PORTUGALIAE MATHEMATICA
Vol. 57 Fasc. 1 – 2000

MEAN SQUARE ERROR FOR HISTOGRAMS WHEN ESTIMATING RADON–NIKODYM DERIVATIVES

P.E. Oliveira *

Abstract: The use of histograms for estimation of Radon–Nikodym derivatives is addressed. Some results concerning the convergence have been established with no reference about the behaviour of the error. In this paper we study the mean square convergence rate of this error. The optimization of the partitions on which the histograms are based thus obtained recovers the $n^{-2/(p+2)}$ rate known for some problems that are included in this more general framework.

1 – Introduction

Inference for point processes has been the object of a very wide literature, including problems such as regression estimation, Palm distributions or density estimation, among others. This note intends to complement results by Jacob, Oliveira [9, 11], where histograms were considered to estimate Radon–Nikodym derivatives between means of random measures. Although we may think in terms of random measures, which includes point processes, most of the examples presented deal with point processes and their intensities. These include many classical functional problems and provide a better interpretation of the assumptions we will introduce. In [9] the construction of histograms was based on an embedded sequence of partitions, whereas in [11] the embedding property was replaced by a decomposition of product measures (condition **(M)** below), which enables us to recover the same convergence results as in the embedded case. The gen-

Received: April 8, 1998; *Revised*: July 28, 1998.
AMS Subject Classification: 62G05, 62G20.
Keywords: Point process; Histogram; Convergence rate.
* Author supported by grants PRAXIS/2/2.1/MAT/19/94, PRAXIS/2/2.1/MAT/400/94 from FCT and Instituto de Telecomunicações – Pólo de Coimbra.

eral idea is to define two integrable random measures ξ and η with mean ν and μ, respectively, such that $\mu \ll \nu$ and estimate $\frac{d\mu}{d\nu}$. This framework has been used in Ellis [6] for a particular choice of ξ and η, in order to address density estimation, Jacob, Mendes Lopes [7], where the authors considered absolutely continuous random measures, thus setting the problem in terms of the random densities associated, Jacob, Oliveira [9, 10, 11] with the same framework as stated here. All the references quoted above suppose independent sampling and, with the exception of [9] and [11] where histograms are considered, kernel estimates. Some work has been produced for non independent sampling in this framework: Bensaïd, Fabre, [1] considered strong mixing samples, Ferrieux [4, 5] and Roussas [12, 13, 14] considered associated sampling for the case where ξ is almost surely fixed. All these extensions deal with kernel estimates.

The convergence of the estimators is the main problem studied in the above mentioned references. Only some of them address the convergence rate of the mean square error of the estimator. Results are mentioned in Bensaïd, Jacob [2] and Ferrieux [4], both for kernel estimates. It is interesting to note that, although the setting is quite general, the results derived recover the known rates in some classical estimation problems that are included in this setting. Thus, this seems not to produce a loss in the power of analysis, although we will need some moment conditions avoidable in some cases. The complements of the results of Jacob, Oliveira [9, 11] mentioned above will mean that we seek the mean squared convergence rate for the histogram based on independent sampling. Again, we will find the optimal convergence rates for the classical problems included in this setting. The methods used here are quite close to those used by Bensaïd, Jacob [2].

2 – Auxiliary results

Although in [9] and [11] the authors considered random measures on some metric space, here we will take them to be on $\mathbb{R}^p$, for some fixed $p \geq 1$. Some more generality could be achieved, but then we would be limited on the subsequent analysis of the convergence rates. The main tool, as expected, are Taylor expansions and these may be used in a more general setting than $\mathbb{R}^p$, leading to conditions somewhat weaker than the differentiability conditions we will use, but with no real gain on the results. So, we choose to work in $\mathbb{R}^p$, to gain readability and also because for the usual examples this is quite satisfactory. We will suppose that $\mu \ll \nu \ll \lambda$, the Lebesgue measure on $\mathbb{R}^p$.

Just for sake of completeness we recall here how to reduce our setting to obtain some classical estimation problems. We will denote by $\mathbb{1}_A$ the indicator function of the set A.

- (Ellis [6]) Density estimation: take $\xi=\nu$ almost surely, $\eta=\delta_X$, where X is a random variable with distribution absolutely continuous with respect to ν. Then $\frac{d\mu}{d\nu}$ is the density of X with respect to ν.
- Regression: suppose Y is an almost surely non-negative real random variable and X a random variable on $\mathbb{R}^p$. Then, if $\xi=\delta_X$ and $\eta=Y\delta_X$, the conditional expectation $\mathrm{E}(Y|X=s)$ is a version of $\frac{d\mu}{d\nu}$.
- Thinning: suppose $\xi=\sum_{i=1}^N \delta_{X_i}$, where the X_n, $n\in\mathbb{N}$, are random variables on $\mathbb{R}^p$, α_n, $n\in\mathbb{N}$, are Bernoulli variables, conditionally independent given the sequence X_n, $n\in\mathbb{N}$, with parameters $p(X_n)$, and put $\eta=\sum_{i=1}^N \alpha_i\,\delta_{X_i}$. Then $\frac{d\mu}{d\nu}$ is the thinning function giving the probability of suppressing each point.
- Marked point processes: let $\zeta=\sum_{i=1}^N \delta_{(X_i,T_i)}$ be a point process on $\mathbb{R}^p\times\mathrm{T}$ such that the margin $\xi=\sum_{i=1}^N \delta_{X_i}$ is itself a point process. If $B\subset\mathrm{T}$ is measurable, choosing $\alpha_n=\mathbb{1}_B(T_n)$, and $\eta=\sum_{i=1}^N \alpha_i\,\delta_{X_i}$, we have

  $$\mathrm{E}\zeta(A\times B)=\int_A \frac{d\mu}{d\nu}(s)\ \mathrm{E}\zeta(ds\times\mathbb{R})\ ,$$

