Density estimation for associated sampling: a point process influenced approach

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

Abstract

Let X_n, $n \in \mathbb{N}$, be a sequence of associated variables with common density function. We study the kernel estimation of this density, based on the given sequence of variables. Sufficient conditions are given for the consistency and asymptotic normality of the kernel estimator. The assumptions made require that the distribution of pairs (X_i, X_j) decompose as the sum of an absolutely continuous measure with another measure concentrated on the diagonal of $\mathbb{R} \times \mathbb{R}$ satisfying a further absolute continuity with respect to the Lebesgue measure on this diagonal. For the convergence in probability we find the usual convergence rate on the bandwidth, whereas for the almost sure convergence we need to require that the bandwidth does not decrease to fast and that the kernel is of bounded variation. This assumption on the kernel is also required for the asymptotic normality, together with a slightly strengthened version of the usual decrease rate on the bandwidth. The assumption of bounded variation on the kernel is needed as a consequence of the dependence structure we are dealing with, as association is only preserved by monotone transformations.

1 Introduction

The estimation of the density of random variables or random vectors is a quite classical statistical problem for which there exists a well established theory and a very wide literature. There are several methods proposed and we will be here interested in the nonparametric approach. Naturally, the problem was first studied supposing that the data collected satisfied an independence assumption. A nice account of results and existing literature is given in Bosq and Lecoutre [6]. Eventually the same problem was treated for depend sampling, considering mixing samples as in Roussas [18], Pham and Tran [17] or Masry [14] (see the monograph Bosq [5] for existing literature and a presentation of the asymptotic properties of the kernel estimator). More recently there was some interest in considering data satisfying a positive dependence assumption, namely supposing the data to be associated. For this dependence structure Bagai and Prakasa Rao [1, 2] and Roussas [19] established the consistency of the kernel estimator for the density while Dewan and Prakasa Rao [7] proved an uniform exponential rate of convergence supposing a convenient decrease rate on the covariances of the variables. This last reference deals with a general estimator that includes both the kernel and the histogram estimator, among others, although the conditions imposed essentially exclude the histogram from their results. The asymptotic normality, for the kernel estimator, was established in Roussas [20]. We note that in Roussas [20], besides some more or less usual conditions on the covariance structure of the data (which is a natural way of controlling the dependence for

*This author was supported by Centro de Matemática da Universidade de Coimbra

associated variables, as it follows from Theorem 10 in Newman [15]), it is supposed that the pairs of random variables are absolutely continuous, while in this article we will suppose the distribution of these pairs of variables to have some mass concentrated on the diagonal of $\mathbb{R} \times \mathbb{R}$. This property appears in simple models of associated random variables.

The assumptions we will make are suggested by a point process approach to estimation problems. This approach consisted on defining convenient point processes or random measures ξ and η with mean measures such that $\mathbb{E}\xi \ll \mathbb{E}\eta$ and then estimate the Radon-Nikodym derivative $\frac{d\mathbb{E}\xi}{d\mathbb{E}\eta}$. This Radon-Nikodym derivative may be conveniently interpreted representing, for a suitable choice of ξ and η, the density we want to estimate. We may estimate this Radon-Nikodym derivative using histograms or kernels and this correspond to the usual estimators defined in the classical framework. A point process approach to density estimation was first used in Ellis [8] in a context that was slightly different from ours. For an approach corresponding to the one that has directly influenced this article there exists some literature available: for independent sampling, Jacob and Oliveira [11] considered the histogram, while Jacob and Oliveira [12] considered the kernel; for strong mixing sampling, Bensaïd and Fabre [3] studied the kernel and Bensaïd and Oliveira [4] the histogram; for associated sampling, Ferrieux [10] treated the kernel estimator and Jacob and Oliveira [13] the histogram. In all these references it is supposed that one samples the point processes themselves, that is, a sample is a collection of possible realizations of the point processes. This is, in fact, equivalent to the usual sampling in density estimation, as well as in other models that may be included in this point process approach, as explained in those references. We refer the interested reader to Jacob and Oliveira [11, 12, 13] for more details on how the point process approach models some more classical estimation problems. Now, while independence or strong mixing of the sampled point processes corresponds to independence or strong mixing of the underlying variables (the ones that define the points), the same is not true for association. In fact, as noted by Ferrieux [10] and Jacob and Oliveira [13] association of point processes seems to have no relation with association of the underlying variables, thus the results in Ferrieux [10] or Jacob and Oliveira [13] do not overlap with those in Bagai and Prakasa Rao [1, 2], Roussas [19, 20], Dewan and Prakasa Rao [7]. To be a little more precise, a random variable is always associated with itself while the point process with just one point defined by this random variable is associated (as point process) with itself only in degenerate cases. This is due to the fact that the order defined in the space of discrete measures is only a partial order. Thus, the idea of approaching the same problem as in Jacob and Oliveira [13], but supposing instead that the points are associated. Several proofs in Jacob and Oliveira [13] are no longer valid in the new context. Moreover, the assumptions used there make little sense in the framework we are placing ourselves throughout this article. When dealing with point processes we are lead to the control of second order moment measures, which are defined on the product of the base space by itself, and it becomes natural to suppose that those measures may decomposed as the sum of an absolutely continuous measure with a measure which is concentrated on the diagonal of the product space. When regarding this decomposition in terms of the distribution of a pair of random variables this means that there is some probability concentrated on the diagonal, so there is no absolute continuity with respect to the product measure if this measure is, for example, diffuse. The need for such a decomposition appeared even when dealing with independent sampling of point processes. In fact, this sort of decomposition was first used in Jacob and Oliveira [11] and later adapted to dependent sampling in Ferrieux [10], Bensaïd and Fabre [3], Jacob and Oliveira [13] and Bensaïd and Oliveira [4]. The conditions we will use here are suggested by this point process approach but we will try to keep the language on the usual level, thus avoiding the use of point processes as much as possible. Further, this article will concentrate on the kernel estimator. The same approached may be used to find conditions for similar results for the histogram.

