Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 04–06

ALMOST OPTIMAL CONVERGENCE RATES FOR KERNEL DENSITY ESTIMATION UNDER ASSOCIATION

CARLA HENRIQUES AND PAULO EDUARDO OLIVEIRA

ABSTRACT: Exponential inequalities for associated variables are derived under an assumption milder than the absolute continuity of joint distributions of the sample variables. This inequality is used to prove convergence rates for the kernel estimator for the density which are just slightly slower than the optimal rates known form independent samples.

KEYWORDS: kernel estimation, association, exponential rates, diagonal decomposition.
AMS SUBJECT CLASSIFICATION (2000): 62G07, 62G20.

1. Introduction

The convergence rate of the kernel estimator for the density function based on associated samples has been addressed in some recent articles. Exponential rates were announced in Dewan, Prakasa Rao [2], but their result assumes a convergence rate on the covariance structure that is unattainable under association. These authors used an approximation to independence technique. With this same technique and the blocking decomposition as in Ioannides, Roussas [9], the above mentioned problem was corrected in Henriques, Oliveira [7]. The rates that follow from the results in this later reference are significantly slower than the best known for independent samples. More recently Masry [11] using a quite different approach obtained convergence rates that are just slightly slower than those for independent samples. His method is based on Rosenthal's inequalities and requires the absolute continuity of joint distributions of random vectors of all dimensions with their margins being sample variables. The present article improves the results in Henriques, Oliveira [7], using the same approximation to independence and the blocking decomposition, assuming some particular behaviour of the the joint distributions of pairs of the sample variables, allowing for same mass to be concentrated in the diagonal of $I\!R \times I\!R$. This assumption was used in estimation problems in point processes in Jacob, Oliveira [10] under independence, and extended to the treatment of associated samples in

Both authors supported by Centro de Matemática da Universidade de Coimbra.

the same context in Ferrieux [4, 5] and for the study of the kernel density estimator under association on Oliveira [13].

2. Definitions and assumptions

Let $X_1, X_2, \dots$ be random variables with the same distribution as X for which there exists a density function f. Let K be a fixed probability density and h_n a sequence on nonnegative real numbers converging to zero. The kernel estimator of the density function f is, as usual, defined as

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right),$$

which is well known to be asymptotically unbiased, if there exists a bounded and continuous version of the density. Moreover, the convergence of $\mathbb{E}[\widehat{f}_n(x)]$ to $f(x)$ is, under these assumptions on f, uniform on compact sets.

The random variables $X_1, X_2, \dots$ will be supposed associated, which means, as defined in Esary, Proschan, Walkup [3], that for every $n \in \mathbb{N}$ and $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ coordinatewise increasing

$$\mathrm{Cov}\Big(f(X_1, \dots, X_n),\ g(X_1, \dots, X_n)\Big) \geq 0,$$

whenever this covariance exists.

A technical problem arises when dealing with $\widehat{f}_n(x)$ for associated variables. In fact, association is only preserved under monotone transformations, which means that, in general, the variables $K\left(\frac{x-X_1}{h_n}\right), K\left(\frac{x-X_2}{h_n}\right), \dots$ are not associated. One way to resolve this problem is supposing the kernel K to be of bounded variation. We way also control this drawback of the association structure with a different kind of assumption on the kernel function K as it was done in Ferrieux [4, 5].

For easier future reference, we introduce now a set of assumptions.

(A1) $X_1, X_2, \dots$ are strictly stationary and associated random variables with common bounded and continuous density function f; let $B_0 = \sup_{x \in \mathbb{R}} |f(x)|$;

(A2) The kernel function K is a probability density such that $\int K^2(u)\, du < \infty$; further, either it is of bounded variation and, if $K = K_1 - K_2$ where K_1 and K_2 are nondecreasing functions, the derivatives K_1' and K_2' exist and are integrable;

(A3) Let λ^* represent the Lebesgue measure on the diagonal Δ of $I\!R \times I\!R$ and λ^2 represents the Lebesgue measure on $I\!R^2$; then

- for each $j \in I\!N$, the distribution of (X_1, X_j) is the sum of a measure $m_{1,j}$ on $I\!R \times I\!R \setminus \Delta$ with a measure $m_{2,j}$ on Δ, such that $m_{1,j} \ll \lambda^2$ and $m_{2,j} \ll \lambda^*$;

- for each $j \in I\!N$, there exist a bounded version $g_{1,j}$ of $\frac{dm_{1,j}}{d\lambda^2}$ and a bounded and continuous version $g_{2,j}$ of $\frac{dm_{2,j}}{d\lambda^*}$;

- $\sum_{j=1}^{\infty} |g_{1,j}(x, y) - f(x)f(y)|$ converges uniformly to a bounded function g_1;

- $\sum_{j=1}^{\infty} g_{2,j}$ converges uniformly to a bounded and continuous function g_2.