  thus $\frac{d\mu}{d\nu}$ is the marking function.
- Cluster point processes: suppose $\zeta=\sum_{i=1}^N\sum_{j=1}^{N_i}\delta_{(X_i,Y_{i,j})}$ is a point process on $\mathbb{R}^p\times\mathbb{R}^p$ such that $\sum_{i=1}^N\sum_{j=1}^{N_i}\delta_{Y_{i,j}}$ is also a point process (for which it suffices that, for example, N and the N_n, $n\in\mathbb{N}$, are almost surely finite). The process $\xi=\sum_{i=1}^N\delta_{X_i}$ identifies the cluster centers and the processes $\zeta_{X_i}=\sum_{i=1}^{N_i}\delta_{Y_{i,j}}$ identify the points. The distribution of ζ may be characterized by a markovian kernel of distributions $(\pi_x,\,x\in\mathbb{R}^p)$ with means $(a_x,\,x\in\mathbb{R}^p)$ such that, conditionally on $\xi=\sum_{i=1}^N\delta_{x_i}$, $(\zeta_{x_1},\ldots,\zeta_{x_n})$ has distribution $\pi_{x_1}\otimes\cdots\otimes\pi_{x_n}$. Defining $\eta(A)=\zeta(A\times B)$, with B a fixed bounded Borel subset of $\mathbb{R}^p$, we have $\frac{d\mu}{d\nu}(x)=a_x(B)$ ν-almost everywhere.
- Markovian shifts: this is a special case of the previous example, when $N_i=1$ a.s., $i\geq 1$. Looking at the previous example, the conclusion is that the random vector $(Y_1,\ldots,Y_n)$ has distribution $a_{x_1}\otimes\cdots\otimes a_{x_n}$ (we replaced the double index of the Y variables by a single one as, for each i fixed, there is only one such variable). Then it would follow that $\frac{d\mu}{d\nu}(x)=a_x(B)=\mathrm{P}(Y\in B|X=x)$.

So, as illustrated by the examples above, we will be concerned with the estimation of the Radon–Nikodym derivative $\frac{d\mu}{d\nu}$.

To define the histogram we introduce a sequence of partitions Π_k, $k \in \mathbb{N}$, of a fixed compact set B, verifying

- **(P1)** for each $k \in \mathbb{N}$, the sets in Π_k are bounded Borel measurable;
- **(P2)** for each $k \in \mathbb{N}$, Π_k is finite;
- **(P3)** $\sup\left\{\mathrm{diam}(I) : I \in \Pi_k\right\} \longrightarrow 0$;
- **(P4)** for each $k \in \mathbb{N}$ and $I \in \Pi_k$, $\nu(I) > 0$;
- **(P5)** for each $k \in \mathbb{N}$ the Lebesgue measure of the sets in Π_k is constant and equal to h_k^p. Further $\lim_{k \to +\infty} h_k = 0$.

Given a point $s \in B$ we denote by $I_k(s)$ the unique set of Π_k containing the point s and define, for each $k \in \mathbb{N}$,

$$g_k(s) = \sum_{I \in \Pi_k} \frac{\mu(I)}{\nu(I)} \, \mathbb{I}_I(s) = \frac{\mu(I_k(s))}{\nu(I_k(s))} \; .$$

It is well known that if φ is a version of $\frac{d\mu}{d\nu}$ continuous on B, then

$$\sup_{s \in B} |g_k(s) - \varphi(s)| \longrightarrow 0 \; .$$

Given $((\xi_1, \eta_1), \ldots, (\xi_n, \eta_n))$ an independent sample of (ξ, η) and defining $\overline{\xi}_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, $\overline{\eta}_n = \frac{1}{n} \sum_{i=1}^n \eta_i$, the histogram is

$$\varphi_n(s) = \sum_{I \in \Pi_k} \frac{\overline{\eta}_n(I)}{\overline{\xi}_n(I)} \, \mathbb{I}_I(s) = \frac{\overline{\eta}_n(I_k(s))}{\overline{\xi}_n(I_k(s))}$$

(we define $\varphi_n(s)$ as zero whenever the denominator vanishes, as usual), where the dependence of k on n is to be specified to obtain the convergence.

As we are not working with embedded partitions we need the following assumptions, as in Jacob, Oliveira [11]. A measure m on $\mathbb{R}^p \times \mathbb{R}^p$ satisfies condition **(M)** with respect to the measure ν on $\mathbb{R}^p$ if $m = m_1 + m_2$ where m_2 is a measure on Δ, the diagonal of $\mathbb{R}^p \times \mathbb{R}^p$ and m_1 is a measure on $\mathbb{R}^p \times \mathbb{R}^p \setminus \Delta$, verifying

- **(M1)** $m_1 \ll \nu \otimes \nu$ and there exists a version γ_1 of the Radon–Nikodym derivative $\frac{dm_1}{d\nu \otimes \nu}$ which is bounded;
- **(M2)** $m_2 \ll \nu^*$, where ν^* is the measure on Δ defined by lifting ν, that is, $\nu^*(A^*) = \nu(A)$ with $A^* = \{(s,s) : s \in A\}$, and there exits a continuously differentiable version γ_2 of the Radon–Nikodym derivative $\frac{dm_2}{d\nu^*}$.

In [11] the function γ_2 was only supposed continuous, as only the convergence of the estimator was considered. The differentiability will allow the use of the Taylor expansion that serves as a tool for establishing the convergence rates.

We will be using decomposition **(M)** throughout our results, so we present some examples showing it is reasonable to suppose that this decomposition is satisfied.

1. We begin with a simple situation. Suppose that $\xi = \delta_X + \delta_Y$, where X and Y are independent $\mathbb{R}^p$-valued variables with distributions P_X and P_Y, respectively. Then $\nu = \mathrm{E}(\xi) = \mathrm{P}_X + \mathrm{P}_Y$, so

$$\nu \otimes \nu \;=\; \mathrm{P}_X \otimes \mathrm{P}_X + \mathrm{P}_Y \otimes \mathrm{P}_Y + \mathrm{P}_X \otimes \mathrm{P}_Y + \mathrm{P}_Y \otimes \mathrm{P}_X \;\;.$$

On the other hand $\mathrm{E}(\xi \otimes \xi) = \mathrm{P}_X \otimes \mathrm{P}_Y + \mathrm{P}_Y \otimes \mathrm{P}_X + \mathrm{P}_X^* + \mathrm{P}_Y^*$, which satisfies **(M)** as long as X and Y do not have common atoms.