2 Definitions, assumptions and preliminaries

Let $X_1, X_2, \ldots$ be an associated sequence of 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 of real numbers converging to zero. The kernel estimator of the density function f is defined as

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

This estimator is well known to be asymptotically unbiased if there exists a suitable version of the density, as a consequence of the Bochner's Lemma.

LEMMA 2.1. *If f is bounded and continuous then $I\!E(\widehat{f}_n(x)) \longrightarrow f(x)$ uniformly on any compact set.*

This means also that, in order to establish the convergence of $\widehat{f}_n(x)$, it is enough to prove that $\widehat{f}_n(x) - I\!E\widehat{f}_n(x) \longrightarrow 0$ in the appropriate mode of convergence.

Let us now present the assumptions we will suppose throughout this article. On what follows λ^* represents the Lebesgue measure on the diagonal Δ of $I\!R \times I\!R$, and λ^2 represents the Lebesgue measure on $I\!R^2$.

(D)

- For each $j, k \in I\!N$, the distribution of (X_j, X_k) is the sum of a measure $m_{1,j,k}$ on $I\!R \times I\!R \setminus \Delta$ with a measure $m_{2,j,k}$ on Δ, such that $m_{1,j,k} \ll \lambda^2$ and $m_{2,j,k} \ll \lambda^*$;
- For each $j, k \in I\!N$, there exists a bounded version $g_{1,j,k}$ of $\frac{dm_{1,j,k}}{d\lambda^2}$;
- For each $j, k \in I\!N$, there exists a bounded and continuous version $g_{2,j,k}$ of $\frac{dm_{2,j,k}}{d\lambda^*}$.

Note that we allow the pairs to have some mass concentrated on the diagonal, thus they need not to be absolutely continuous with respect to λ^2, as it was supposed in Roussas [20]. A simple example where this property is required arises from a method used to construct sequences of associated variables from an independent sequence of variables, as described next. Suppose that Z_n, $n \in I\!N$, are independent and identically distributed absolutely continuous random variables. Let m be some fixed integer and define, for each $n \in I\!N$, $X_n = \min(Z_n, \cdots, Z_{n+m})$. The variables X_n, $n \in I\!N$, are associated, have a common absolutely continuous distribution but the random pairs (X_n, X_{n+j}), $j = 0, \ldots, m$, are not absolutely continuous. Nevertheless their distribution decomposes according to **(D)**.

In order to prove the convergence results we will need some control on Radon-Nikodym derivatives introduced in the previous condition.

Suppose **(D)** is satisfied and

(A)

- $\frac{1}{n} \sum_{j,k=1}^{n} |g_{1,j,k}(x,y) - f(x)f(y)|$ converges uniformly to a bounded function g_1;
- $\frac{1}{n} \sum_{j,k=1}^{n} g_{2,j,k}(x,x)$ converges uniformly to a bounded and continuous function g_2.

Remark that, if the variables $X_1, X_2, \ldots$ are strictly stationary, it is easy to check that

$$\frac{1}{n}\sum_{j,k=1}^{n} |g_{1,j,k} - f \otimes f| \leq \sum_{j=0}^{n-1} |g_{1,1,j} - f \otimes f|$$

and analogously for the other summation in **(A)**. Then, as in Bensaïd and Fabre [3] or Bensaïd and Oliveira [4], it is enough to assume the convergence of the upper bounds.

The main tool for proving convergence is stated in the following lemma.

LEMMA 2.2. *Suppose **(D)** and **(A)** are satisfied. If the kernel K is bounded then*

$$\frac{1}{nh_n}\sum_{j,k=1}^{n} \mathrm{Cov}\left(K\left(\frac{x-X_j}{h_n}\right), K\left(\frac{x-X_k}{h_n}\right)\right) \longrightarrow g_2(x,x)\int K^2(u)\,du$$

uniformly on any compact set.

Proof. Write

$$\frac{1}{h_n}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_n}\right), K\left(\frac{x-X_k}{h_n}\right)\right) = \tag{1}$$

$$= \frac{1}{h_n}\int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right)\Big(g_{1,j,k}(u,v) - f(u)f(v)\Big)\,dudv + \frac{1}{h_n}\int_{\Delta} K^2\left(\frac{x-u}{h_n}\right) g_{2,j,k}\,du.$$

Now the first integral in (1) is bounded above by

$$\int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) |g_{1,j,k}(u,v) - f(u)f(v)|\,dudv,$$

so that, for $n \in \mathbb{N}$ large enough

$$\frac{1}{nh_n}\sum_{j,k=1}^{n} \int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right)\Big(g_{1,j,k}(u,v) - f(u)f(v)\Big)\,dudv \leq$$

$$\leq 2h_n \frac{1}{h_n^2}\int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) g_1(u,v)\,dudv \leq 2h_n \sup_{u,v}|g_1(u,v)| \longrightarrow 0,$$

as $h_n \longrightarrow 0$. For the second integral in (1), it leads to

$$\frac{1}{h_n}\int_{\Delta} K^2\left(\frac{x-u}{h_n}\right)\frac{1}{n}\sum_{j,k=1}^{n} g_{2,j,k}(u,u)\,du \longrightarrow g_2(x,x)\int K^2(u)\,du$$

using the Bochner's Lemma after renormalizing K^2 to find a density.