In Ferrieux [4], instead of the bounded variation assumption it is supposed that the kernel function K is bounded and, for each $x \in I\!R$, the function $\lambda \longmapsto K(\lambda x)$ is decreasing for $\lambda > 0$, and K' is integrable. The usual kernel functions satisfy this assumption. Each of the following results may be restated under this alternative assumption.

The assumption (A3) allows the random pairs (X_1, X_j) to have some mass concentrated on the diagonal, thus no absolute continuity with respect to λ^2 is required. Distributions satisfying this kind of diagonal decomposition appear, for example, when using a simple method to construct sequences of associated variables from an independent sequence: let $W_1, W_2, \ldots$ be independent and identically distributed absolutely continuous variables, $m \in I\!N$ some fixed integer, and define $X_n = \min(W_n, \ldots, W_{n+m})$. The sequence of variables $X_1, X_2, \ldots$ is associated and the distribution of random pairs satisfies a diagonal decomposition as described in the first two items of (A3).

Under (A1), we have that

$$\sup_{u,v \in I\!R} \left| F_{X_1,X_j}(u, v) - F_{X_1}(u)F_{X_1}(v) \right| \leq B_2 \operatorname{Cov}^{1/3}(X_1, X_j) \tag{2.1}$$

where F_{X_1,X_j} and F_{X_1} represent the distribution functions of (X_1, X_j) and X_1, respectively, and $B_2 = \max(4/\pi^2, 90B_0)$ (see Lemma 2.6 in Roussas [15] for details). This inequality provides an upper bound for the covariances between the variables $K_q\left(\frac{x-X_j}{h_n}\right)$, $q = 1, 2$, $j = 1, 2, \ldots$.

Lemma 2.1. *Suppose the variables* $X_1, X_2, \ldots$ *satisfy* (A1) *and the kernel function satisfy* (A2). *Then*

$$\begin{aligned}\operatorname{Cov}\left(K_q\left(\frac{x-X_1}{h_n}\right), K_q\left(\frac{x-X_j}{h_n}\right)\right) &\leq \\ &\leq B_2 \operatorname{Cov}^{1/3}(X_1, X_j)\left(\int H'(u)\, du\right)^2, \quad q=1,2.\end{aligned}$$

and analogously for K_2.

Proof: Apply integration by parts and use (2.1). ∎

The next result is the basis of the development used to establish the exponential inequality as it provides a mean to control the approximation to independence. It appears under the present form in Dewan, Parkasa Rao [2] and is a version for generating functions of Newman's [12] inequality for characteristic functions.

Lemma 2.2. *Let* $X_1, \ldots, X_n$ *be associated random variables bounded by a constant* M. *Then, for every* $\theta > 0$,

$$\left| I\!\!E\left(e^{\theta\sum_{i=1}^n X_i}\right) - \prod_{i=1}^n I\!\!E\left(e^{\theta X_i}\right)\right| \leq \theta^2 e^{n\theta M} \sum_{1\leq i<j\leq n} \operatorname{Cov}(X_i, X_j).$$

This inequality was used by Dewan, Prakasa Rao [2] to control the distance between the joint distribution of the variables and what one would find in case of independence. The way they manipulated the upper bound conducted them to an assumption that is too strong for associated variables. We refer the reader to Henriques, Oliveira [7] for details.

Now we introduce the notation that will be used in the sequel. Given (A2) define

$$\widehat{f}_{n,1}(x) = \frac{1}{nh_n}\sum_{j=1}^n K_1\left(\frac{x-X_j}{h_n}\right), \qquad \widehat{f}_{n,2}(x) = \frac{1}{nh_n}\sum_{j=1}^n K_2\left(\frac{x-X_j}{h_n}\right),$$

so that $\widehat{f}_n(x) = \widehat{f}_{n,1}(x) - \widehat{f}_{n,2}(x)$. For each $n \in I\!\!N$, $j = 1, \ldots, n$, and $q = 1, 2$, let

$$T_{n,q,j} = \frac{1}{h_n}\left(K_q\left(\frac{x-X_j}{h_n}\right) - I\!\!E\, K_q\left(\frac{x-X_j}{h_n}\right)\right). \tag{2.2}$$

Note that these variables are associated if the $X_1, X_2, \ldots$ are associated, as they are nonincreasing transformations of these variables.

Given a natural number $p < \frac{n}{2}$, let r be the greatest natural number less or equal to $\frac{n}{2p}$ and define, for each $j = 1, \ldots, 2r$, and $q = 1, 2$

$$Y_{n,q,j} = \sum_{l=(j-1)p+1}^{jp} T_{n,q,l}\,. \tag{2.3}$$

Note that, if the kernel$K$ satisfies (A2), the functions K_1 and K_2 may be chosen bounded so that each variable $Y_{n,q,j}$ is bounded by $\frac{2p\|K_q\|_\infty}{h_n}$, where $\|\cdot\|_\infty$ represents the supremum norm.

