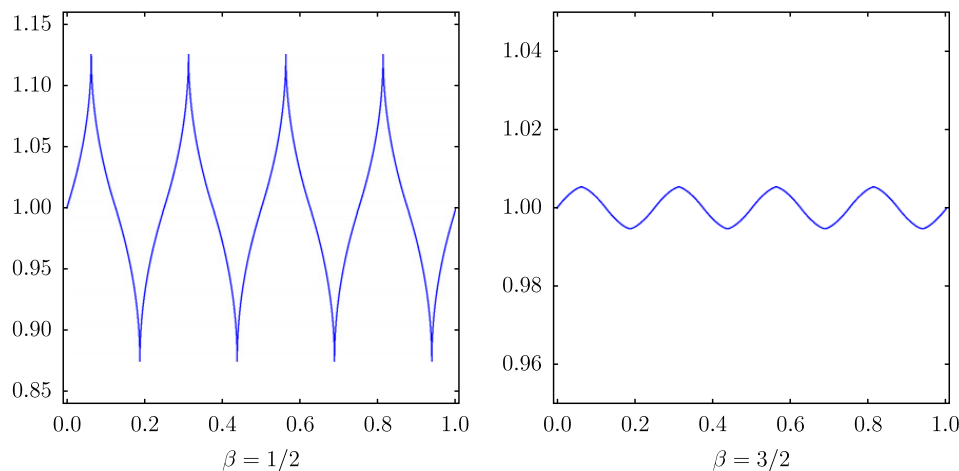


Fig. 2. Function f_β : $\beta = 1/2$ in the left side, $\beta = 3/2$ in the right side. In both cases, $L=1$ and $b=0.0005$.

The function $f(b) = b^{\beta/2} + |\log b| / (nb^{1/2})^{-1/2}$ attains its minimum b_n^0 on $(0,1)$ for n large enough. The minimum satisfies $b_n^0 = c_1 n^{-2/(2\beta+1)} (c_2 + |\log(b_n^0)|)^{4/(2\beta+1)}$ with c_1 and c_2 positive constants. It can be easily proved that the sequence b_n^0 tends to 0 as n tends to ∞ . This implies that

$$\inf_{b \in (0,1)} \left\{ b^{\beta/2} + \frac{|\log b|}{\sqrt{nb^{1/2}}} \right\} = (b_n^0)^{\beta/2} + \frac{|\log b_n^0|}{\sqrt{n(b_n^0)^{1/2}}} \asymp \varphi_n(\beta) |\log(b_n^0)|^{2\beta/(2\beta+1)}.$$

This last result implies Theorem 3.

3.3. Proof of Lemma 1

To prove Eq. (3), it is sufficient to prove that $|B_t| \leq CLb^{\beta/2}$, for all $t \in [0,1]$. In order to prove this last inequality, we will use the following result (see Johnson et al., 1994).

Lemma 4. *If ξ is a variable with density $K_{t,b}$ ($\xi \sim K_{t,b}$) then there exist two functions Δ_1 and Δ_2 , and a constant $M > 0$ such that*

$$E(\xi) - t = b(1 - 2t) + \Delta_1(t, b) \tag{7}$$

and

$$\sigma^2(\xi) = bt(1 - t) + \Delta_2(t, b), \tag{8}$$

where

$$\sup_{0 \leq t \leq 1} |\Delta_j(t, b)| \leq Mb^2 \quad \text{if } j = 1, 2. \tag{9}$$

The variable ξ introduced in this lemma appears in the study of the bias term. Indeed we have

$$B_t = E(\hat{f}_b(t)) - f(t) = E\left(\frac{1}{n} \sum_{k=1}^n K_{t,b}(X_k)\right) - f(t) = \int_0^1 K_{t,b}(x) f(x) dx - f(t) = E(f(\xi) - f(t)).$$

Now, we have to distinguish two cases depending on the position of β with respect to 1.

a. Assume that $0 < \beta \leq 1$. Since f belongs to $\Sigma(\beta, L)$, one can write

$$|B_t| \leq E(|f(\xi) - f(t)|) \leq LE(|\xi - t|^\beta) \leq LE(|\xi - t|)^\beta.$$

Last inequality holds because $\beta \leq 1$. Set $\eta > 0$.