2. Let ξ be a Poisson process represented as $\sum_{i=1}^N \delta_{X_i}$, where N is a Poisson random variable and the X_i are independent with common distribution P_X and are independent of N. Then $\nu = \mathrm{E}(\xi) = \mathrm{E}(N)\,\mathrm{E}(\delta_{X_1}) = \mathrm{E}(N)\,\mathrm{P}_X$ and

$$\begin{aligned}\mathrm{E}(\xi \otimes \xi) \;&=\; \mathrm{E}\left(\sum_{i=1}^N \delta_{X_i} \otimes \delta_{X_i} \;+\; \sum_{\substack{i,j=1\\ i\neq j}}^N \delta_{X_i} \otimes \delta_{X_j}\right) \\ &=\; \mathrm{E}(N)\,\mathrm{P}_X^* + \mathrm{E}(N^2 - N)\,\mathrm{P}_X \otimes \mathrm{P}_X \;\;.\end{aligned}$$

3. In most of our examples the point processes are represented as $\xi = \sum_{i=1}^N \delta_{X_i}$, $\eta = \sum_{i=1}^N \alpha_i\, \delta_{X_i}$, with the variables α_n, $n \in \mathbb{N}$, being 0-1 valued. In such cases $\mathrm{E}(\eta \otimes \eta)$ will satisfy **(M)** whenever $\mathrm{E}(\xi \otimes \xi)$ does. If ξ is a Poisson process with a representation such as the one on the previous example then **(M)** was shown to hold. For a more general situation, note that

$$\mathrm{E}(\xi \otimes \xi) \;=\; \mathrm{E}\left(\sum_{\substack{i,j=1\\ i\neq j}}^N \delta_{X_i} \otimes \delta_{X_j}\right) + \mathrm{E}\left(\sum_{i=1}^N \delta_{(X_i,X_i)}\right)$$

so, at least, a decomposition with a diagonal and a non-diagonal component holds. Besides reducing these expressions to the poissonian representation, the absolute continuity relations will depend on the possible dependence between the variables involved. For instance, if N is equal to k with probability 1, we would find $\nu = \sum_{i=1}^k \mathrm{P}_{X_i}$ and

$$\mathrm{E}(\xi \otimes \xi) \;=\; \sum_{\substack{i,j=1\\ i\neq j}}^k \mathrm{P}_{(X_i,X_j)} + \sum_{i=1}^k \mathrm{P}_{X_i}^* \;,$$

so **(M)** will be satisfied if the variables are independent and have no common atoms or if, for every distinct $i, j = 1, ..., k$, $\mathrm{P}_{(X_i, X_j)}(\Delta) = 0$.

4. The example of regression is not included in the previous one. In this case $\nu = \mathrm{E}(\xi) = \mathrm{E}(\delta_X) = \mathrm{P}_X$ and $\mathrm{E}(\xi \otimes \xi) = \mathrm{E}(\delta_{(X,X)}) = \mathrm{P}_X^*$ so there is no non-diagonal component of $\mathrm{E}(\xi \otimes \xi)$. As for η we have

$$\begin{aligned} \mathrm{E}(\eta \otimes \eta)(A^*) &= \mathrm{E}[Y^2\, \mathbb{I}_A(X)] \\ &= \mathrm{E}[Y^2\, \mathbb{I}_{A^*}(X, X)] \\ &= \int_{A^*} \mathrm{E}(Y^2 | X = s)\ \mathrm{P}_X^*(ds) \end{aligned}$$

so **(M)** holds for $\mathrm{E}(\eta \otimes \eta)$ if $\mathrm{E}(Y^2 | X = s)$ exists and is continuous. Further, if ν is absolutely continuous with respect to the Lebesgue measure, λ, on $\mathbb{R}^p$, we may use densities. Let $g(s, y)$ be the density of (X, Y) with respect to $\lambda^p \otimes \lambda$, where λ is the Lebesgue measure on $\mathbb{R}$. Then

$$\mathrm{E}(\eta \otimes \eta)(A^*) = \int_A \int y^2\, g(s, y)\ \lambda(dy)\ \lambda^p(ds)$$

so **(M)** follows from a continuity and boundness assumption on g.

The following is essential for the analysis of the estimator.

Theorem 2.1 (Jacob, Oliveira [11]). *Suppose m is a measure on $\mathbb{R}^p \times \mathbb{R}^p$ that verifies condition* **(M)** *with respect to ν and the sequence of partitions Π_k, $k \in \mathbb{N}$, verifies* **(P1)**–**(P5)**. *Then*

$$\sum_{I \in \Pi_k} \frac{m(I \times I)}{\nu(I)}\, \mathbb{I}_I(s) \longrightarrow \gamma_2(s, s)$$

uniformly on B. ∎

Then, as shown in [11] if

(1) $$n\, h_n^p \to +\infty$$

and $\mathrm{E}(\xi \otimes \xi)$, $\mathrm{E}(\eta \otimes \eta)$ both satisfy **(M)**, $\varphi_n(s)$ converges in probability to $\varphi(s)$. If further, there exists $R > 0$, such that, for every $I \subset B$ and $k \geq 2$,

$$\begin{aligned} \mathrm{E}[\xi^k(I)] &\leq R^{k-2}\, k!\ \mathrm{E}[\xi^2(I)]\ , \\ \mathrm{E}[\eta^k(I)] &\leq R^{k-2}\, k!\ \mathrm{E}[\eta^2(I)]\ , \end{aligned}$$

then $\sup_{x \in B} |\varphi_n(s) - g_k(s)|$ convergences almost completely to zero.

When supposing $\mathrm{E}(\zeta_1 \otimes \zeta_2)$ verifies **(M)**, with $\zeta_1, \zeta_2 \in \{\xi, \eta\}$, we introduce measures that will be denoted by $m_1^{\zeta_1,\zeta_2}$ and $m_2^{\zeta_1,\zeta_2}$, respectively, with the corresponding densities denoted by $\gamma_1^{\zeta_1,\zeta_2}$ and $\gamma_2^{\zeta_1,\zeta_2}$.

Let f and g be versions of $\frac{d\nu}{d\lambda}$ and $\frac{d\mu}{d\lambda}$, respectively. Then, if $\mathrm{E}(\xi\otimes\xi)$, $\mathrm{E}(\eta\otimes\eta)$ satisfy **(M)** and (1) holds,

$$f_n(s) \;=\; \frac{1}{n\,h_n^p} \sum_{i=1}^n \xi_i(I_n(s)) \;\longrightarrow\; f(s)\;,$$

$$g_n(s) \;=\; \frac{1}{n\,h_n^p} \sum_{i=1}^n \eta_i(I_n(s)) \;\longrightarrow\; g(s)\;.$$

As we have $\varphi_n(s) = \frac{g_n(s)}{f_n(s)}$ and $\varphi(s) = \frac{g(s)}{f(s)}$, we will look at the convergences $g_n(s) \to g(s)$ and $f_n(s) \to f(s)$.

To finish with the auxiliary results, we quote a lemma enabling the separation of variables in the quotient φ_n.