Finally define, for $q = 1, 2$,

$$\begin{aligned} Z_{n,q}^{od} &= Y_{n,q,1} + Y_{n,q,3} + \cdots + Y_{n,q,2r-1}, \\ Z_{n,q}^{ev} &= Y_{n,q,2} + Y_{n,q,4} + \cdots + Y_{n,q,2r}, \end{aligned} \tag{2.4}$$

and analogously without the subscript q. With these definitions, whenever $n = 2pr$ we have $\widehat{f}_{n,q}(x) - I\!E[\widehat{f}_{n,q}(x)] = \frac{1}{n}(Z_{n,q}^{ev} + Z_{n,q}^{od})$.

The assumption (A3) provides control on the variance terms mentioned above, as proved in Lemma 2.2 of Oliveira [13].

Lemma 2.3 (Oliveira [13]). *Suppose* (A1) *and* (A3) *are satisfied. If the kernel K is bounded then, for each $x \in I\!R$,*

$$nh_n \operatorname{Var}(\widehat{f}_n(x)) \longrightarrow g_2(x,x) \int K^2(u)\, du.$$

Further, the convergence is uniform on compact sets.

Notice that, for the variables we have introduced, this means that

$$\frac{h_n}{p_n} \operatorname{Var}(Y_{n,q,1}) \longrightarrow c_q(x) = g_2(x,x) \int K_q^2(u)\, du, \quad q = 1, 2.$$

Define now, for each $x \in I\!R$, $c(x) = c_1(x) + c_2(x)$.

3. Main results

In this section we first prove some preparatory lemmas that pave the way to the proof of the main results.

The next lemma states how we may control both terms $I\!E\left(e^{\frac{\lambda}{n} Z_{n,q}^{od}}\right)$ and $I\!E\left(e^{\frac{\lambda}{n} Z_{n,q}^{ev}}\right)$ with respect to the independent case.

Lemma 3.1. *Suppose* (A1) *and* (A2) *are satisfied. With the definitions made before, for every* $\lambda > 0$,

$$\left| I\!\!E\left(e^{\frac{\lambda}{n} Z_{n,q}^{od}}\right) - \prod_{j=1}^{r} I\!\!E\left(e^{\frac{\lambda}{n} Y_{n,q,2j-1}}\right)\right| \leq$$

$$\leq \frac{\lambda^2}{2n} \exp\left(\frac{\lambda \| K_q \|_\infty}{h_n}\right) \sum_{j=p+2}^{(2r-1)p} \mathrm{Cov}(T_{n,q,1}, T_{n,q,j}), \quad q = 1, 2, \tag{3.1}$$

and analogously for the term corresponding to $Z_{n,q}^{ev}$.

Proof: As $Z_{n,q}^{od} = \sum_{j=1}^{r} Y_{n,q,2j-1}$ and the summands are bounded, we may apply Lemma 2.2. Then use the stationarity twice and the fact that $2pr \leq n$. ∎

The preceding results considered the integers p and r fixed. Let us now make the choices of these integers dependent on n, thus obtaining sequences p_n and r_n such that $\frac{n}{2p_n r_n} \longrightarrow 1$. We show now how (A3) may be used to control the independent like term.

Lemma 3.2. *Suppose* (A1), (A2) *and* (A3) *are satisfied. Given a sequence of positive numbers* c_n, *it follows, for* $q = 1, 2$ *and every* $\lambda \in \left(0, \frac{n h_n c_n}{2p_n \|K_q\|_\infty}\right)$, *that*

$$I\!\!E\left(e^{\frac{\lambda}{n} Y_{n,q,j}}\right) \leq \exp\left(\frac{\lambda^2 c_n^*}{n^2} \mathrm{Var}(Y_{n,q,1})\right),$$

where $c_n^* = \sum_{k=2}^{\infty} \frac{c_n^{k-2}}{k!}$.