□

Note that the limit obtained in this lemma is exactly the diagonal density. So, if we had supposed the absolute continuity of the random pairs (X_j, X_k) this limit would have been zero. As this term appears when dealing with the convergence of the estimator, if it converges to zero most of the convergences would follow with relaxed assumptions on the bandwidth sequence h_n.

3 Convergence of the estimator

In this section we state conditions for convergence in probability and almost surely of $\widehat{f_n}(x)$.

THEOREM 3.1. *Suppose* ***(D)*** *and* ***(A)*** *are satisfied. If the kernel K is bounded and*

$$nh_n \longrightarrow +\infty \tag{2}$$

then $\widehat{f_n}(x)$ converges in probability to $f(x)$, for every $x \in I\!R$.

Proof. As

$$I\!P\left(\left|\widehat{f_n}(x) - I\!E(\widehat{f_n}(x))\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2 n h_n} \frac{1}{n h_n} \sum_{j,k=1}^{n} \mathrm{Cov}\left(K\left(\frac{x-X_j}{h_n}\right), K\left(\frac{x-X_k}{h_n}\right)\right)$$

the theorem follows immediately taking account of Lemmas 2.1 and 2.2.

□

The almost sure convergence requires some further assumptions on the kernel K and on the bandwidth sequence. This is due to the fact that the variables $K\left(\frac{x-X_j}{h_n}\right)$ are, in general, not associated. In order to keep the association we should apply only monotone transformations to the original variables, which is not the case with a general kernel. This may resolved supposing the kernel K to be of bounded variation. There are some additional technical assumptions required by the method of proof, which follows closely the proof of Theorem 2.7.1 in Ferrieux [9].

THEOREM 3.2. *Suppose* ***(D)*** *and* ***(A)*** *are satisfied, K is of bounded variation such that $K(\alpha x)$, for fixed $x \in I\!R$, is a decreasing function of $\alpha > 0$, and*

$$h_n \searrow 0, \qquad \sum_{n=1}^{\infty} \frac{1}{n^2 h_{n^2}} < \infty, \qquad \frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1.$$

Then $\widehat{f_n}(x)$ converges almost surely to $f(x)$, for every $x \in I\!R$.

Proof. We first check that the subsequence corresponding to terms of order n^2 converges almost surely. This is immediate consequence of the Borel-Cantelli Lemma, Lemma 2.2 and the assumptions made, as

$$I\!P\left(\left|\widehat{f_{n^2}}(x) - I\!E(\widehat{f_{n^2}}(x))\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2 n^2 h_{n^2}} \frac{1}{n^2 h_{n^2}} \sum_{j,k=1}^{n^2} \mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right), K\left(\frac{x-X_k}{h_{n^2}}\right)\right).$$

For the remaining terms we write, for an integer $k \in (n^2, (n+1)^2]$,

$$\left|\left(\widehat{f_k}(x) - I\!E\widehat{f_k}(x)\right) - \left(\widehat{f}_{n^2}(x) - I\!E\widehat{f}_{n^2}(x)\right)\right| \leq$$

$$\leq \frac{1}{n^2} \max_{n^2 < k \leq (n+1)^2} \left| \frac{1}{h_k} \sum_{j=1}^{k} \left[K\left(\frac{x-X_j}{h_k}\right) - I\!E\, K\left(\frac{x-X_j}{h_k}\right)\right] \right.$$

$$\left. - \frac{1}{h_{n^2}} \sum_{j=1}^{n^2} \left[K\left(\frac{x-X_j}{h_{n^2}}\right) - I\!E\, K\left(\frac{x-X_j}{h_{n^2}}\right)\right] \right| +$$

$$+ \frac{1}{n^2 h_{n^2}} \left| \sum_{j=1}^{n^2} \left[K\left(\frac{x-X_j}{h_{n^2}}\right) - I\!E\, K\left(\frac{x-X_j}{h_{n^2}}\right)\right] \right|,$$

thus leaving the first term to be treated. Now

$$\frac{1}{h_k}\sum_{j=1}^{k}\left[K\left(\frac{x-X_j}{h_k}\right)-I\!E\,K\left(\frac{x-X_j}{h_k}\right)\right]-\frac{1}{h_{n^2}}\sum_{j=1}^{n^2}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]=$$

$$=\sum_{j=1}^{k}\left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_k}\right)-I\!E\,K\left(\frac{x-X_j}{h_k}\right)\right]+$$

$$+\sum_{j=1}^{k}\frac{1}{h_{n^2}}\left(\left[K\left(\frac{x-X_j}{h_k}\right)-I\!E\,K\left(\frac{x-X_j}{h_k}\right)\right]-\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]\right)+$$

$$+\sum_{j=n^2+1}^{k}\frac{1}{h_{n^2}}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right].$$

Let us denote the maximum over $k\in(n^2,(n+1)^2]$ of each of the three preceding terms by a_n, b_n and c_n, respectively. The almost sure convergence to zero of $\frac{a_n}{n^2}$, $\frac{b_n}{n^2}$ and $\frac{c_n}{n^2}$, that concludes the proof of the theorem, is proved in the following lemmas.