Lemma 2.2 (Jacob, Niéré [8]). *Let X and Y be non-negative integrable random variables then, for $\varepsilon > 0$ small enough,*

$$\left\{\left|\frac{X}{Y} - \frac{\mathrm{E}(X)}{\mathrm{E}(Y)}\right| > \varepsilon\right\} \subset \left\{\left|\frac{X}{\mathrm{E}(X)} - 1\right| > \frac{\varepsilon}{4}\,\frac{\mathrm{E}(Y)}{\mathrm{E}(X)}\right\} \cup \left\{\left|\frac{Y}{\mathrm{E}(Y)} - 1\right| > \frac{\varepsilon}{4}\,\frac{\mathrm{E}(Y)}{\mathrm{E}(X)}\right\}\,. \blacksquare$$

3 – The convergence rates

On the sequel f and g will be versions of $\frac{d\nu}{d\lambda}$ and $\frac{d\mu}{d\lambda}$, respectively, which will be supposed continuously differentiable on the compact set B. Also the sequence of partitions Π_k, $k\in\mathbb{N}$, will always be supposed to satisfy **(P1)**–**(P5)**.

According to the final lemma of the preceding section we will separate the variables, so we start with the convergence rates for the histograms f_n, g_n and also for their product.

Theorem 3.1. *If the moment measures $\mathrm{E}(\xi\otimes\xi)$, $\mathrm{E}(\eta\otimes\eta)$ both satisfy* **(M)** *and (1) holds. Then*

$$\mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^2\Big] \;=\; \frac{\gamma_2^{\xi,\xi}(s,s)}{n\,h_n^p} + O(h_n^2) \sum_{k,l=1}^p \frac{\partial f}{\partial x_k}(s)\,\frac{\partial f}{\partial x_l}(s) + o\Big(h_n^{2p} + \frac{1}{n\,h_n^p}\Big)\;,$$

$$\mathrm{E}\Big[\Big(g_n(s)-g(s)\Big)^2\Big] \;=\; \frac{\gamma_2^{\eta,\eta}(s,s)}{n\,h_n^p} + O(h_n^2)\sum_{k,l=1}^p \frac{\partial g}{\partial x_k}(s)\,\frac{\partial g}{\partial x_l}(s) + o\Big(h_n^{2p}+\frac{1}{n\,h_n^p}\Big)\ ,$$

$$\mathrm{E}\Big[\Big(f_n(s)-f(s)\Big)\Big(g_n(s)-g(s)\Big)\Big] \;=\;$$

$$=\; \frac{\gamma_2^{\xi,\eta}(s,s)}{n\,h_n^p} + O(h_n^2)\sum_{k,l=1}^p \frac{\partial f}{\partial x_k}(s)\,\frac{\partial g}{\partial x_l}(s) + o\Big(h_n^{2p}+\frac{1}{n\,h_n^p}\Big)\ .$$

Proof: As usual put $\mathrm{E}[(f_n(s)-f(s))^2]=\mathrm{Var}[f_n(s)]+\mathrm{E}^2[f_n(s)-f(s)]$, and write

$$\mathrm{E}[f_n(s)] \;=\; \frac{1}{h_n^p}\,\nu(I_n(s)) \;=\; \frac{1}{h_n^p}\int_{I_n} f(t)\,\lambda(dt)$$

(we drop the mention to the point s on the set $I_n(s)$ whenever confusion does not arise). Now, as f is continuously differentiable, we may write, with $t=(t_1,...,t_p)$, $f(t)=f(s)+\langle\nabla f(s),t-s\rangle+O(\|t-s\|^2)$, thus

$$\begin{aligned}\mathrm{E}\Big[f_n(s)-f(s)\Big] \;&=\; \frac{1}{h_n^p}\int_{I_n}\Big\langle\nabla f(s),\,t-s\Big\rangle\,\lambda(dt) + \frac{1}{h_n^p}\int_{I_n} O\Big(\|t-s\|^2\Big)\,\lambda(dt)\\ &=\; O(h_n)\sum_{k=1}^p\frac{\partial f}{\partial x_k}(s)+O(h_n^{2p})\end{aligned}$$

as $\|t-s\|\le h_n^p$ and $\lambda(I_n)=h_n^p$. On the other hand writing $\mathrm{E}[\xi^2(I_n)]=\mathrm{E}(\xi\otimes\xi)(I_n\times I_n)=m_1^{\xi,\xi}(I_n\times I_n)+m_2^{\xi,\xi}(I_n^*)$, it follows that

$$\frac{m_1^{\xi,\xi}(I_n\times I_n)}{h_n^p} \;=\; \frac{1}{h_n^p}\int_{I_n\times I_n}\gamma_1^{\xi,\xi}\,d(\lambda\otimes\lambda) \;\le\; \sup_{x\in B}|\gamma_1^{\xi,\xi}(x,x)|\,\lambda(I_n) \;\longrightarrow\; 0$$

and, for some $\theta\in\mathbb{R}^p$ with $\|\theta\|\le 1$,

$$\begin{aligned}\frac{m_2^{\xi,\xi}(I_n^*)}{h_n^p} \;&=\; \frac{1}{h_n^p}\int_{I_n^*}\gamma_2^{\xi,\xi}(t,t)\,\lambda^*(dt)\\ &=\; \frac{1}{h_n^p}\int_{I_n^*}\gamma_2^{\xi,\xi}(s,s)\,\lambda^*(dt) \;+\; \frac{1}{h_n^p}\int_{I_n^*}\Big\langle\nabla\gamma_2^{\xi,\xi}\Big(s+\langle\theta,t-s\rangle\Big),\,t-s\Big\rangle\,\lambda^*(dt)\\ &=\; \gamma_2^{\xi,\xi}(s,s)+\sum_{k=1}^p\frac{\partial\gamma_2^{\xi,\xi}}{\partial x_k}(s)\,O(h_n)+o(1)\ ,\end{aligned}$$

so, according to (1),

$$\frac{m_2^{\xi,x\xi}(I_n^*)}{n\,h_n^{2p}} \;=\; \frac{\gamma_2^{\xi,\xi}(s,s)}{n\,h_n^p}+o\Big(\frac{1}{n\,h_n^p}\Big)\ .$$

As $\frac{1}{nh_n^2}\,\mathrm{E}^2[\xi(I_n)] = \frac{1}{n}(\frac{\nu(I_n)}{h_n})^2$ is clearly an $O(\frac{1}{n})$, the result follows gathering all these approximations. The other two approximations are proved analogously. ∎