$$|B_t| \leq LE(|\xi - t| (I_{|\xi - t| \leq \eta} + I_{|\xi - t| > \eta}))^\beta \leq L(\eta + \sqrt{\sigma^2(\xi) \cdot P(|\xi - t| > \eta)})^\beta \leq L\left(\eta + \sqrt{\sigma^2(\xi) \cdot \frac{\sigma^2(\xi)}{\eta^2}}\right)^\beta \leq L\left(\eta + \frac{Cb}{\eta}\right)^\beta.$$

Except for the last inequality which follows from Eqs. (8) and (9), all the others follow from classical probability inequalities. Optimizing in η , we obtain that $|B_t| \leq CLb^{\beta/2}$, which allows to conclude.

b. Assume that $1 < \beta \leq 2$. Using the mean-value theorem, we obtain

$$|B_t| = |E(f(\xi) - f(t))| = |E(f'(\xi)(\xi - t) + (\xi - t)(f'(\xi) - f'(t)))|,$$

where $|\xi - t| \leq |\xi - t|$. Thus we obtain:

$$|B_t| = |f'(t)(E(\xi) - t) + E((\xi - t)(f'(\xi) - f'(t)))| \leq |f'(t)| \cdot |E(\xi) - t| + E(|\xi - t| \cdot |f'(\xi) - f'(t)|) \leq A(M + 1)b + LE(|\xi - t|^\beta),$$

where $A = \sup_{f \in \Sigma(\beta, L)} \|f'\|_\infty < +\infty$ (see, for example, Tsybakov, 2004, Section 1.2.1). Remark that last inequality follows from Eq. (7).

Let us consider the term $E(|\xi - t|^\beta)$. By introducing the mean of ξ we obtain

$$E(|\xi - t|^\beta) = E(|\xi - E(\xi) + E(\xi) - t|^\beta) \leq C(E(|\xi - E(\xi)|^\beta) + |E(\xi) - t|^\beta).$$

As $1 < \beta \leq 2$, we obtain, using Hölder's inequality:

$$E(|\xi - t|^\beta) \leq C(\sigma^\beta(\xi) + |E(\xi) - t|^\beta) \leq C((Mb)^{\beta/2} + (Mb)^\beta) \leq Cb^{\beta/2}.$$

Finally, we obtain again that $|B_t| \leq CLb^{\beta/2}$ that gives the lemma in the second case.

3.4. Proof of Lemma 2

Note that $Z_t = \hat{f}_b(t) - E(\hat{f}_b(t)) = 1/n \sum_{k=1}^n \eta_k$, where the $\eta_k = K_{t,b}(X_k) - E(K_{t,b}(X_k))$'s are i.i.d. centered variables. Thanks to the following lemma (see Bretagnolle and Huber, 1979) it is possible to control precisely the p -th moment of Z_t .

Lemma 5. If y_1, \dots, y_n are n i.i.d. variables such that $E(y_1) = 0$ and $\sigma^2(y_1) = \nu$, then the following inequalities hold:

(a) If $p \leq 2$ then:

$$E\left(\left|\frac{1}{n} \sum_{k=1}^n y_k\right|^p\right) \leq \left(\frac{\nu}{n}\right)^{p/2}.$$

(2) If $p > 2$ and moreover $\|y_1\|_\infty < +\infty$, then:

$$E\left(\left|\frac{1}{n} \sum_{k=1}^n y_k\right|^p\right) \leq C_p \left(\frac{\nu \|y_1\|_\infty^{p-2}}{n^{p-1}} + \left(\frac{\nu}{n}\right)^{p/2}\right).$$

This lemma will be apply with $y_k = \eta_k$, ($k = 1, \dots, n$). Thus we have to control two terms: $\sigma^2(\eta_1)$ on the one hand and $\|\eta_1\|_\infty$ on the other hand.

a. In order to control of $\sigma^2(\eta_1)$, let us compute:

$$\sigma^2(\eta_1) \leq E((K_{t,b}(X_1))^2) = \int_0^1 \frac{x^{2t/b}(1-x)^{2(1-t)/b}}{B^2\left(\frac{t}{b}+1, \frac{1-t}{b}+1\right)} f(x) dx = A_b(t)E[f(Z)],$$

where

$$A_b(t) = \frac{B\left(2\frac{t}{b}+1, 2\frac{1-t}{b}+1\right)}{B^2\left(\frac{t}{b}+1, \frac{1-t}{b}+1\right)} \tag{10}$$

and Z has a beta distribution with parameters $2t/b+1$ and $2(1-t)/b+1$. Since f belongs to $\Sigma(\beta, L)$ and thanks to the fact that the supremum of $\|f\|_\infty$ over $\Sigma(\beta, L)$ is finite, we obtain that $\sigma^2(\eta_1) \leq CA_b(t)$. Moreover, it is known (see [Chen, 2000](#)) that for b small enough:

$$A_b(t) \leq \frac{Cb^{-1/2}}{\sqrt{t(1-t)}}.$$

Thus, for all b small enough:

$$\sigma^2(\eta_1) \leq \frac{Cb^{-1/2}}{\sqrt{t(1-t)}}.$$

b. Control of $\|\eta_1\|_\infty$. As t is the mode of $K_{t,b}$ and thanks to Stirling's formula, it can be shown (see [Chen, 2000](#)) that:

$$\sup_{x \in [0,1]} K_{t,b}(x) \leq \frac{Cb^{-1/2}}{\sqrt{t(1-t)}}.$$

Thus we obtain

$$\begin{aligned} \|\eta_1\|_\infty &= \|K_{t,b}(X_1) - E(K_{t,b}(X_1))\|_\infty \leq \|K_{t,b}(X_1)\|_\infty + |E(f(\xi))| \leq \sup_{x \in [0,1]} K_{t,b}(x) + |f(t)| + |E(f(\xi)) - f(t)| \\ &\leq \frac{Cb^{-1/2}}{\sqrt{t(1-t)}} + \sup_{f \in \Sigma(\beta, L)} \|f\|_\infty + CLb^{\beta/2}. \end{aligned}$$

Thus, for small b we have

$$\|\eta_1\|_\infty \leq \frac{Cb^{-1/2}}{\sqrt{t(1-t)}}.$$

Now let us complete the proof of this lemma. We have to distinguish two cases. The first one concerns the case where $p \leq 2$. The second one, the case where $p > 2$.

First, let us assume that $p \leq 2$. Applying Lemma 5 and using the bound on $\sigma^2(\eta_1)$ just obtained, we have

$$E(|Z_t|^p) = E\left(\left|\frac{1}{n} \sum_{k=1}^n \eta_k\right|^p\right) \leq \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2} \leq C \left(\frac{1}{nb^{1/2}}\right)^{p/2} (t(1-t))^{-p/4}.$$

Last, if $p > 2$, we have, thanks to Lemma 5:

$$E(|Z_t|^p) = E\left(\left|\frac{1}{n} \sum_{k=1}^n \eta_k\right|^p\right) \leq C_p \left\{ \frac{\sigma^2(\eta_1) \|\eta_1\|_\infty^{p-2}}{n^{p-1}} + \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2} \right\}.$$

Thanks to our bound on $\|\eta_1\|_\infty$ we obtain

$$\begin{aligned} E(|Z_t|^p) &\leq C_p \left\{ \frac{\sigma^2(\eta_1)\sigma^{2p-4}(\eta_1)}{n^{p-1}} + \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2} \right\} \leq C_p \left\{ \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p-1} + \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2} \right\} \\ &\leq C_p \left\{ \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p-1-p/2} + 1 \right\} \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2} \leq C \left(\frac{\sigma^2(\eta_1)}{n}\right)^{p/2}. \end{aligned}$$

Using our bound on $\sigma^2(\eta_1)$ we obtain

$$E(|Z_t|^p) \leq C \left(\frac{1}{nb^{1/2}}\right)^{p/2} (t(1-t))^{-p/4}.$$

Taking all together lemma follows. Indeed, for any $p < 4$ and b small enough we have

$$\left(\int_0^1 E(|Z_t|^p) dt\right)^{1/p} \leq C \left(\frac{1}{nb^{1/2}}\right)^{1/2} \left(\int_0^1 (t(1-t))^{-p/4} dt\right)^{1/p}.$$

3.5. Proof of Lemma 3

Let us consider a preliminary lemma which will be proved in Subsection 3.6.

Lemma 6. For all $p \geq 2$ and $b \in (0,1)$ and all density f , we have

$$E(\|\hat{f}_b - f\|_p^p) \geq \int_0^1 |B_t|^p dt \tag{11}$$

$$\text{and } E(\|\hat{f}_b - f\|_p^p) \geq 2^{-p} \int_0^1 (E(Z_t^2))^{p/2} dt. \tag{12}$$

Proof of (i). In order to prove Eq. (4) it is enough, thanks to Eq. (12), to lower bound $\int_0^1 (E_0(Z_t^2))^{p/2} dt$. Note that $E_0(Z_t^2)$ can be written in the following way:

$$E_0(Z_t^2) = \frac{1}{n} (E_0(\zeta_t^2) - (E_0(\zeta_t))^2) = \frac{1}{n} (E_0(\zeta_t^2) - 1)$$

where $\zeta_t = K_{t,b}(X_1)$. Thus our goal is to minorate $E_0(\zeta_t^2)$. We have

$$E_0(\zeta_t^2) = E(K_{t,b}^2(X_1)) = \int_0^1 K_{t,b}^2(x) dx = A_b(t)$$

where $A_b(t)$ is defined by Eq. (10).