□

LEMMA 3.3. *Under the assumptions of Theorem 3.2,* $\dfrac{a_n}{n^2}\longrightarrow 0$ *almost surely.*

Proof. As h_n is decreasing and using the decreasing assumption on the kernel, it follows that

$$0\le\left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)K\left(\frac{x-u}{h_k}\right)\le\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)K\left(\frac{x-u}{h_{n^2}}\right).$$

Now, for every $k\in(n^2,(n+1)^2]$,

$$\left|\sum_{j=1}^{k}\left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_k}\right)-I\!E\,K\left(\frac{x-X_j}{h_k}\right)\right]\right|\le$$

$$\le\frac{1}{n^2}\sum_{j=1}^{(n+1)^2}\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_{n^2}}\right)+I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]. \tag{3}$$

Let us first look at the terms with the mathematical expectation. On one side we have

$$\frac{1}{n^2h_{(n+1)^2}}\sum_{j=1}^{(n+1)^2}I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)=\frac{(n+1)^2}{n^2}\,\frac{h_{n^2}}{h_{(n+1)^2}}\,\frac{1}{h_{n^2}}I\!E\,K\left(\frac{x-X}{h_{n^2}}\right)\longrightarrow f(x)$$

using Lemma 2.1. Analogously, it follows

$$\frac{1}{n^2h_{n^2}}\sum_{j=1}^{(n+1)^2}I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\longrightarrow f(x).$$

Thus, in (3) the terms corresponding to the mathematical expectations converge to zero. This allows replacing the sign + on the right hand side of (3) by a sign - as, given $\delta > 0$, for n large enough, the right hand side in (3) becomes smaller than

$$\frac{1}{n^2}\sum_{j=1}^{(n+1)^2}\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]+\delta,$$

and it is enough to verify that this term converges almost surely to zero. Like this, Chebyshev's inequality gives an upper bound with a variance term: for any $\varepsilon > 0$,

$$I\!P\left(\frac{1}{n^2}\left|\sum_{j=1}^{(n+1)^2}\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]\right|>\varepsilon\right)\leq$$

$$\leq\frac{1}{\varepsilon^2 n^2}\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)^2\sum_{j,j'=1}^{(n+1)^2}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right),K\left(\frac{x-X_{j'}}{h_{n^2}}\right)\right)=$$

$$=\frac{1}{\varepsilon^2 n^2}\left(\frac{h_{n^2}}{h_{(n+1)^2}}-1\right)^2\frac{1}{n^2 h_{n^2}}\sum_{j,j'=1}^{(n+1)^2}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right),K\left(\frac{x-X_{j'}}{h_{n^2}}\right)\right),$$

which defines a convergent series, taking account of Lemma 2.2 and the assumptions made on the bandwidth sequence.

□

LEMMA 3.4. *Under the assumptions of Theorem 3.2,* $\dfrac{b_n}{n^2}\longrightarrow 0$ *almost surely.*

Proof. Using the decreasing assumption on the kernel, it follows that, for every $k\in(n^2,(n+1)^2]$,

$$\frac{|b_n|}{n^2}\leq\frac{1}{n^2h_{n^2}}\sum_{j=1}^{(n+1)^2}\left(\left[K\left(\frac{x-X_j}{h_{n^2}}\right)+I\!E\,K\left(\frac{x-X_j}{h_{n^2}}\right)\right]-\left[K\left(\frac{x-X_j}{h_{(n+1)^2}}\right)+I\!E\,K\left(\frac{x-X_j}{h_{(n+1)^2}}\right)\right]\right).$$

As in the previous lemma, it is easy to check that the terms with the mathematical expectations cancel each other in the limit. So, we are left with checking the almost sure convergence to zero of

$$\frac{1}{n^2h_{n^2}}\sum_{j=1}^{(n+1)^2}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-K\left(\frac{x-X_j}{h_{(n+1)^2}}\right)\right].$$

Given $\varepsilon > 0$,

$$I\!P\left(\frac{1}{n^2h_{n^2}}\left|\sum_{j=1}^{(n+1)^2}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-K\left(\frac{x-X_j}{h_{(n+1)^2}}\right)\right]\right|>\varepsilon\right)\leq$$

$$\leq\frac{1}{\varepsilon n^4h_{n^2}^2}\sum_{j,j'=1}^{(n+1)^2}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right)-K\left(\frac{x-X_j}{h_{(n+1)^2}}\right),K\left(\frac{x-X_{j'}}{h_{n^2}}\right)-K\left(\frac{x-X_{j'}}{h_{(n+1)^2}}\right)\right).$$

Using Lemma 2.1 and $\frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1$, it follows that the sum of these covariances, divided by $n^2 h_{n^2}$ is convergent (to $4g_2(x,x)\int K^2(u)\,du$). As there remains the term $\frac{1}{n^4 h_{n^2}^2}$ we have in fact a convergent series defined by the probabilities.

□

Notice that the association and the fact that K is of bounded variation has not yet been used. These assumptions are required to prove the almost sure convergence of the remaining term.