It is possible to be more precise about the factor $O(h_n^2)$ that multiplies the sum of derivatives if we have a more accurate description of the sets involved. Suppose that $I_n = \prod_{k=1}^p (a_{n,k},\, a_{n,k}{+}h_{n,k}]$ with $h_n = h_{n,1}\cdots h_{n,p}$, then looking back to the expansion of $\mathrm{E}[f_n(s)-f(s)]$ we would find the integral

$$\frac{1}{h_{n,1}\cdots\, h_{n,p}}\int_{a_{n,1}}^{a_{n,1}+h_{n,1}}\!\!\cdots\int_{a_{n,p}}^{a_{n,p}+h_{n,p}}\!\Big\langle \nabla f(s),\, t{-}s\Big\rangle\, dt_1\cdots dt_p \;=\; = \sum_{k=1}^{p}\frac{h_{n,k}^2 - 2\,h_{n,k}(s_k - a_{n,k})}{2\,h_{n,k}}\ .$$

To look at the convergence rate of $\mathrm{E}[(\varphi_n(s)-\varphi(s))^2]$ we will write, as in Bosq, Cheze [3],

$$\begin{aligned}
\mathrm{E}\Big[\Big(\varphi_n(s)-\varphi(s)\Big)^2\Big] \;&=\; \frac{\varphi^2(s)}{f^2(s)}\,\mathrm{E}\Big[\Big(f_n(s)-f(s)\Big)^2\Big] \;+\\
&+\;\frac{1}{f^2(s)}\,\mathrm{E}\Big[\Big(g_n(s)-g(s)\Big)^2\Big] - \frac{2\,\varphi(s)}{f^2(s)}\,\mathrm{E}\Big[\Big(g_n(s)-g(s)\Big)\Big(f_n(s)-f(s)\Big)\Big]\\
&+\;\frac{1}{f^2(s)}\,\mathrm{E}\Big[\Big(\varphi_n^2(s)-\varphi^2(s)\Big)\Big(f_n(s)-f(s)\Big)^2\Big]\\
&-\;\frac{2}{f^2(s)}\,\mathrm{E}\Big[\Big(\varphi_n(s)-\varphi(s)\Big)\Big(f_n(s)-f(s)\Big)\Big(g_n(s)-g(s)\Big)\Big]\ .
\end{aligned}\tag{2}$$

Thus, when expanding the last two terms, we will need the convergence rate of $\mathrm{E}[(f_n(s)-f(s))^4]$.

Lemma 3.2. *Suppose the moment measure* $\mathrm{E}(\xi\otimes\xi)$ *satisfies* **(M)**, *that there exists* $R>0$ *such that, for every* $I\subset B$ *and* $k=3,4$,

$$\mathrm{E}[\xi^k(I)] \;\le\; R\,\mathrm{E}[\xi^2(I)]\ .\tag{3}$$

Finally, if (1) holds,

$$\mathrm{E}\Big[\Big(f_n(s)-f(s)\Big)^4\Big] \;=\; O\Big(h_n^4 + \frac{h_n^2}{n\,h_n^p} + \frac{1}{n^2\,h_n^{2p}}\Big)\ .$$

Proof: Write

$$\begin{aligned}
\mathrm{E}\Big[\Big(f_n(s)-f(s)\Big)^4\Big] &= \\
&= \mathrm{E}\Big[\Big(f_n(s)-\mathrm{E}f_n(s)\Big)^4\Big] + 4\,\mathrm{E}\Big[\Big(f_n(s)-\mathrm{E}f_n(s)\Big)^3\Big]\,\mathrm{E}\Big[f_n(s)-f(s)\Big] \\
&\quad + 6\,\mathrm{E}\Big[\Big(f_n(s)-\mathrm{E}f_n(s)\Big)^2\Big]\,\Big(\mathrm{E}\Big[f_n(s)-f(s)\Big]\Big)^2 + \Big(\mathrm{E}\Big[f_n(s)-f(s)\Big]\Big)^4
\end{aligned}$$

and look at each term. From the proof of Theorem 3.1,

$$\Big(\mathrm{E}[f_n(s)-f(s)]\Big)^4 = \Big(O(h_n)\sum_{k=1}^{p}\frac{\partial f}{\partial x_k}(s)+O(h_n^{2p})\Big)^4 = O(h_n^4)$$

and

$$\mathrm{E}\Big[\Big(f_n(s)-f(s)\Big)^2\Big]\,\Big(\mathrm{E}[f_n(s)-f(s)]\Big)^2 = O\Big(\frac{h_n^2}{n\,h_n^p}\Big)\ .$$

Expanding now the third order moment, we find

$$\begin{aligned}
\mathrm{E}\Big[\Big(f_n(s)-\mathrm{E}f_n(s)\Big)^3\Big] &= \\
&= \frac{1}{n^2\,h_n^{3p}}\,\mathrm{E}[\xi^2(I_n)] - \frac{3}{n^2\,h_n^{3p}}\,\mathrm{E}[\xi^2(I_n)]\,\nu(I_n) + \frac{2}{n^2\,h_n^{3p}}\,\nu^3(I_n)\ .
\end{aligned}$$

The last term is an $O(\frac{1}{n^2})$, while the others, using (3)

$$\frac{\mathrm{E}[\xi^3(I_n)]}{n^2\,h_n^{3p}} \le R\,\frac{\mathrm{E}[\xi^2(I_n)]}{n^2\,h_n^{3p}} = \frac{R}{n^2\,h_n^{2p}}\Big(\frac{m_1^{\xi,\xi}(I_n\times I_n)}{h_n^p}+\frac{m_2^{\xi,\xi}(I_n^*)}{h_n^p}\Big) = O\Big(\frac{1}{n^2\,h_n^{2p}}\Big)\ ,$$

$$\frac{\mathrm{E}[\xi^2(I_n)]\,\nu(I_n)}{n^2\,h_n^{3p}} = \frac{1}{n^2\,h_n^{p}}\Big(\frac{m_1^{\xi,\xi}(I_n\times I_n)}{h_n^p}+\frac{m_2^{\xi,\xi}(I_n^*)}{h_n^p}\Big)\,\frac{\nu(I_n)}{h_n^p} = O\Big(\frac{1}{n^2 h_n^p}\Big)\ ,$$

so the sum behaves like $O(\frac{1}{n^2 h_n^{2p}})$. After multiplying by $\mathrm{E}[f_n(s)-f(s)]$ we find then an $O(\frac{h_n}{n^2 h_n^{2p}})$. As for the remaining term, we again expand

$$\begin{aligned}
\mathrm{E}\Big[\Big(f_n(s)-\mathrm{E}f_n(s)\Big)^4\Big] &= \\
&= \frac{1}{n^3\,h_n^{4p}}\,\mathrm{E}[\xi^4(I_n)] - \frac{4}{n^3\,h_n^{4p}}\,\mathrm{E}[\xi^3(I_n)]\,\nu(I_n) + \frac{6}{n^3\,h_n^{4p}}\,\mathrm{E}[\xi^2(I_n)]\,\nu^2(I_n) \\
&\quad - \frac{4}{n^3\,h_n^{4p}}\,\nu^4(I_n) + \frac{3(n-1)}{n^3\,h_n^{4p}}\Big(\mathrm{E}\Big[\Big(\xi(I_n)-\nu(I_n)\Big)^2\Big]\Big)^2\ .
\end{aligned}$$