Proof: As for every $n \in I\!\!N$, $q = 1, 2, j = 1, \ldots, r_n$, $I\!\!E(Y_{n,q,j}) = 0$, it follows, using a Taylor expansion, that

$$I\!\!E\left(e^{\frac{\lambda}{n} Y_{n,q,j}}\right) = 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{n^k k!} I\!\!E(Y_{n,q,j}^k).$$

As the kernel K is bounded, we have, for each $k \geq 2$,

$$\left|Y_{n,q,j}^k\right| \leq \left(\frac{2p_n \| K_q \|_\infty}{h_n}\right)^{k-2} Y_{n,q,j}^2,$$

so, inserting this in the Taylor expansion, it follows that, using also the stationarity of the variables,

$$\mathbb{E}\left(e^{\frac{\lambda}{n}Y_{n,q,j}}\right) \leq 1+\left(\frac{\lambda}{n}\right)^2 \operatorname{Var}(Y_{n,q,1}) \sum_{k=2}^{\infty} \frac{1}{k!}\left(\frac{2p_n\|K_q\|_\infty \lambda}{nh_n}\right)^{k-2} \leq$$

$$\leq 1+\left(\frac{\lambda}{n}\right)^2 \operatorname{Var}(Y_{n,q,1}) c_n^* \leq \exp\left(\frac{\lambda^2 c_n^*}{n^2}\operatorname{Var}(Y_{n,q,1})\right).$$

■

Notice that there is an evident relation between c_n^* and c_n: $c_n^* \leq e^{c_n}$. This will be useful to make some explicit construction of upper bounds.

This lemma provides immediately control of the independent like terms: under the assumptions of Lemma 3.2 we have

$$\begin{aligned} &\prod_{j=1}^{r} \mathbb{E}\left(e^{\frac{\lambda}{n}Y_{n,q,2j-1}}\right) \leq \exp\left(\frac{\lambda^2 c_n^* r_n}{n^2}\operatorname{Var}(Y_{n,q,1})\right), \quad q=1,2, \\ &\prod_{j=1}^{r} \mathbb{E}\left(e^{\frac{\lambda}{n}Y_{n,q,2j}}\right) \leq \exp\left(\frac{\lambda^2 c_n^* r_n}{n^2}\operatorname{Var}(Y_{n,q,1})\right), \quad q=1,2. \end{aligned} \tag{3.2}$$

We are now in position to prove exponential rates for the convergence in probability of the centered estimator. We will need to optimize the choice of λ where the generating functions is to be computed.

Lemma 3.3. *Suppose* (A1), (A2) *and* (A3) *are satisfied. If* $\sup_{n\in\mathbb{N}} \frac{p_n}{c_n c_n^*} < M$ *and*

$$\frac{nh_n^2}{c_n^{*2}} \exp\left(\frac{n}{c_n^*}\right) \sum_{j=p_n+2}^{\infty} \operatorname{Cov}(T_{n,q,1}, T_{n,q,j}) \leq C_0 < \infty, \tag{3.3}$$

then, for every $\varepsilon \in \left(0, \frac{1}{4M}\min\left(c_1(x), c_2(x), \frac{c_1(x)}{\|K_1\|_\infty}, \frac{c_2(x)}{\|K_2\|_\infty}\right)\right)$,

$$\mathrm{P}\left(\frac{1}{n}\left|Z_{n,q}^{od}\right| > \varepsilon\right) \leq 2(1+C_0)\exp\left(-\frac{\varepsilon^2 nh_n}{4c_n^* c_q(x)}\right), \quad q=1,2,$$

and analogously for $Z_{n,q}^{ev}$.

Proof: Markov's inequality and (3.2) lead to

$$\mathrm{P}\left(\frac{1}{n}Z^{od}_{n,q}>\varepsilon\right)\leq$$
$$\leq \exp\left(\frac{\lambda^2 c_n^* r_n}{n^2}\operatorname{Var}(Y_{n,q,1})-\lambda\varepsilon\right)+e^{-\lambda\varepsilon}\left|I\!\!E\left(e^{\frac{\lambda}{n}Z^{od}_{n,q}}\right)-\prod_{j=1}^{r}I\!\!E\left(e^{\frac{\lambda}{n}Y_{n,q,2j-1}}\right)\right|.$$

The optimization of the first term leads to $\lambda=\frac{\varepsilon n^2}{2c_n^* r_n \operatorname{Var}(Y_{n,q,1})}\sim\frac{\varepsilon n p_n}{c_n^*\operatorname{Var}(Y_{n,q,1})}$. In Lemma 3.2 there is an upper bound for the choice of λ, which means that $\frac{\varepsilon p_n}{c_n^*\operatorname{Var}(Y_{n,q,1})}\leq\frac{h_n c_n}{2p_n\|K_q\|_\infty}$, or equivalently, $\varepsilon\frac{p_n}{c_n c_n^*}\leq\frac{1}{2\|K_q\|_\infty}\frac{h_n}{p_n}\operatorname{Var}(Y_{n,q,1})$. For ε small enough this is verified, since $\sup_{n\in I\!\!N}\frac{p_n}{c_n c_n^*}<M$ and, according to Lemma 2.3, $\frac{h_n}{p_n}\operatorname{Var}(Y_{n,q,1})$ is convergent, so the result follows. ∎

In order to state the main result we have to deal with the terms in $\widehat{f}_{n,q}(x)$ that are not in $\frac{1}{n}(Z^{od}_{n,q}+Z^{ev}_{n,q})$. But these are, as expected, negligible. For easier reference, define $R_{n,q}=\widehat{f}_{n,q}(x)-I\!\!E[\widehat{f}_{n,q}(x)]-\frac{1}{n}(Z^{od}_{n,q}+Z^{ev}_{n,q})$, $q=1,2$.