Let us introduce the R -function defined as follows, for $z \geq 0$:

$$R(z) = \frac{1}{\Gamma(z+1)} \left(\frac{z}{e}\right)^z \sqrt{2\pi z}.$$

It is well-known that R is an increasing function such that $R(z) < 1$ and, $R(z) \rightarrow 1$ as $z \rightarrow +\infty$.

Following [Chen \(1999\)](#) let us write $A_b(t)$ in terms of R -functions:

$$A_b(t) = c(b) \frac{1}{\sqrt{t(1-t)}} \frac{R^2\left(\frac{t}{b}\right) R^2\left(\frac{1-t}{b}\right) R\left(\frac{2}{b}+1\right)}{R\left(\frac{2t}{b}\right) R\left(\frac{2(1-t)}{b}\right) R^2\left(\frac{1}{b}+1\right)}$$

where

$$c(b) = \frac{e^{-1}}{2\sqrt{\pi}} b \left(1 + \frac{1}{b}\right)^{3/2} \left(1 + \left(\frac{2}{b} + \frac{5}{2}\right)^{-1}\right)^{2/b+3} \geq Cb^{-1/2},$$

for b small enough. Moreover, as R is increasing we have

$$\frac{R^2\left(\frac{t}{b}\right) R^2\left(\frac{1-t}{b}\right) R\left(\frac{2}{b}+1\right)}{R\left(\frac{2t}{b}\right) R\left(\frac{2(1-t)}{b}\right) R^2\left(\frac{1}{b}+1\right)} \geq \frac{R^2\left(\frac{t}{b}\right) R^2\left(\frac{1-t}{b}\right)}{R\left(\frac{2t}{b}\right) R\left(\frac{2(1-t)}{b}\right) R\left(\frac{1}{b}+1\right)}.$$

Using the fact that $R(z) < 1$, we obtain

$$\frac{R^2\left(\frac{t}{b}\right)R^2\left(\frac{1-t}{b}\right)R^2\left(\frac{2}{b}+1\right)}{R\left(\frac{2t}{b}\right)R\left(\frac{2(1-t)}{b}\right)R^2\left(\frac{1}{b}+1\right)} \geq R^2\left(\frac{t}{b}\right)R^2\left(\frac{1-t}{b}\right).$$

Thus we have

$$A_b(t) \geq C \frac{b^{-1/2}}{\sqrt{t(1-t)}} R^2\left(\frac{t}{b}\right) R^2\left(\frac{1-t}{b}\right) \geq C \frac{b^{-1/2}}{\sqrt{t(1-t)}} R^2\left(\frac{t}{b}\right) R^2\left(\frac{1-t}{b}\right) I_{[1/2, 1-b]}(t) \geq C \frac{b^{-1/2}}{\sqrt{t(1-t)}} R^2\left(\frac{1}{2b}\right) R^2(1) I_{[1/2, 1-b]}(t).$$

For b small enough, $R^2(1/2b)$ is greater than $R^2(1)$. Hence we obtain

$$A_b(t) \geq C \frac{b^{-1/2}}{\sqrt{t(1-t)}}. \tag{13}$$

From Eq. (13) we obtain

$$E_0(\zeta_t^2) \geq C b^{-1/2} \frac{1}{\sqrt{t(1-t)}} I_{[1/2, 1-b]}(t) \geq C b^{-1/2} \frac{1}{\sqrt{1-t}} I_{[1/2, 1-b]}(t).$$

Thus, for b small enough we obtain

$$E_0(Z_t^2) \geq C \frac{1}{n b^{1/2}} \frac{1}{\sqrt{1-t}} I_{[1/2, 1-b]}(t) - \frac{1}{n}.$$

Finally, we can write

$$\int_0^1 (E_0(Z_t^2))^{p/2} dt \geq C \left(\frac{1}{n b^{1/2}}\right)^{p/2} d_n(b, p).$$

Proof of (ii). We define for $k=1, \dots, N$, $t_k = (2k-1)/(4N)$ and the following intervals: $T_k = [t_k - \varepsilon b^{1/2}, t_k + \varepsilon b^{1/2}]$, $I_k = [t_k - b^{1/2}, t_k + b^{1/2}]$ and $J_k = [t_k - 1/(4N), t_k + 1/(4N)]$ where $0 < \varepsilon < 1/2$ will be chosen. Let us recall that $1/(4N)$ is very close to $5b^{1/2}$. Note that we have the following inclusions: $T_k \subset I_k \subset J_k$.