Applying again (3) and reproducing the same arguments as above, it is easily checked that the sum of the first four terms is an $O(\frac{1}{n^3 h_n^{3p}})$. The last term, again after expansion and using **(M)** is easily found to be an $O(\frac{1}{n^2 h_n^{2p}})$. So as (1) holds we finally get $\mathrm{E}[(f_n(s) - \mathrm{E} f_n(s))^4] = O(\frac{1}{n^2 h_n^{2p}})$, which after summing with the convergence rates of the other terms proves the lemma. ∎

We are now ready to study $\mathrm{E}[(\varphi_n(s) - \varphi(s))^2]$ using the decomposition (2).

Theorem 3.3. *Suppose the moment measures* $\mathrm{E}(\xi \otimes \xi), \mathrm{E}(\eta \otimes \eta)$ *both satisfy* **(M)** *and that there exists* $R > 0$ *such that, for every* $I \subset B$ *and* $k = 3, 4$,

$$\mathrm{E}[\xi^k(I)] \le R\, \mathrm{E}[\xi^2(I)], \qquad \mathrm{E}[\eta^k(I)] \le R\, \mathrm{E}[\eta^2(I)] \tag{4}$$

holds. Further, suppose that there exist real numbers $\beta > \alpha > 0$ *such that*

$$n\, h_n^{4\alpha + 2\beta + p} \to +\infty \tag{5}$$

and that

$$\mathrm{E}\Big(\varphi_n^4(s)\, \mathbb{I}_{\{\varphi_n(s) > h_n^{-\alpha}\}}\Big) \to 0\ , \tag{6}$$

then

$$\begin{aligned}
\mathrm{E}\Big[\Big(\varphi_n(s) - \varphi(s)\Big)^2\Big] &= \\
&= \frac{O(h_n^2)}{f^2(s)} \left(\varphi(s) \sum_{k=1}^{p} \frac{\partial f}{\partial x_k}(s) - \sum_{k=1}^{p} \frac{\partial g}{\partial x_k}(s) \right)^2 \\
&\quad + \frac{1}{n\, h_n^p\, f^2(s)} \Big(\varphi^2(s)\, \gamma_2^{\xi,\xi}(s,s) - 2\, \varphi(s)\, \gamma_2^{\xi,\eta}(s,s) + \gamma_2^{\eta,\eta}(s,s) \Big) \\
&\quad + o\left(h_n^2 + \frac{h_n}{n^{1/2} h_n^{p/2}} + \frac{1}{n\, h_n^p} + \frac{h_n^{1/2}}{n^{3/4}\, h_n^{3p/4}} + \frac{h_n^{3/2}}{n^{1/4}\, h_n^{p/4}} \right) .
\end{aligned}$$

Proof: We will go through each term in (2) to derive the convenient rates for each one. The first three are easily treated as a consequence of the rates derived in the proof on theorem 3.1. In fact, according to that proof, it remains to verify that the two last terms in (2) are an $o\Big(h_n^2 + \frac{h_n}{n^{1/2} h_n^{p/2}} + \frac{1}{n h_n^p} + \frac{h_n^{1/2}}{n^{3/4} h_n^{3p/4}} + \frac{h_n^{3/2}}{n^{1/4} h_n^{p/4}}\Big)$. For this put $\widetilde{\varphi}_n(s) = \frac{\mathrm{E}[g_n(s)]}{\mathrm{E}[f_n(s)]}$ and write

$$\mathrm{E}\Big[\Big(\varphi_n^2(s) - \varphi^2(s)\Big)\Big(f_n(s) - f(s)\Big)^2\Big] =$$

$$= \mathrm{E}\Big[\Big(\varphi_n^2(s) - \widetilde{\varphi}_n^2(s)\Big)\Big(f_n(s) - f(s)\Big)^2\Big] + \Big(\widetilde{\varphi}_n^2(s) - \varphi^2(s)\Big)\, \mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^2\Big] .$$

Obviously $\widetilde{\varphi}_n^2(s) \to \varphi(s)$, and, as seen in the proof of Theorem 3.1, $\mathrm{E}[(f_n(s)-f(s))^2] = O(h_n^2 + \frac{1}{nh_n^p})$, so

$$\Big(\widetilde{\varphi}_n^2(s) - \varphi^2(s)\Big)\,\mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^2\Big] = o\Big(h_n^2 + \frac{1}{n\,h_n^p}\Big)\,.$$

Let $\varepsilon_n = h_n^\beta \to 0$, $\alpha_n = h_n^{-\alpha} \to +\infty$, and write

$$\begin{aligned}
&\mathrm{E}\Big[\Big(\varphi_n^2(s) - \widetilde{\varphi}_n^2(s)\Big)\,\Big(f_n(s) - f(s)\Big)^2\Big] \leq \\
&\quad\leq \Big(\alpha_n + \widetilde{\varphi}_n(s)\Big)\,\mathrm{E}\Big[\Big|\varphi_n(s) - \widetilde{\varphi}_n(s)\Big|\,\Big(f_n(s) - f(s)\Big)^2\,\mathbb{I}_{\{\varphi_n(s)\leq\alpha_n,\,|\varphi_n(s)-\widetilde{\varphi}_n(s)|\leq\varepsilon_n\}}\Big] \\
&\quad + \Big(\alpha_n + \widetilde{\varphi}_n(s)\Big)\,\mathrm{E}\Big[\Big|\varphi_n(s) - \widetilde{\varphi}_n(s)\Big|\,\Big(f_n(s) - f(s)\Big)^2\,\mathbb{I}_{\{\varphi_n(s)\leq\alpha_n,\,|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon_n\}}\Big] \qquad (7)\\
&\quad + \mathrm{E}\Big[\Big(\varphi_n^2(s) - \widetilde{\varphi}_n^2(s)\Big)\,\Big(f_n(s) - f(s)\Big)^2\,\mathbb{I}_{\{\varphi_n(s)>\alpha_n\}}\Big]\,.
\end{aligned}$$

The first term of this expansion is bounded above by

$$\Big(\alpha_n + \widetilde{\varphi}_n(s)\Big)\,\varepsilon_n\;\mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^2\Big] = o\Big(h_n^2 + \frac{1}{n\,h_n^p}\Big)\,,$$

according to the proof of Theorem 3.1, as $\alpha_n\varepsilon_n \to 0$.