Lemma 3.4. *Suppose* (A1), (A2) *are satisfied and*

$$\frac{nh_n}{p_n}\longrightarrow+\infty. \tag{3.4}$$

Then, for n large enough and every $\varepsilon>0$, $\mathrm{P}(|R_{n,q}|>\varepsilon)=0$, $q=1,2$.

Proof: Write $R_{n,q}=\frac{1}{n}\sum_{j=2r_np_n+1}^{n}T_{n,q,j}$. As the functions K_q, $q=1,2$, are bounded, it follows that $R_{n,q}$ is bounded by $2\frac{n-2p_nr_n}{nh_n}\|K_q\|_\infty\leq\frac{4p_n\|K_q\|_\infty}{nh_n}$ according to the construction of the sequences p_n and r_n. It follows from (3.4) that eventually this upper bound becomes less than ε, so the lemma is proved. ∎

Now, the main work has been completed. It remains to collect the various partial results in order to obtain the exponential rate for the kernel estimator centered at its mean.

Theorem 3.5. *Suppose* (A1), (A2), (A3) *and (3.4) are satisfied. Further assume that* $\sup_{n\in I\!\!N}\frac{p_n}{c_n c_n^*}<M$ *and*

$$\frac{n}{c_n^{*\,2}}\exp\left(\frac{n}{c_n^*}\right)\sum_{k=p_n+2}^{\infty}\operatorname{Cov}^{1/3}(X_1,X_j)\leq C_1<\infty\,, \tag{3.5}$$

then, for every $\varepsilon \in \left(0, \frac{3}{2M} \min\left(c_1(x), c_2(x), \frac{c_1(x)}{\|K_1\|_\infty}, \frac{c_2(x)}{\|K_2\|_\infty}\right)\right)$ *and* n *large enough,*

$$\mathrm{P}\left(\left|\widehat{f}_n(x) - \mathbb{E}[\widehat{f}_n(x)]\right| > \varepsilon\right) \leq D \exp\left(-\frac{\varepsilon^2 n h_n}{144 c_n^* c(x)}\right), \tag{3.6}$$

where

$$C = \max(\| K_1 \|_\infty, \| K_2 \|_\infty),$$

$$D = 4\left[2 + B_2 C_1 \left(\left(\int K_1'(u)du\right)^2 + \left(\int K_2'(u)du\right)^2\right)\right],$$

and $c(x)$ *is defined at the end of Section 2.*

Proof: We have

$$\begin{aligned}
&\widehat{f}_n(x) - \mathbb{E}[\widehat{f}_n(x)] = \\
&\quad = \left(\widehat{f}_{n,1}(x) - \mathbb{E}[\widehat{f}_{n,1}(x)]\right) + \left(\widehat{f}_{n,2}(x) - \mathbb{E}[\widehat{f}_{n,2}(x)]\right) = \\
&\quad = \frac{1}{n}\left(Z_{n,1}^{od} + Z_{n,1}^{ev}\right) + \frac{1}{n}\left(Z_{n,2}^{od} + Z_{n,2}^{ev}\right) + R_{n,1} + R_{n,2}.
\end{aligned}$$

According to Lemma 3.4, for n large enough we have that $\mathrm{P}\left(|R_{n,1}| > \frac{\varepsilon}{6}\right) = \mathrm{P}\left(|R_{n,2}| > \frac{\varepsilon}{6}\right) = 0$, so we need to concentrate only on the first terms. In order to apply Lemma 3.3 we must check that (3.3) is verified. For this purpose notice that, according to Lemma 2.1,

$$\begin{aligned}
&\mathrm{Cov}(T_{n,q,1}, T_{n,q,j}) = \\
&\quad = \frac{1}{h_n^2} \mathrm{Cov}\left(K_q\left(\frac{x - X_1}{h_n}\right), K_q\left(\frac{x - X_j}{h_n}\right)\right) \leq \\
&\quad \leq \frac{1}{h_n^2} B_2 \left(\int K_q'(u)du\right)^2 \mathrm{Cov}^{1/3}(X_1, X_j).
\end{aligned}$$