Now, let us lower bound the integrated bias:

$$\int_0^1 |B_t| dt \geq \sum_{\ell=1}^N \int_{T_{2\ell-1}} \left| \int_0^1 K_{t,b}(x)(f_\beta(t) - f_\beta(x)) dx \right| dt \geq \sum_{\ell=1}^N \int_{T_{2\ell-1}} (A_\ell(t) - B(t)) dt$$

where

$$A_\ell(t) = \int_{J_{2\ell-1} \cap I_{2\ell-1}^c} K_{t,b}(x)(f_\beta(t) - f_\beta(x)) dx$$

and

$$B(t) = \int_{\{f_\beta(x) \geq f_\beta(t)\}} K_{t,b}(x)(f_\beta(x) - f_\beta(t)) dx.$$

Since for $t \in T_{2\ell-1}$ and $x \in J_{2\ell-1} \cap I_{2\ell-1}^c$ we have

$$f_\beta(t) - f_\beta(x) = L_\beta(|x - t|^\beta - |t - t|^\beta) \geq L_\beta(b^{\beta/2} - \varepsilon^\beta b^{\beta/2}) \geq C b^{\beta/2},$$

it follows that

$$A_\ell(t) \geq C b^{\beta/2} \int_{J_\ell \cap I_\ell^c} K_{t,b}(x) dx = C b^{\beta/2} P(\xi \in J_{2\ell-1} \cap I_{2\ell-1}^c),$$

where $\xi \sim K_{t,b}$. This probability can be estimated:

$$P(\xi \in J_{2\ell-1} \cap I_{2\ell-1}^c) \geq P(t_{2\ell-1} + b^{1/2} < \xi < t_{2\ell-1} + 5b^{1/2}) \geq P(t_{2\ell-1} - t + t + b^{1/2} < \xi < t_{2\ell-1} - t + t + 5b^{1/2}) \\ \geq P(t + 2b^{1/2} < \xi < t + 4b^{1/2}) \geq 2b^{1/2} K_{t,b}(t + 4b^{1/2})$$

For $1/4 \leq t \leq 3/4$ we have

$$K_{t,b}(t + 4b^{1/2}) \underset{b \rightarrow 0}{\sim} K_{t,b}(t) \exp\left(-\frac{8}{t(1-t)}\right) \geq C b^{-1/2}.$$

Indeed, using the R function we obtain

$$K_{t,b}(t) = \frac{R\left(\frac{t}{b}\right)R\left(\frac{1-t}{b}\right)\sqrt{b}(1+1/b)}{R(1/b)\sqrt{2\pi t(1-t)}} \geq C b^{-1/2}$$

for $1/4 \leq t \leq 3/4$ and b small enough. Thus, for ℓ such that $T_{2\ell-1} \subset [1/4, 3/4]$, $t \in T_{2\ell-1}$ and for b small enough, we have

$$A_\ell(t) \geq Cb^{\beta/2},$$

where C is an absolute constant independent of b and ε .

For $t \in T_{2\ell-1}$ we have

$$B(t) = \int_{\bigcup_{k=1}^N \{|x-t_{2k-1}| \leq |t-t_{2\ell-1}|\}} K_{t,b}(x)(f_\beta(x)-f_\beta(t)) dx \leq \sum_{k=1}^N \int_{\{|x-t_{2k-1}| \leq |t-t_{2\ell-1}|\}} K_{t,b}(x)L_\beta(|t-t_{2\ell-1}|^\beta - |x-t_{2k-1}|^\beta) dx \leq C\varepsilon^\beta b^{\beta/2}.$$

Taking all together we obtain the following lower bound:

$$\int_0^1 |B_t| dt \geq \sum_{\ell=1}^N \int_{T_{2\ell-1}} A_\ell(t) dt - \sum_{\ell=1}^N \int_{T_{2\ell-1}} B(t) dt \geq CN\varepsilon b^{1/2} b^{\beta/2} - N\varepsilon^{\beta+1} b^{1/2} b^{\beta/2} \geq b^{\beta/2} (c_1 \varepsilon - c_2 \varepsilon^{\beta+1})$$

where c_1 and c_2 are absolute positive constants. Thus, for $\varepsilon = \min(1/2, (c_1/c_2)^{1/\beta}/2)$ we obtain

$$\int_0^1 |B_t| dt \geq Cb^{\beta/2},$$

and thus, using the convexity of $x \mapsto x^p$,

$$\int_0^1 |B_t|^p dt \geq Cb^{\beta p/2}.$$

Using Eq. (11), this achieves the proof of (ii).