The second term in (7) is bounded above by

$$\begin{aligned}
&\Big(\alpha_n + \widetilde{\varphi}_n(s)\Big)^2\,\mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^2\,\mathbb{I}_{\{|\varphi_n(s)-\widetilde{\varphi}_n(s)|>\varepsilon_n\}}\Big] \leq \qquad (8)\\
&\quad\leq \Big(\alpha_n + \widetilde{\varphi}_n(s)\Big)^2\,\Big(\mathrm{E}\Big[\Big(f_n(s) - f(s)\Big)^4\Big]\Big)^{1/2}\,\Big(\mathrm{P}\Big(|\varphi_n(s) - \widetilde{\varphi}_n(s)| > \varepsilon_n\Big)\Big)^{1/2}\,.
\end{aligned}$$

According to Lemma 2.2, we have

$$\begin{aligned}
&\mathrm{P}\Big(\Big|\varphi_n(s) - \widetilde{\varphi}_n(s)\Big| > \varepsilon_n\Big) \leq \\
&\leq \mathrm{P}\Big(\Big|g_n(s) - \mathrm{E}[g_n(s)]\Big| > \frac{\varepsilon_n}{4}\,\mathrm{E}[f_n(s)]\Big) + \mathrm{P}\Big(\Big|f_n(s) - \mathrm{E}[f_n(s)]\Big| > \frac{\varepsilon_n}{4}\,\frac{(\mathrm{E}[f_n(s)])^2}{\mathrm{E}[g_n(s)]}\Big)\,.
\end{aligned}$$

We shall look at the first term arising from this inequality, the other being treated analogously.

$$
\begin{aligned}
&\Bigl(\alpha_n+\widetilde{\varphi}_n(s)\Bigr)^2 \Bigl(\mathrm{E}\Bigl[\Bigl(f_n(s)-f(s)\Bigr)^4\Bigr]\Bigr)^{1/2} \Bigl(\mathrm{P}\Bigl(\Bigl|g_n(s)-\mathrm{E}[g_n(s)]\Bigr| > \frac{\varepsilon_n}{4}\,\mathrm{E}[f_n(s)]\Bigr)\Bigr)^{1/2} \leq \\
&\quad \leq \Bigl(\alpha_n+\widetilde{\varphi}_n(s)\Bigr)^2 \Bigl(\mathrm{E}\Bigl[\Bigl(f_n(s)-f(s)\Bigr)^4\Bigr]\Bigr)^{1/2} \left(\frac{16\,\mathrm{E}\Bigl[\Bigl(g_n(s)-\mathrm{E}[g_n(s)]\Bigr)^2\Bigr]}{\varepsilon_n^2\Bigl(\mathrm{E}[f_n(s)]\Bigr)^2}\right)^{1/2} \\
&\quad = \Bigl(\alpha_n+\widetilde{\varphi}_n(s)\Bigr)^2 \frac{4}{\varepsilon_n\,\mathrm{E}[f_n(s)]}\, O\Bigl(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}\Bigr)\, O\Bigl(\frac{1}{n^{1/2}\,h_n^{p/2}}\Bigr)
\end{aligned}
$$

according to the proof of Theorem 3.1 and Lemma 3.2. As $\widetilde{\varphi}_n(s)\to\varphi(s)$ and $\mathrm{E}[f_n(s)]\to f(s)$, the asymptotic behaviour is given by

$$
\frac{\alpha_n^2}{\varepsilon_n}\, O\Bigl(h_n^2+\frac{h_n}{n^{1/2}\,h_n^{p/2}}+\frac{1}{nh_n^p}\Bigr)\, O\Bigl(\frac{1}{n^{1/2}\,h_n^{p/2}}\Bigr)\ .
$$

The choice of the sequences α_n and ε_n implies that

$$
\frac{\alpha_n^2}{\varepsilon_n}\, O\Bigl(\frac{1}{n^{1/2}\,h_n^{p/2}}\Bigr) \longrightarrow 0\ ,
$$

so the second term in (7) is an $o\Bigl(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}\Bigr)$.

We look now at the third term in (7). Applying Hölder's inequality, this term is bounded above by

$$
\begin{aligned}
&\Bigl(\mathrm{E}\Bigl[\Bigl(f_n(s)-f(s)\Bigr)^4\Bigr]\Bigr)^{1/2} \Bigl(\mathrm{E}\Bigl(\varphi_n^4(s)\,\mathbb{I}_{\{\varphi_n(s)>\alpha_n\}}\Bigr)\Bigr)^{1/2} + \\
&\qquad + \Bigl(\mathrm{E}\Bigl[\Bigl(f_n(s)-f(s)\Bigr)^4\Bigr]\Bigr)^{1/2} \widetilde{\varphi}_n^2(s)\, \Bigl(\mathrm{P}\Bigl(\varphi_n(s)>\alpha_n\Bigr)\Bigr)^{1/2}
\end{aligned}
$$

and this is an $o\Bigl(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}\Bigr)$ according to (6).

To finish our proof, we still have to treat that last term arising in (2). We first apply Hölder's inequality,

$$
\begin{aligned}
&\mathrm{E}\Bigl[\Bigl(\varphi_n(s)-\varphi(s)\Bigr)\Bigl(f_n(s)-f(s)\Bigr)\Bigl(g_n(s)-g(s)\Bigr)\Bigr] \leq \\
&\qquad \leq \Bigl(\mathrm{E}\Bigl[\Bigl(\varphi_n(s)-\varphi(s)\Bigr)^2\Bigl(f_n(s)-f(s)\Bigr)^2\Bigr]\Bigr)^{1/2} \Bigl(\mathrm{E}\Bigl[\Bigl(g_n(s)-g(s)\Bigr)^2\Bigr]\Bigr)^{1/2} .
\end{aligned}
$$

The first factor is further bounded by

$$\sqrt{2}\Big(\mathrm{E}\Big[\big(\varphi_n(s)-\widetilde{\varphi}_n(s)\big)^2\big(f_n(s)-f(s)\big)^2\Big]\Big)^{1/2}+\\ +\sqrt{2}\Big(\mathrm{E}\Big[\big(\widetilde{\varphi}_n(s)-\varphi(s)\big)^2\big(f_n(s)-f(s)\big)^2\Big]\Big)^{1/2},$$

the analysis of which proceeds as the one made for the second term from (2), showing a convergence rate of $o^{1/2}\Big(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}\Big)$. The factor $\mathrm{E}[(g_n(s)-g(s))^2]=O(\frac{1}{nh_n^p}+h_n^2)$, so we finally have a convergence rate $o\Big(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}+\frac{h_n^{1/2}}{n^{3/4}h_n^{3p/4}}+\frac{h_n^{3/2}}{n^{1/4}h_n^{p/4}}\Big)$, which concludes the proof. ∎

Note that condition (6) may be replaced by a more tractable condition leading to a somewhat weaker result, the proof of which goes through the same arguments as the preceding theorem.