As K_1' and K_2' are assumed integrable, it follows that (3.5) implies (3.3). Applying then Lemma 3.3, we find, for $q = 1, 2$,

$$\mathrm{P}\left(\frac{1}{n}\left|Z_{n,q}^{od}\right| > \frac{\varepsilon}{6}\right) \leq 2\left(1 + B_2 C_1 \left(\int K_q'(u)du\right)^2\right) \exp\left(-\frac{\varepsilon^2 n h_n}{144 c_n^* c(x)}\right),$$

and analogously for $Z_{n,q}^{ev}$, from where the result follows. ∎

Strengthening the conditions on the kernel we may use a decomposition of a compact interval to derive an uniform exponential rate for the centered estimator. Let, on what follows, $[a,b]$ be a fixed interval and decompose $[a,b]=\cup_{j=1}^{s_n}[z_{n,j}-t_n,z_{n,j}+t_n]$ into s_n intervals of length $2t_n$. Then, obviously,

$$\begin{aligned}\sup_{x\in[a,b]}\left|\widehat{f}_n(x)-\mathbb{E}[\widehat{f}_n(x)]\right| &\leq \\ &\leq \max_{1\leq j\leq s_n}\left|\widehat{f}_n(z_{n,j})-\mathbb{E}[\widehat{f}_n(z_{n,j})]\right| + \\ &+ \max_{1\leq j\leq s_n}\sup_{x\in[z_{n,j}-t_n,z_{n,j}+t_n]}\left|\widehat{f}_n(x)-\widehat{f}_n(z_{n,j})-\mathbb{E}[\widehat{f}_n(x)-\widehat{f}_n(z_{n,j})]\right|.\end{aligned}$$

If we suppose the kernel K to be Lipschitzian, it follows that there exists a constant $\theta>0$ such that

$$\left|\widehat{f}_n(x)-\widehat{f}_n(z_{n,j})-\mathbb{E}[\widehat{f}_n(x)-\widehat{f}_n(z_{n,j})]\right|\leq 2\,\frac{\theta\,|x-z_{n,j}|}{h_n^2}\leq\frac{2\theta t_n}{h_n^2}.$$

A correct choice of the sequence of radii will verify $\frac{t_n}{h_n^2}\longrightarrow 0$. Supposing this condition satisfied it follows then that

$$\begin{aligned}&\mathrm{P}\left(\sup_{x\in[a,b]}\left|\widehat{f}_n(x)-\mathbb{E}[\widehat{f}_n(x)]\right|>\varepsilon\right)\\ &\leq \mathrm{P}\left(\max_{1\leq j\leq s_n}\left|\widehat{f}_n(z_{n,j})-\mathbb{E}[\widehat{f}_n(z_{n,j})]\right|>\varepsilon-\frac{2\theta t_n}{h_n^2}\right)\leq\\ &\leq s_n\max_{1\leq j\leq s_n}\mathrm{P}\left(\left|\widehat{f}_n(z_{n,j})-\mathbb{E}[\widehat{f}_n(z_{n,j})]\right|>\frac{\varepsilon}{2}\right).\end{aligned}$$

Thus we may state a result summarizing what was derived before.

Theorem 3.6. *Suppose that* (A1), (A2), (A3), *(3.4), (3.5) are satisfied,* $\sup_{n\in\mathbb{N}}\frac{p_n}{c_nc_n^*}<M$, *the kernel* K *is Lipschtzian and* $\frac{t_n}{h_n^2}\longrightarrow 0$. *Let* C *and* D *be defined as in Theorem 3.5. Then, for* n *large enough, each interval* $[a,b]$ *and every* $\varepsilon\in\left(0,\frac{3}{2M}\min\left(c_1(x),c_2(x),\frac{c_1(x)}{\|K_1\|_\infty},\frac{c_2(x)}{\|K_2\|_\infty}\right)\right)$,

$$\mathrm{P}\left(\sup_{x\in[a,b]}\left|\widehat{f}_n(x)-\mathbb{E}[\widehat{f}_n(x)]\right|>\varepsilon\right)\leq D\,\frac{b-a}{2t_n}\exp\left(-\frac{\varepsilon^2nh_n}{576c_n^*c(x)}\right). \quad (3.7)$$

4. Some examples and convergence rates

In the preceding we derived some sufficient conditions allowing the proof of exponential inequalities for the kernel estimator for the density. We will show that our assumptions on the sequences h_n and p_n may be fulfilled and that the convergence rates that follow are just slightly slower than those known for independent samples. These compare with the results proved by Masry [11] for geometrically decreasing covariances. For the case of polynomially decreasing covariances our inequalities seem to be not strong enough.