LEMMA 3.5. *Under the assumptions of Theorem 3.2,* $\dfrac{c_n}{n^2} \longrightarrow 0$ *almost surely.*

Proof. Write $K = K_1 - K_2$, with K_1, K_2 increasing functions. The variables $K_1\Big(\frac{x-X_j}{h_{n^2}}\Big)$, $j = 1, 2, \ldots$, being monotone transformations of associated variables, are associated. Then, we may apply the generalization of Kolmogorov's inequality for associated variables proved by Newman and Wright [16], to obtain that, given $\varepsilon > 0$,

$$I\!P\left(\frac{1}{n^2}\max_{n^2<k\le(n+1)^2}\left|\sum_{j=n^2+1}^{k}\left[K_1\Big(\frac{x-X_j}{h_{n^2}}\Big) - I\!E\,K_1\Big(\frac{x-X_j}{h_{n^2}}\Big)\right]\right| > \varepsilon\right) \le$$

$$\le \frac{2}{\varepsilon^2 n^4 h_{n^2}^2}\sum_{j,j'=n^2+1}^{(n+1)^2}\mathrm{Cov}\left(K_1\Big(\frac{x-X_j}{h_{n^2}}\Big), K_1\Big(\frac{x-X_{j'}}{h_{n^2}}\Big)\right).$$

Using the association this sum is bounded above by the sum with j, j' ranging from 1 to $(n+1)^2$ and then the proof is completed repeating the arguments of the two previous lemmas. The terms corresponding to K_2 are treated analogously.

□

4 Asymptotic normality of the estimator

We now look at the asymptotic normality of the kernel estimator. We will prove the asymptotic normality under conditions similar to those of Theorem 4.2 in Jacob and Oliveira [13], which is somewhat weaker than second order stationarity. As the proof is quite long and technical we will divide it into several lemmas that will be put together at the end of the section to state the general result. For the remaining of the section let k be an integer smaller than n and $m = \lfloor \frac{n}{k} \rfloor$ the greatest integer less or equal than n/k. Let us further define the random variables

$$T_{n,i} = \frac{1}{\sqrt{h_n}}\left(K\Big(\frac{x-X_i}{h_n}\Big) - I\!E\,K\Big(\frac{x-X_i}{h_n}\Big)\right) \qquad i = 1, \ldots, n,\ n \in I\!N$$

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{n,i}, \qquad \text{and} \qquad Y_{n,j} = \frac{1}{\sqrt{k}}\sum_{l=(j-1)k+1}^{jk} T_{n,l}, \quad j = 1, \ldots, m.$$

Note that $T_{mk} = \frac{1}{\sqrt{m}}\sum_{j=1}^{m} Y_{mk,j} = \frac{1}{\sqrt{mk}}\sum_{i=1}^{mk} T_{mk,i}$. The first lemma describes the behaviour of the variances and is just a restating of Lemma 2.2.

LEMMA 4.1. *Suppose* ***(D)*** *and* ***(A)*** *are satisfied. Then*

$$\sigma_{mk}^2 = \mathrm{Var}(T_{mk}) \longrightarrow g_2(x,x)\int K^2(u)\,du.$$

This first step towards the asymptotic normality is proving that it is enough to look at those values of n that are multiples of k.

LEMMA 4.2. *Suppose* ***(D)*** *and* ***(A)*** *are satisfied. Let k be fixed and suppose further that*

$$h_n \longrightarrow 0, \qquad \frac{h_{n+1}}{h_n} \longrightarrow 1,$$

$$K \text{ is bounded} \qquad \text{and} \qquad \lim_{|u|\longrightarrow\infty} K(u) = 0.$$

Then

$$\left|I\!E e^{iuT_n} - I\!E e^{iuT_{mk}}\right| \longrightarrow 0.$$

Proof. As $\left|I\!E e^{iuT_n} - I\!E e^{iuT_{mk}}\right| \leq |u| \operatorname{Var}^{1/2}(T_n - T_{mk})$ it is enough to prove the convergence to zero of this variance. Now we write

$$\begin{aligned}\operatorname{Var}^{1/2}(T_n - T_{mk}) \leq \operatorname{Var}^{1/2}&\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{mk}(T_{n,j} - T_{mk,j})\right] + \\ &+\operatorname{Var}^{1/2}\left[\left(\frac{1}{\sqrt{mk}} - \frac{1}{\sqrt{n}}\right)\sum_{j=1}^{mk} T_{mk,j}\right] + \operatorname{Var}^{1/2}\left[\frac{1}{\sqrt{n}}\sum_{j=mk+1}^{n} T_{n,j}\right],\end{aligned} \tag{4}$$

and prove that each of these terms converges to zero. As for the first term

$$\begin{aligned}\frac{1}{n}\operatorname{Var}&\left[\sum_{j=1}^{mk}(T_{n,j} - T_{mk,j})\right] = \\ &= \frac{1}{n}\sum_{j,j'=1}^{mk}\left[\operatorname{Cov}\left(\frac{1}{\sqrt{h_n}}K\left(\frac{x - X_j}{h_n}\right), \frac{1}{\sqrt{h_n}}K\left(\frac{x - X_{j'}}{h_n}\right)\right)\right. \\ &\quad - \operatorname{Cov}\left(\frac{1}{\sqrt{h_n}}K\left(\frac{x - X_j}{h_n}\right), \frac{1}{\sqrt{h_{mk}}}K\left(\frac{x - X_{j'}}{h_{mk}}\right)\right) \\ &\quad - \operatorname{Cov}\left(\frac{1}{\sqrt{h_{mk}}}K\left(\frac{x - X_j}{h_{mk}}\right), \frac{1}{\sqrt{h_n}}K\left(\frac{x - X_{j'}}{h_n}\right)\right) \\ &\quad \left. + \operatorname{Cov}\left(\frac{1}{\sqrt{h_{mk}}}K\left(\frac{x - X_j}{h_{mk}}\right), \frac{1}{\sqrt{h_{mk}}}K\left(\frac{x - X_{j'}}{h_{mk}}\right)\right)\right].\end{aligned}$$