Proof of (iii). For b small enough, we have

$$B_t = E_3(K_{t,b}(X_1)) - f_3(t) = E(f_3(\xi)) - f_3(t) = 2(E(\xi) - t) \geq Cb(1-2t),$$

where ξ is defined as in Lemma 4 and last inequality is a consequence of Eq. (7). As $|1-2t|^p$ is integrable on $[0,1]$, it follows:

$$\int_0^1 |B_t|^p dt \geq Cb^p.$$

Using Eq. (11) we deduce Eq. (6).

3.6. Proof of Lemma 6

Since $x \mapsto |x|^p$ is a convex function

$$|B_t|^p = |E(\hat{f}_b(t) - f(t))|^p \leq E(|\hat{f}_b(t) - f(t)|^p),$$

which implies Eq. (11) taking the integral on $[0,1]$ in both sides of the last inequality.

Since $\|Z_t\|_p \leq \|B_t\|_p + \|\hat{f}_b - f\|_p$, we deduce, using Eq. (11) that, for all function f ,

$$E(\|Z_t\|_p^p) \leq 2^{p-1} E(\|B_t\|_p^p + \|\hat{f}_b - f\|_p^p) \leq 2^p E(\|\hat{f}_b - f\|_p^p).$$

Then, we have

$$E(\|\hat{f}_b - f\|_p^p) \geq 2^{-p} E(\|Z_t\|_p^p) = 2^{-p} E\left(\int_0^1 |Z_t|^p dt\right) = 2^{-p} \int_0^1 E(|Z_t|^p) dt \geq 2^{-p} \int_0^1 (E(Z_t^2))^{p/2} dt.$$

This proves (Eq. (12)).

Acknowledgements

Karine Bertin and Nicolas Klutchnikoff are supported by Project Fondecyt 1090285. Karine Bertin has been supported by project Laboratory ANESTOC PBCT ACT 13.

References

Bouezmarni, T., Rolin, J.-M., 2003. Consistency of the beta kernel density function estimator. *Canad. J. Statist.* 31 (1), 89–98. doi:10.2307/3315905
 ISSN 0319-5724. URL <http://dx.doi.org/10.2307/3315905>.
 Bouezmarni, T., Van Bellegem, S., 2009. Nonparametric beta kernel estimator for long memory time series. Technical Report.
 Bretagnolle, J., Huber, C., 1979. Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* 47 (2), 119–137. doi:10.1007/BF00535278
 ISSN 0044-3719. URL <http://dx.doi.org/10.1007/BF00535278>.
 Charpentier, A., Oulidi, A., 2010. Beta kernel quantile estimators of heavy-tailed loss distributions. *Statist. Comput.* 20 (1), 35–55. doi:10.1007/s11222-009-9114-2.
 Chen, S.X., 1999. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* 31 (2), 131–145. doi:10.1016/S0167-9473(99)00010-9
 ISSN 0167-9473. URL <http://dx.doi.org/10.1016/S0167-9473(99)00010-9>.

- Chen, S.X., 2000. Beta kernel smoothers for regression curves. *Statist. Sinica* 10 (1), 73–91 ISSN 1017-0405.
- Ibragimov, I.A., Has'minskiĭ, R.Z., 1981. *Statistical estimation. Applications of Mathematics*, vol. 16. Springer-Verlag, New York ISBN 0-387-90523-5. Asymptotic theory, Translated from the Russian by Samuel Kotz.
- Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. *Continuous univariate distributions. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*, second edition, vol. 1. John Wiley & Sons Inc., New York ISBN 0-471-58495-9. A Wiley-Interscience Publication.
- Tsybakov, A.B., 2004. *Introduction à l'estimation non-paramétrique. Mathématiques & Applications (Berlin) [Mathematics & Applications]*, vol. 41. Springer-Verlag, Berlin ISBN 3-540-40592-5.