4.1. Geometric decreasing covariances. Suppose that $\mathrm{Cov}(X_1, X_n) = \rho_0\rho^n$ for some $\rho_0 > 0$ and $\rho \in (0, 1)$, so that

$$\sum_{j=p_n+2}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_j) \sim \rho^{\frac{p_n}{3}}.$$

Let us first check that the assumptions we used are attainable. The term in (3.5) behaves like

$$\frac{n}{c_n^{*\,2}} \exp\left(\frac{n}{c_n^*} + \frac{p_n}{3}\log\rho\right) = \exp\left(\frac{n}{c_n^*} + \frac{p_n}{3}\log\rho + 2\log\frac{\sqrt{n}}{c_n^*}\right) \leq C_0. \qquad (4.1)$$

Proposition 4.1. *Suppose that* (A1), (A2), (A3), *(3.4) are satisfied and* $\mathrm{Cov}(X_1, X_j) = \rho_0\rho^n$ *for some* $\rho > 0$ *and* $\rho \in (0, 1)$. *If* $\sup_{n\in\mathbb{N}} \frac{p_n}{c_n c_n^*} < M$, $\sup_{n\in\mathbb{N}} \frac{n}{p_n c_n^*} < M'$ *and* $\rho \in (0, e^{-9M'})$ *then (3.6) holds. Moreover, if the kernel* K *is Lipschitzian and* $\frac{t_n}{h_n^2} \longrightarrow 0$ *the inequality (3.7) also holds.*

Proof: In order to the exponent in (4.1) to be bounded we must have $\log\rho \leq \frac{3A}{p_n} - \frac{3n}{p_n c_n^*} - \frac{6}{p_n}\log\frac{\sqrt{n}}{c_n^*}$. Since, $\frac{1}{p_n}\log\frac{\sqrt{n}}{c_n^*} \leq \frac{n}{p_n c_n^*}$, it is enough that $\sup_{n\in\mathbb{N}} \frac{n}{p_n c_n^*} < M'$ and $\log\rho \leq -9M'$. ■

It is easy to construct sequences p_n, h_n and c_n that fulfil the assumptions of the last proposition: for example, $p_n = n^\delta$, $h_n = n^{-\beta}$ and $c_n = \gamma\log n$, with δ, β, $\gamma \in (0, 1)$ such that $\delta < \gamma$, $\beta + \gamma < 1 < \delta + \gamma$.

In order to derive an almost sure convergence rate we will not use (3.5) directly. Indeed, in Theorem 3.5 we tried to make an assumption that would be independent of the ε appearing in the inequality. This means some loss of power when trying to optimize the choice of such an ε. To take full advantage of the method we state the conditions we would find in Lemma 3.3 and Theorem 3.5. It is easily seen that (3.5) may be replaced by

$$\frac{\varepsilon^2 n}{c_n^{*2}} \exp\left(\frac{\varepsilon n}{c_n^*}\right) \sum_{j=p_n+2}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_j) \le C_1 < \infty. \tag{4.2}$$

To find a convergence rate we allow ε to depend on n according to

$$\varepsilon_n^2 = \frac{\alpha c_n^* \log n}{n h_n}, \tag{4.3}$$

where $\alpha > 0$ should be conveniently chosen to obtain a convergent series. With this choice and the geometric decrease rate of the covariances, (4.2) will follow if we assume that

$$\left(\frac{c_n^* \log n}{n h_n}\right)^{1/2} \frac{n}{c_n^*} + \frac{p_n}{3} \log \rho$$

is bounded. This leads to the choice of the sequence p_n with growth of order $\left(\frac{n \log n}{c_n^* h_n}\right)^{1/2}$. The characterization of the convergence rate depends now on the behaviour of c_n^*, which must be done taking into account the choice and the upper bound assumed on λ in Lemma 3.3: $\frac{2p_n \|K_q\|_\infty \lambda}{n h_n} \le c_n$ and $\lambda \sim \frac{\varepsilon_n n p_n}{c_n^* \mathrm{Var}(Y_{n,q,1})}$. With the choices made for ε_n and p_n we have $\frac{p_n \lambda}{n h_n} \sim \frac{\log n}{h_n c_n^*}$, remembering that, according to Lemma 2.3, $\frac{h_n}{p_n} \mathrm{Var}(Y_{n,q,1})$ is convergent. So to satisfy the upper bound on λ we need to have $h_n \ge \frac{\log n}{c_n c_n^*}$, which means that we must choose the sequences c_n and c_n^* such that $c_n c_n^* \longrightarrow +\infty$ sufficiently fast to allow $h_n \longrightarrow 0$. If we choose $c_n^* \sim n^\delta$ for some $\delta \in (0,1)$ (and $c_n = \delta \log n$), which meets the needs just stated, it will follow that the almost sure convergence rate for the kernel estimator is $\frac{\log^{1/2} n}{n^{(1-\delta)/2} h_n^{1/2}}$, that is, somewhat slower than the best known rate for the independent case. Notice that we did not assume any absolute continuity of any random vector with margins chosen from the sample variables. There is though a price to pay: the bandwidth sequence h_n must not decrease to zero to fast, but the restriction is not to strong as it allows for choices of the form $h_n \sim n^{-a}$, $a > 0$, which include the known optimal rates for the independent case.