Corollary 3.4. *Suppose that the conditions of Theorem 3.3 are satisfied with (6) replaced by*

$$\sup_{s\in B}\mathrm{E}[\varphi_n^4(s)]<\infty\,. \tag{9}$$

Then $\mathrm{E}[(\varphi_n(s)-\varphi(s))^2]=O\Big(h_n^2+\frac{h_n}{n^{1/2}h_n^{p/2}}+\frac{1}{nh_n^p}+\frac{h_n^{1/2}}{n^{3/4}h_n^{3p/4}}+\frac{h_n^{3/2}}{n^{1/4}h_n^{p/4}}\Big)$. ∎

The optimization of h_n indicates one should choose $h_n=c\,n^{-1/(p+2)}$, whether one bases this optimization on the convergence rate given in Lemma 3.2 or in Theorem 3.3. In this case, we get $\mathrm{E}[(f_n(s)-f(s))^2]=O(n^{-2/(p+2)})$, $\mathrm{E}[(g_n(s)-g(s))^2]=O(n^{-2/(p+2)})$, $\mathrm{E}[(f_n(s)-f(s))(g_n(s)-g(s))]=O(n^{-2/(p+2)})$, $\mathrm{E}[(f_n(s)-f(s))^4]=O(n^{-4/(p+2)})$ and $\mathrm{E}[(\varphi_n(s)-\varphi(s))^2]=O(n^{-2/(p+2)})$, thus finding the $n^{-2/(p+2)}$ convergence rate which is well known for density or regression estimation, for example, although (4), (6) or (9) mean some restrictions in each case. Besides the existence of fourth order moments implied by (4) let us discuss condition (9) for some of our examples.

1. The case of regression: we have

$$\mathrm{E}[\varphi_n^4(s)]=\sum_{m=1}^n\mathrm{E}\Big[\varphi_n^4(s)\,\Big|\,\sum_{i=1}^n\xi_i(I_k)=m\Big]\,\mathrm{P}\Big(\sum_{i=1}^n\xi_i(I_k)=m\Big)=$$

$$= \varphi^4(s)\left(1-\mathrm{P}\Big(\sum_{i=1}^n \xi_i(I_k)=0\Big)\right)$$

$$= \varphi^4(s)\left(1-\Big(1-\mathrm{P}(X_i\in I_k)\Big)^n\right) = \varphi^4(s)\left(1-\Big(1-\nu(I_k)\Big)^n\right),$$

so (9) holds if φ, the regression function, is bounded on B.

2. The case of density estimation: we have $\varphi_n(s)=\frac{1}{nh_n^p}\sum_{i=1}^n \delta_{X_i}(I_n)$. Expanding the fourth order power and taking account of the independence and the fact that the variables are 0-1 valued, it is easily checked that

$$\mathrm{E}[\varphi_n^4(s)] = \frac{1}{n^4 h_n^{4p}}\left(\sum_{i=1}^n \nu(I_n) - 10\sum_{i\neq j}\nu^2(I_n) + 6\sum_{\substack{i\neq j,\, i\neq k\\ j\neq k}}\nu^3(I_n) + \sum \nu^4(I_n)\right)$$

where the last summation is over all 4-uples (i,j,k,l) with all four coordinates different. So

$$\mathrm{E}[\varphi_n^4(s)] \;=\; \frac{1}{n^3\,h_n^{3p}} + \frac{10\,c_1}{n^2\,h_n^{2p}} + \frac{12\,c_2}{n\,h_n^p} + c_3$$

which is finite and independent of the point s.

3. Looking back at the computation made for the case of regression, it still holds if the point process ξ is represented as $\sum_{i=1}^N \delta_{X_i}$, meaning some counting of points, as it is the case of the examples mentioned in the beginning, except for the density estimation. That is, in all such cases (9) will hold if the function we try to estimate is bounded on the compact set B. So, for instance in the thinning case, if the thinning function p is continuous (9) holds.

REFERENCES

[**1**] Bensaïd, N. and Fabre, J. – Convergence de l'estimateur à noyau de dérivées de Radon–Nikodym générales dans le cas mélangeant, *Canadian J. Statist.,* 26 (1998), 267–282.

[**2**] Bensaïd, N. and Jacob, P. – Erreur quadratique asymptotique optimale pour les estimateurs à noyau de dérivées de Radon–Nikodym générales, *Ann. Inst. Statist. Univ. Paris,* XXXXII (1998), 3–14.

[**3**] Bosq, D. and Cheze, N. – Erreur quadratique asymptotique optimale de l'estimateur non paramétrique de la régression pour des observations discrétisées d'un processus stationnaire à temps continu, *C. R. Acad. Sci. Paris,* 317 (1993), Série I, 891–894.

[**4**] Ferrieux, D., – *Estimation de Denstités de Mesures Moyennes de Processus Ponctuels Associés*, PhD dissertation, Univ. Montpellier II, 1996.

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

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

[**7**] Jacob, P., Mendes Lopes, N. – Un théorème central limite fonctionnel pour un estimateur d'un noyau de transition, *Port. Math.,* 48 (1991), 217–231.

[**8**] Jacob, P. and Niéré, L. – Contribution à l'estimation des lois de Palm d'une mesure aléatoire, *Pub. Inst. Stat. Univ. Paris,* 35 (1990), 39–49.

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

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

[**11**] Jacob, P. and Oliveira, P.E. – *Using non embedded partitions in histogram estimation*, Preprint, Pré-publicações Dep. Matemática, Univ. Coimbra, 97–25, 1997.

[**12**] Roussas, G.G. – Kernel estimates under association: strong uniform consistency, *Stat. and Probab. Letters,* 12 (1991), 393–403.

[**13**] Roussas, G.G. – Curve estimation in random fields of associated processes, *J. Nonparam. Stat.,* 2 (1993), 215–224.

[**14**] Roussas, G.G. – Asymptotic normality of random fields of positively or negatively associated processes, *J. Multiv. Analysis,* 50 (1994), 152–173.

Paulo Eduardo Oliveira,
Univ. Coimbra, Dep. Matemática,
Apartado 3008, 3000 Coimbra – PORTUGAL