From Lemma 2.2, as $\frac{n}{mk} \longrightarrow 1$, it follows that the summation over the first and last terms of this expansion is convergent to $2g_2(x,x)\int K^2(u)\,du$. The remaining terms are of the form

$$\begin{aligned}\frac{1}{n}\sum_{j,j'=1}^{mk}\frac{1}{\sqrt{h_n h_{mk}}}&\left(\int K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-v}{h_{mk}}\right)(g_{1,j,j'} - f\otimes f)\,dudv + \right. \\ &\left. + \int K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-u}{h_{mk}}\right)g_{2,j,j'}(u,u)\,du\right).\end{aligned}$$

The part corresponding to the first integral converges to zero as it is easily checked reproducing the arguments in the proof of Lemma 2.2. As for the second integral, rewrite it as

$$\frac{1}{n}\sqrt{\frac{h_n}{h_{mk}}}\sum_{j,j'=1}^{mk}\int K(z)K\left(z\frac{h_n}{h_{mk}}\right)g_{2,j,j'}(x-h_nz,x-h_nz)\,dz,$$

which, by the Lebesgue Dominated Convergence Theorem, **(D)** and the assumption $\frac{h_{n+1}}{h_n}\longrightarrow 1$, converges to $g_2(x,x)\int K^2(u)\,du$. So, it follows that

$$\frac{1}{n}\mathrm{Var}\left[\sum_{j=1}^{mk}(T_{n,j}-T_{mk,j})\right]\longrightarrow 0.$$

The second term in (4) may be rewritten as

$$\left(\frac{1}{\sqrt{mk}}-\frac{1}{\sqrt{n}}\right)^2\mathrm{Var}\left[\sum_{j=1}^{mk}T_{mk,j}\right]=\left(1-\frac{\sqrt{mk}}{\sqrt{n}}\right)^2\frac{1}{mkh_{mk}}\sum_{j,j'=1}^{mk}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{mk}}\right),K\left(\frac{x-X_{j'}}{h_{mk}}\right)\right)$$

As $\left(1-\frac{\sqrt{mk}}{\sqrt{n}}\right)\longrightarrow 0$, the convergence to zero of this term follows from Lemma 2.2. Finally, for the third term in (4):

$$\begin{aligned}
&\frac{1}{n}\mathrm{Var}\left[\sum_{j=mk+1}^{n}T_{n,j}\right]=\\
&\quad=\frac{1}{nh_n}\sum_{j,j'=mk+1}^{n}\left|\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_n}\right),K\left(\frac{x-X_{j'}}{h_n}\right)\right)\right|\leq\\
&\quad\leq\frac{1}{nh_n}\sum_{j,j'=1}^{n}\left(\int_{I\!R^2}K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-v}{h_n}\right)\left|g_{1,j,j'}(u,v)-f(u)f(v)\right|\,dudv+\right.\\
&\qquad\qquad\left.+\int_{\Delta}K^2\left(\frac{x-u}{h_n}\right)g_{2,j,j'}(u,u)\,du\right),
\end{aligned}$$

which converges to zero due to the nonnegativity of the terms and $\frac{h_{n+1}}{h_n}\longrightarrow 1$.

□

The variables T_{mk} are the sum of $m^{-1/2}Y_{mk,1},\ldots,m^{-1/2}Y_{mk,m}$, thus we need to prove a Central Limit Theorem for this sum. Note that the variables $Y_{mk,j}$, $j=1,\ldots,m$, are not associated as they are not obtained by monotone transformations of the variables X_j, $j\in I\!N$. Thus, in order to be able to use this association, we need, as before, some extra assumption on the kernel: we will suppose K to be of bounded variation. This means that $K=K_1-K_2$ with K_1 and K_2 monotone functions. Corresponding to these two functions we define, for each $n\in I\!N$ the analogous of the auxiliary variables introducing in this section:

$$T_{n,i,q}=\frac{1}{\sqrt{h_n}}\left(K_q\left(\frac{x-X_i}{h_n}\right)-I\!E\,K_q\left(\frac{x-X_i}{h_n}\right)\right),\qquad q=1,2,\ i=1,\ldots,n,$$

$$Y_{n,j,q}=\tfrac{1}{\sqrt{k}}\textstyle\sum_{l=(j-1)k+1}^{jk}T_{n,l,q},\qquad q=1,2,\ j=1,\ldots,m.$$

Remark that Lemma 2.2 is still applicable when replacing K by any of the functions K_1 or K_2 obtaining, evidently, a different normalization on the limit.

First we control the difference between the situation we are dealing with and what we would find if the variables were independent.