4.2. Polynomial decreasing covariances. Suppose now that the covariances decrease at a polynomial rate, that is, $\mathrm{Cov}(X_1, X_n) = A n^{-a}$, for some

$A > 0$ and $a > 3$. Then

$$\sum_{j=p_n+2}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_j) \sim p_n^{\frac{3-a}{3}} .$$

Inserting the behaviour of the series into (3.5) we find a term like

$$\frac{n}{c_n^{*2}} \exp\left(\frac{n}{c_n^*} - a_1 \log p_n\right) , \tag{4.4}$$

where $a_1 = \frac{a-3}{3}$, as before. If this term is to be bounded we need that, for some $c_1 > 0$, $c_n^* \geq c_1 \frac{n}{\log p_n}$, that is c_n^* should be chosen so that is goes to infinity fast enough. Remember that the upper bound assumed on λ in Lemma 3.3 leads to $p_n \leq M c_n c_n^*$, for some $M > 0$, but this is satisfied if p_n is chosen such that $p_n \log p_n \leq c_1 n$ for some $c_1 > 0$. Thus, for polynomially decreasing covariances we have, at least, exponential convergence rate for the convergence in probability, according to (3.6) and (3.7).

As what concerns almost sure convergence rates for the kernel estimator our exponential inequalities are not strong enough. With ε_n chosen as (4.3), the boundedness of (4.4) and the upper bound on λ implies that $c_n^* \geq c_1 \frac{n}{h_n \log n}$, for some $c_1 > 0$, which means that ε_n does not converge to zero and so it is not possible to derive convergence rates.

References

[1] L. Devroye (1991). *Exponential inequalities in nonparametric estimation.* In: *Nonparametric Functional Estimation and Related Topics*, G. Roussas, ed., Kluwer Academic Publishers, Dordrecht, 31–44.

[2] I. Dewan, B. L. S. Prakasa Rao (1999). A general method of density estimation for associated random variables. *J. Nonparametr. Statist.* 10, 405–420.

[3] J. D. Esary, F. Proschan, D. W. Walkup (1967). Association of random variables, with applications. *Ann. Math. Statist.* 38, 1466–1474.

[4] D. Ferrieux (1996). *Estimation de densités de mesures moyennes de processus ponctuels associés*. PhD. Thesis, University of Montpellier II.

[5] D. Ferrieux (1998). Estimation à noyau de densités moyennes de mesures aléatoires associées. *Compte Rendus Acad. Sciences de Paris* 326, Série I, 1131–1134.

[6] C. Henriques, P. E. Oliveira (2002). Convergence rates for the estimation of two-dimensional distribution functions under association and estimation of the covariance of the limit empirical process. Preprint. *Pré-Publicações do Departamento de Matemática da Universidade de Coimbra* 02–05.

[7] C. Henriques, P. E. Oliveira (1993). Exponential rates for kernel density estimation under association. Preprint. *Pré-Publicações do Departamento de Matemática da Universidade de Coimbra*, 02–23.

[8] C. Henriques, P. E. Oliveira (2003). Estimation of a two-dimensional distribution function under association. *J. Statist. Planning Inf.* 113, 137–150.

[9] D. A. Ioannides, G. G. Roussas (1998). Exponencial inequality for associated random variables. *Statist. Probab. Lett.* 42, 423–431.

[10] P. Jacob, P. E. Oliveira (1997). Kernel estimators of general Radon-Nikodym derivatives. *Statistics* 30, 25–46.

[11] E. Masry (2002). Mulitvariate probability density estimation for associated processes: strong consistency and rates. *Statist. Probab. Letters* 58, 205–219.

[12] C. Newman (1984). *Asymptotic independence and limit theorems for positively and negatively dependent random variables*. In Y. L. Tong (ed) Inequalities in Statistics and Probability, IMS Lecture Notes - Monograph Series Vol. 5, 127–140.

[13] P. E. Oliveira (2002). Density estimation for associated sampling: a point process influenced approach. *J. Nonparametr. Statist.* 14, 495–509.

[14] G. G. Roussas (1991). Kernel estimates under association: strong uniform consistency. *Statist. Probab. Lett.* 12, 393–403.

[15] G. G. Roussas (1995). Asymptotic normality of a smooth estimate of a random field distribution function under association. *Statist. Probab. Letters* 24, 77–90.

Carla Henriques
Escola Superior de Tecnologia de Viseu, 3504-510 Viseu, Portugal

Paulo Eduardo Oliveira
Dep. Matemática, Univ. Coimbra, Apartado 3008, 3001 - 454 Coimbra, Portugal

E-mail address: `paulo@mat.uc.pt`

URL: `http://www.mat.uc.pt`