LEMMA 4.3. *Suppose* ***(D)*** *and* ***(A)*** *are satisfied. Suppose the kernel K is of bounded variation with K_1 and K_2 being bounded. If, for each k fixed,*

$$\lim_{m\to+\infty} \frac{1}{m} \sum_{j=1}^{m} \sum_{l,l'=(j-1)k+1}^{jk} g_{2,l,l'} = g_{2,k} \tag{5}$$

uniformly, where $g_{2,k}$ is bounded and continuous, then

$$\limsup_{m\to+\infty} \left| I\!E e^{iuT_{mk}} - \prod_{j=1}^{m} I\!E e^{i\frac{u}{\sqrt{m}}Y_{mk,j}} \right| \leq Mu^2(g_2(x,x) - g_{2,k}(x,x)),$$

where $M = \int K_1^2(u) + K_2^2(u) + 2K_1(u)K_2(u)\,du$.

Proof. As the variables $Y_{mk,j}$, $j = 1,\ldots,m$, are not associated but, nevertheless, functions of associated variables, we apply the extended version of Newman's inequality (see Theorem 16 in Newman [15]) to obtain

$$\left| I\!E e^{iuT_{mk}} - \prod_{j=1}^{m} I\!E e^{i\frac{u}{\sqrt{m}}Y_{mk,j}} \right| \leq \frac{u^2}{2m} \sum_{\substack{j,j'=1 \\ j\neq j'}}^{m} \mathrm{Cov}(Y_{mk,j,1} + Y_{mk,j,2}, Y_{mk,j',1} + Y_{mk,j',2}).$$

After expanding this covariance, we find four terms which are controlled in the same way as the following one:

$$\frac{1}{m} \sum_{\substack{j,j'=1 \\ j\neq j'}}^{m} \mathrm{Cov}(Y_{mk,j,1}, Y_{mk,j',1}) = \frac{1}{m} \sum_{j,j'=1}^{m} \mathrm{Cov}(Y_{mk,j,1}, Y_{mk,j',1}) - \frac{1}{m} \sum_{j=1}^{m} \mathrm{Var}(Y_{mk,j,1}).$$

From Lemma 2.2, it follows that

$$\frac{1}{m} \sum_{j,j'=1}^{m} \mathrm{Cov}(Y_{mk,j,1}, Y_{mk,j',1}) \longrightarrow g_2(x,x) \int K_1^2(u)\,du.$$

As for the other term,

$$\frac{1}{m} \sum_{j=1}^{m} \mathrm{Var}(Y_{mk,j,1}) = \frac{1}{mk} \sum_{j,j'=1}^{m} \mathrm{Cov}(Y_{mk,j,1}, Y_{mk,j',1}) \longrightarrow g_{2,k}(x,x) \int K_1^2(u)\,du$$

reproducing the arguments in the proof of Lemma 2.2. So the lemma is proved.

□

Note that, if the variables X_n, $n \in I\!N$, are second order stationary, condition (5) is satisfied.

The next step towards the proof of the asymptotic normality of the estimator is to prove a Central Limit Theorem for the variables $Y_{mk,j}$, $j = 1,\ldots,m$, treating them as if they where independent, in view of the preceding lemma (in fact we should introduce a new collection of independent variables with the same distributions as the ones we have; we shall not do so to avoid further notation).

LEMMA 4.4. *Suppose* ***(D)***, ***(A)***, *(2) and (5) are satisfied. If, for any constant* $c > 0$,

$$\int_{\{K^2\left(\frac{x-X}{h_n}\right) > cnh_n\}} \frac{1}{h_n} K^2\left(\frac{x-X}{h_n}\right) I\!P \longrightarrow 0, \tag{6}$$

then $m^{-1/2} \sum_{j=1}^{m} Y_{mk,j}$ *converges in distribution to a centered gaussian random variable with covariance* $g_{2,k}(x,x) \int K^2(u)\, du$.

Proof. As before, it is easy to check that

$$\frac{1}{m} \mathrm{Var} \left[\sum_{j=1}^{m} Y_{mk,j} \right] \longrightarrow g_{2,k}(x,x) \int K^2(u)\, du,$$

so that the Lindeberg condition reduces to verifying

$$\sum_{j=1}^{m} \int_{\{|Y_{mk,j}| > cg_{2,k}(x,x)\sqrt{m}\}} \frac{1}{m} Y_{mk,j}^2 \, I\!P \longrightarrow 0.$$

This integral is, using Lemma 4 in Utev [21], bounded above by

$$\frac{2}{m} \sum_{j=1}^{m} \sum_{l=(j-1)k+1}^{jk} \int_{\{|T_{mk,l}| > \frac{cg_{2,k}(x,x)}{2}\sqrt{\frac{m}{k}}\}} T_{mk,l}^2 \, I\!P = \frac{2}{m} \sum_{j=1}^{mk} \int_{\{|T_{mk,j}| > \frac{cg_{2,k}(x,x)}{2}\sqrt{\frac{m}{k}}\}} T_{mk,j}^2 \, I\!P$$

and this sum converges to zero using a Cesàro argument and (6) (see Theorem 10 in Jacob and Oliveira [12] for the computational details).

□

It remains now to gather all the partial results to find a set of conditions ensuring the asymptotic normality of the kernel estimator, as announced earlier. Given a real number v, we define $\sigma^2(v) = v \int K^2(u)\, du$.

THEOREM 4.5. *Suppose* ***(D)***, ***(A)*** *and (5)are satisfied. If*

$$h_n \longrightarrow 0, \quad nh_n \longrightarrow +\infty, \quad \frac{h_{n+1}}{h_n} \longrightarrow 1,$$

K is of bounded variation, with K_1, K_2 bounded

$$\lim_{|u| \to +\infty} K(u) = 0$$

and

$$\lim_{k \to +\infty} g_{2,k} = g_2 \tag{7}$$

uniformly, where the functions $g_{2,k}$ are those introduced in (5), then

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} \left(K\left(\frac{x-X_i}{h_n}\right) - I\!E\, K\left(\frac{x-X_i}{h_n}\right) \right)$$

converges in distribution to a centered gaussian random variable with covariance $\sigma^2(g_2(x,x)) = g_2(x,x) \int K^2(u)\, du$.

Proof. Write

$$\left| I\!E e^{iuT_n} - e^{-\frac{u^2}{2}\sigma^2(g_2(x,x))} \right| \leq$$

$$\leq \left| I\!E e^{iuT_n} - I\!E e^{iuT_{mk}} \right| + \left| I\!E e^{iuT_{mk}} - \prod_{j=1}^{m} I\!E e^{i\frac{u}{\sqrt{m}}Y_{mk,j}} \right| +$$

$$+ \left| \prod_{j=1}^{m} I\!E e^{i\frac{u}{\sqrt{m}}Y_{mk,j}} - e^{-\frac{u^2}{2}\sigma^2(g_{2,k}(x,x))} \right| + \left| e^{-\frac{u^2}{2}\sigma^2(g_{2,k}(x,x))} - e^{-\frac{u^2}{2}\sigma^2(g_2(x,x))} \right|.$$

For each k fixed, it follows from the preceding lemmas that (note that, as K is bounded and $nh_n \longrightarrow +\infty$, condition (6) is satisfied, for $n \in I\!N$ large enough),

$$\limsup_{n\to+\infty} \left| I\!E e^{iuT_n} - e^{-\frac{u^2}{2}\sigma^2(g_2(x,x))} \right| \leq Mu^2(g_2(x,x) - g_{2,k}(x,x)) + \left| e^{-\frac{u^2}{2}\sigma^2(g_{2,k}(x,x))} - e^{-\frac{u^2}{2}\sigma^2(g_2(x,x))} \right|,$$

where M is defined in Lemma 4.3. Letting now $k \longrightarrow +\infty$ this upper bound converges to zero, according to (7).

□

It is not difficult to adapt the previous proof to a multidimensional version. This will not be done here for sake of brevity, but we will state, nevertheless, the result.

THEOREM 4.6. *Suppose all the conditions of the previous theorem are satisfied. Then, given,* $x_1, \ldots, x_s \in I\!R$,

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{n} \left(K\left(\frac{x_1 - X_i}{h_n}\right) - I\!E\, K\left(\frac{x_1 - X_i}{h_n}\right), \ldots, K\left(\frac{x_s - X_i}{h_n}\right) - I\!E\, K\left(\frac{x_s - X_i}{h_n}\right) \right)$$

converges in distribution toa centered gaussian random vector with diagonal covariance matrix $\Gamma = diag(\sigma^2(g_2(x_1, x_1)), \ldots, \sigma^2(g_2(x_s, x_s)))$.

References

[1] Bagai, I. and Prakasa Rao, B. L. S., Estimation of the survival function for stationary associated processes, Statist. Probab. Lett. 12(1991), 385-291.

[2] Bagai, I. and Prakasa Rao, B. L. S., Kernel-type density and failure rate estimation for associated sequences, Ann. Inst. Statist. Math. 47(1995), 253-266.

[3] Bensaïd N. and Fabre, J., Estimation par noyau d'une derivée de Radon-Nikodym sous des conditions de mélange, Canadian J. Statist. 28(1998), 267-282.

[4] Bensaïd N. and Oliveira, P. E., Histogram estimation of Radon-Nikodym derivatives for strong mixing data, to appear in Statistics

[5] Bosq, D., Nonparametric statistics for stochastic processes, Lecture Notes in Statistics 110, Springer, 1996.

[6] Bosq, D. and Lecoutre, J. P., Théorie de l'estimation fonctionelle, Economica, Paris, 1987.

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

[8] Ellis, S. P., Density estimation for point processes, Stoch. Proc. Appl., 39(1991), 345-358.

[9] Ferrieux, D., Estimation de densités de mesures moyennes de processus ponctuels associés, Thesis, University of Montpellier II, 1996.

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

[11] Jacob, P. and Oliveira, P. E., A general approach to non parametric histogram estimation, Statistics, 27(1995), 73-92.

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

[13] Jacob, P. and Oliveira, P. E., Histogram and associated point processes, Statist. Inf. Stoch. Proc. 2(1999), 227-251.

[14] Masry, E., Recursive Probability Density Estimation for weakly Dependent Stationary processes, IEEE Trans. inform. Theory II, 32(1986), 254-267.

[15] Newman, C., 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(1984), pp. 127-140.

[16] Newman, C. and Wright, A. L., An invariance principle for certain dependent sequences, Ann. Probab. 9(1981), 671-675.

[17] Pham, T. D. and Tran, L. T., Kernel Density Estimation under a locally Mixing condition, in Nonparametric Functional Estimation and Related Topics, Ed. G. Roussas, Nato Series, V. 335(1991), 419-430.

[18] Roussas, G. G., Nonparametric Estimation in mixing sequences of random variables, J. Statist. Plann. Inference 18(1988), 135-149.

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

[20] Roussas, G. G., Asymptotic normality of the kernel estimate of a probability density function under association, Statist. Probab. Lett., 50(2000), 1-12.

[21] Utev, S. A., On the central limit theorem for φ-mixing arrays of random variables, Th. Probab. Appl., 35(1990), 131-139.